ISBN 978-82-575-1022-0
ISSN 1503-1667

Norwegian University of Life Sciences
NO–1432 Ås, Norway
Phone +47 64 96 50 00
www.umb.no, e-mail: postmottak@umb.no

Heli Havukainen

Norwegian University of Life Sciences • Universitetet for miljø- og biovitenskap
Department of Chemistry, Biotechnology and Food Science
Philosophiae Doctor (PhD) Thesis 2011:59

Dissecting molecular properties of honey bee vitellogenin: a protein acting at the intersection between social behavior and aging

Disseksjon av vitellogenins molekylære egenskaper:
et protein med funksjoner i grenseflaten mellom sosial adferd og aldring

Heli Havukainen

Philosophiae Doctor (PhD) Thesis 2011:59

Dissecting molecular properties of honey bee vitellogenin:
a protein acting at the intersection between social behavior and aging

Disseksjon av vitellogenins molekylære egenskaper:
et protein med funksjoner i grenseflaten mellom sosial adferd og aldring

Philosophiae Doctor (PhD) Thesis

Heli Havukainen

Department of Chemistry, Biotechnology and Food Science
Norwegian University of Life Sciences

Ås 2011

Thesis number 2011:59
ISSN 1503-1667
ISBN 978-82-575-1022-0

Table of contents

Preface

First things first: big thanks to my family and the good ol' friends in Finland for your love and support that knows no country boundaries. For all my Biorecognizer colleagues and friends (mainly both): you have made these three years unforgettable. Thank you for the support, I would not have managed without. You are an essential part of my thesis and my heart. Those students/post-docs/technicians who participated the research and the writing in practice were (in alphabetical order): Anne Baumann, Erin Fennern, Daniel Münch, Lars Skjærven, Bente Smedal, Randi Svebak, Knut Teigen, Jarl Underhaug and Florian Wolschin. Special hugs in addition to the "Naboen gang" Helene & Marte & Åge & Ana! Thanks to all members of the Amdam group in Ås and Phoenix for your support and the inspiring conversations! Thanks to my honey-sweet best friends of Norway: Rebecka, Jonna, Anniina & Dante. Thanks to the great proteomics people of PROBE at University of Bergen and Protein Chemistry at University of Helsinki (especially Nisse Kalkkinen), who all did great research in collaboration. Thanks to Claus Kreibich and Alf Søyland who eagerly helped me with the living bees!

For my supervisors

Aurora Martinez: I can never thank you enough for the perfect laboratory environment – equipment and social wise – that you offered for the duration of my thesis. I will praise you wherever I go.

Øyvind Halskau: You always supported me, calmed me down and cheered me up. Thank you for your patience, the NMR-knowledge you shared, and of course, all the jokes and discussions (plus the Iron Maiden gig).

Gro Amdam: Thank you for giving me this incredible chance and sculpting me into a real researcher. The most important thing you made me realize is that science is not about work. It is about passion. Work comes along.

The cover figure was the cover of The Journal of Experimental Biology, where Paper I of this thesis was published (Havukainen et al., 2011). It is a combination of the vitellogenin N-sheet domain model and a photo taken by Christofer Bang. The design and layout was done by the author.

Aristotle (~ 384-322 BC) in History of Animals, Book VIII(IX):

"The older bees work inside, and are hairy because they stay in, while the young ones fetch from outside and are smoother." (Harvard University Press, 1991)

A comment based on the twentieth century science:

It is the opposite: the young bees do the indoor work and the elder bees do the foraging. Briefly, the age-related behavioral shift is caused by a drop in the titer of a major hemolymph (blood) protein in the aging worker bee. This behavior coordinating protein is called vitellogenin.

Abstract

Vitellogenin is a central honey bee (*Apis mellifera*) life-history regulator. This thesis represents an initiative to study this protein, which affects aging and task-division of the bees, at the molecular level. I have used, among others, molecular modeling, nuclear magnetic resonance spectroscopy and surface plasmon resonance techniques to understand the structure and properties of vitellogenin. In addition to the three experimental papers of this thesis, vitellogenin is discussed from the molecular viewpoint in two invited publications (Papers II-III). The studies have resulted in more detailed understanding of the structural organization of the protein and its modifications: the novel findings include the cleavage of honey bee vitellogenin at a phosphorylated polyserine linker site, and the ability of the protein to bind to membranes and to interact with *Escherichia coli*. Thus, this vitellogenin study exceeds its primary molecular scope, and provides a new perspective on the protein as a membrane-active factor in bee physiology. Since little is known of the detailed molecular properties of insect vitellogenins in general and honey bee vitellogenin in particular, the thesis also contributes towards expanding vitellogenin molecular biology beyond the more studied vertebrate vitellogenins.

Abstrakt

Vitellogenin er et sentralt protein i reguleringen av livsutviklingen hos honningbie (*Apis mellifera*). Dette proteinet påvirker aldring og oppgavefordeling hos bier, og avhandlingen er en studie av proteinet på molekylært nivå. Jeg har blant annet brukt molekylær modellering, kjernemagnetisk resonans spektroskopi og en teknikk for å måle overflateplasmonresonans (surface plasmon resonance) for å forstå strukturen og egenskapene til vitellogenin. I tillegg til de tre eksperimentelle artiklene i denne avhandlingen er vitellogenin diskutert fra et molekylært synspunkt i to inviterte publikasjoner (Papers II-III). Studiene har resultert i en økt forståelse av den strukturelle oppbygning av proteinet og dets modifikasjoner. Blant viktige funn i avhandlingen kan det nevnes at vitellogenin blir spaltet på et fosforylert polyserin-linker sted, at det binder seg til membraner og at det interagerer med *Escherichia coli*. Denne studiens omfang dekker dermed mer enn det primært molekylære, og gir nye perspektiv på proteiner som en membran-aktiv faktor i biefysiologi. Detaljert kjennskap om de molekylære egenskapene til vitellogenin hos insekter, og hos honningbie spesielt, er mangelfull, og denne avhandlingen gir derfor et bidrag til å utvide kunnskapen om molekylærbiologien til vitellogenin utover de mer studerte vitellogeninene hos vertebrater.

The Norwegian translation was provided by Helene Bustad Johannessen, Marte Flydal and Jarl Underhaug.

List of papers

I Havukainen, H., Halskau, Ø., Skjaerven, L., Smedal, B., Amdam, G. V. (2011) Deconstructing honey bee vitellogenin: novel 40 kDa fragment assigned to its N terminus. *J. Exp. Biol.* 214, p. 582-592.

II Amdam, G. V., Fennern, E., Havukainen, H. (2011) Vitellogenin in honey bee behavior and lifespan. *Honey bee Neurobiology and Behavior*, Springer. Chapter 1.2.

III Havukainen, H., Halskau, Ø., Amdam, G. V. (2011) Social pleiotropy and the molecular evolution of honey bee vitellogenin. *Mol. Ecol.* In press.

IV Havukainen, H., Underhaug, J., Wolschin, F., Amdam, G. V., Halskau, Ø. A vitellogenin polyserine cleavage site: highly disordered conformation protected from proteolysis by phosphorylation. *Submitted.*

V Havukainen, H., Münch, D., Baumann, A., Halskau, Ø., Amdam, G. V. Membrane binding of vitellogenin. *Manuscript.*

Other publications:

Hunt, J. H., Mutti, N. S., Havukainen, H., Henshaw, M. T., Howe, M. K., Amdam, G. V. (2011) Development of an RNA interference tool, characterization of its target, and an ecological test of caste differentiation in the eusocial wasp Polistes. *PLoS ONE.* In press.

Abbreviations

NMR	nuclear magnetic resonance spectroscopy
SPR	surface plasmon resonance
Vg	vitellogenin (protein)
vg	vitellogenin (gene)
Vgs	vitellogenins
apoB	apolipoprotein B
MTP	microsomal triglyceride transfer protein
vWFD	von Willebrand factor type D
JH	juvenile hormone
kDa	kilo Dalton
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
MALDI-TOF	matrix-assisted laser desorption/ionization time-of-flight
LC-MS	liquid chromatography mass-spectrometry
CD	circular dichroism
DLS	dynamic light scattering
MRI	magnetic resonance imaging
FID	free induction decay
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
LLTP	large lipid transfer proteins

1. Introduction

1.1 Honey bee as a study target

The honey bee has charmed as a study object since the early natural scientists (Wilson, 2004); the bees provide highly appreciated goods and services in the form of honey and crop pollination and maintain a fascinatingly well organized, industrious society. This society can be referred to as a superorganism, since the individuals are highly interdependent and best seen a discrete unit (Tautz, 2008). While studies of the bee society have a long history, only relatively recently has modern biology produced tools for examining the molecular basis of the bee colony. The sequencing of the honey bee genome (Weinstock et al., 2006) elevated the bee to the company of the classical model organisms. A specialty of the bee among the model organisms is the small size in combination with a big brain, learning and memory skills, as well as its highly distinct social organization. The modern tools can be employed in search for answers to the old questions: how do the bees orchestrate their tasks in the colony? Why do the forager bees live only a couple of weeks, while in the winter time, the bees survive many months till the spring? The protein I study is deeply involved in these questions, as I will describe in the following two chapters.

1.2 Vitellogenin and the social organization

Vg is a pleiotropic gene; it affects the fecundity of the queen but also diverse traits in worker bees. But, first, to understand the importance of these traits within the complex structure of a bee society, an introduction to the social organization of the hive will be given.

A colony of honey bees consists of one queen, several drones during the reproductive season, and thousands of worker bees. The queen is the mother of all the individuals, continuously laying eggs during the warm months of the year.

The worker bees are all female as well, but, in contrast to the queen, they typically remain facultatively sterile. An extraordinary feature of worker bees is the highly advanced division of labor, with workers that, during maturation, progress through a sequence of different social tasks

(age polyethism). A newly hatched worker, typically, will first engage in cleaning tasks. Then, within the next days, its food-producing glands (hypopharyngeal glands) will be fully developed, and these "nurse bees" engage in brood care, which includes feeding and cleaning tasks. Fully developed wax glands, moreover, allow bees to participate in constructing the building blocks of the hive – the combs – with hundreds of brood and food storage cells. Finally, at approximately 16-20 days of age, nurses change to "foragers"; now with a fully developed flight machinery that enables them to collect nectar or pollen often far away from the hive (Davis, 2006; Seeley, 2010; Tautz, 2008).

The egg-laying queen with large, fully active ovaries is rich in an egg yolk precursor protein called vitellogenin (Vg). In the majority of oviparous species from insects to fish and birds, some of the amino acid and lipid proportion of the eggs is derived from this lipoprotein (reviewed by Finn, 2007; Tufail and Takeda, 2008). In insects, Vg is produced in the fatbody, an adipose tissue comparable to the liver of the vertebrates. Vg is transported to the hemolymph (blood). From there it is taken up by the oocytes in a receptor-mediated endocytosis (Sappington and Raikhel, 1998). In the developing eggs, Vg is typically cleaved, and called (lipo)vitellin. Notably, the human proteins apolipoprotein B (apoB) and microsomal triglyceride transfer protein (MTP) are evolutionarily related to vitellogenins (Vgs) (Baker, 1988b; Mann et al., 1999), having lost the egg-related function, but kept the lipid-transport property.

Seemingly contradictory to the yolk related function of Vg, the protein is also highly abundant in honey bee individuals that do not lay eggs. This is true for female worker bees, and even drones (Trenczek and Engels, 1986) and larva (Guidugli-Lazzarini et al., 2008) can express *vg*. For my studies, I purified Vg from winter time worker bees (*diutinus*) that may posses up to 100 mg/ml Vg in their hemolymph (Amdam et al., 2004a). Also the fatbody tissue of the *diutinus* bees contains excessive protein stores in the form of Vg granules (Smedal et al., 2009). Such an abundant protein in non-reproductive individuals likely has additional functions other than being an egg-yolk precursor.

Beyond its direct reproductive role, honey bee Vg was shown to affect behavior by regulating the nurse-forager task division (Munch and Amdam, 2010). Specifically, younger bees with high Vg and low juvenile hormone (JH) levels engage in nursing tasks, while the usually older foragers are low in Vg and high in JH (Fig. 1). In a negative feedback circuit, high Vg titer may suppress JH (Amdam and Omholt, 2003). Among other experiments, RNAi mediated *vg*

silencing has verified that Vg and JH together control the behavioral shift (Guidugli et al., 2005; Nelson et al., 2007). Intriguingly, the Vg-JH feedback differs from what is known in other insects, which implies that the honey bee exploits a common insect hormone and an egg-yolk protein in a highly adapted manner (Page et al., 2006). Other, less studied factors are also suggested to be involved in the nurse-forager shift, including genes *amfor* and *malvolio* and the neuromodulator octopamine (reviewed by Ament et al., 2010).

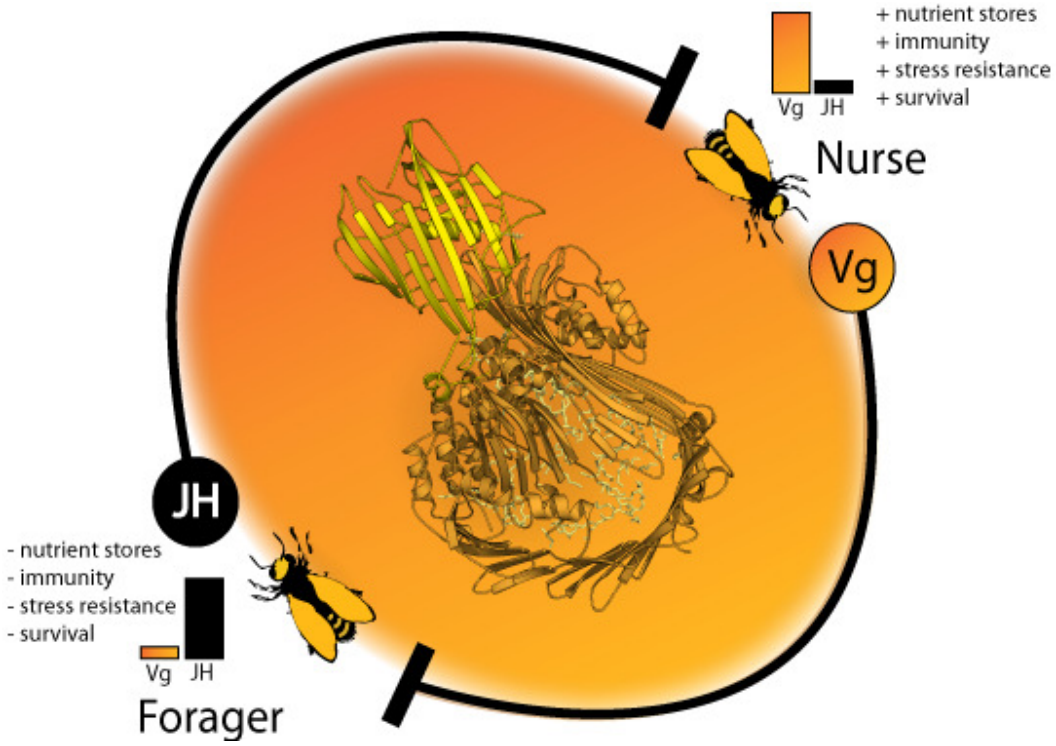


Figure 1. The control of the nurse-forager transition by Vg and JH was explained by a double repressor model, in which high Vg levels suppress JH (nurses) and high JH levels, in turn, reduce Vg (foragers) (Amdam et al., 2003). Importantly, nurses with ample Vg also show longer life span, as compared to very short lived foragers with low Vg titers in their hemolymph. In the middle, lamprey (*Ichthyomyzon unicuspis*) lipovitellin – a homologous protein – is used to illustrate Vg (PDB-ID: 1lsh). The phospholipid cargo is shown as sticks inside the major lipid-binding cavity (orange) of lamprey lipovitellin. The figure is prepared by the author based on Fig. 2 of Paper III (Havukainen, H.) and Fig. 3 of Paper II (Amdam, G.V.).

1.3 Vitellogenin and longevity

Nurse bees possess greater nutrient stores, and their immunity, stress resistance and survival capacity are improved relative to the more frail foragers (for review, see Munch et al., 2008). The winter-time *diutinus* bees have the longest life span of all worker bees. One mechanism, where Vg may explain the extreme life span of the *diutinus* bees is its function as a nutrient storage. Vg as a metabolic resource not only can serve as proteinaceous larval food after

processing in the hypopharyngeal glands (Amdam et al., 2003), but is also exchanged between and used by adult bees (Crailsheim, 1991). In this context, the lipoprotein Vg can be a store of amino acids, lipids and trace elements like zinc, copper and iron (the two latter elements verified in turkey lipovitellin by Richards, 1997). Yet, what is the link between Vg and the other life span promoting characteristics mentioned above?

Bees rich in Vg show improved resistance against oxidative stress, while bees low in Vg are more susceptible to oxidative damage (Seehuus et al., 2006a). When bees are exposed to paraquat, a drug that induces oxidative stress, Vg becomes highly carbonylated. This indicates that Vg itself can scavenge reactive oxygen species (Seehuus et al., 2006a). The antioxidant effect of Vg is not the only benefit the protein has on lifespan. Previous work shows that high Vg titers are linked to more viable hemocytes (insect immune cells) in the hemolymph, a function Vg might perform through donating zinc (Amdam et al., 2004b; Amdam et al., 2005a).

Honey bees show an extraordinary plasticity of aging, where Vg seems to be in a key position. For example, foragers that are low in Vg also experience accelerated age-related decline, exemplified by high apoptosis rate of hemocytes, more abundant oxidative carbonylation of proteins in the brain (Seehuus et al., 2006b) and impaired learning performance (Behrends et al., 2007; Munch and Amdam, 2010). Strikingly, age related decline in bees depends on social task rather than chronological age, as foragers develop signs of cellular and behavioral senescence more rapidly than nurse and winter bees (see, e.g. Dukas, 2008; Munch and Amdam, 2010). However, where the honey bee model might make its most important contribution to understanding aging is its potential to reveal mechanisms of reversed ontogeny and reversal of senescence. Reversed ontogeny occurs when bees change back from an older to a usually younger phenotype. This can be achieved, when nurse bees are removed from the hive and foragers are forced to become nurses again (Robinson et al., 1992). These reverted individuals show characteristics that are typical for nurses, including raised Vg and hemocyte titers – i.e. they undergo reversal of immunosenescence (Amdam et al., 2005b).

In all, Vg confers survival benefits, which include positive effects on immune cells, and increased resistance against oxidative stress. However, exactly how Vg can exert these effects and how it can support extreme longevity in winter bees is not well understood. The same holds true for understanding the way Vg structure is linked to Vg function at the molecular level, which is the motivation for my research.

1.4 The molecular characteristics of vitellogenin

Honey bee Vg (UniProtKB ID: Q868N5) has 1770 amino acid residues, of which the 16 first form a signal sequence that is cut off and not present in the final protein product (Piulachs et al., 2003). Cross-species studies assign Vg a phosphoglycolipoprotein – a member of the group of large lipid-transfer proteins (LLTPs) (Avarre et al., 2007; Finn, 2007; Tufail and Takeda, 2008; Wheeler and Kawooya, 1990). On SDS-PAGE protein gel, Vg appears as a 180 kDa band (Wheeler and Kawooya, 1990), and several articles report an additional 150 kDa band that is recognized by the Vg antibody (Amdam et al., 2003; Seehuus et al., 2007; Wheeler and Kawooya, 1990). The 150 kDa Vg has been suggested to be an unidentified Vg fragment or even another unidentified type of Vg (Bitondi and Simoes, 1996; Pinto et al., 2000), but these previous studies reached no conclusion of the origin of this fragment.

Most of the knowledge on honey bee Vg protein chemistry to date comes from an over 20-year-old article, the only one with purification information for honey bee Vg; there Wheeler and Kawooya (1990) produced pure vitellogenin from queen hemolymph. They reported ample aggregation and fragmentation tendency of Vg – not a promising sign for protein work.

The extensive collection of bioinformatic tools on, e.g., www.expasy.org were employed to produce crude analyses based on the primary structure of vitellogenin and to produce an initial analysis of the domain architecture of Vg. These domains are visualized in Figure 2A. In the N-terminus, there is a domain that is named N-sheet in lamprey (*I. unicuspis*) lipovitellin. This domain is followed by a serine-rich segment called the polyserine tract (Tufail and Takeda, 2008). The next domain in sequence is called α-helical and the C-terminal tail is occupied by a von Willebrand Factor D domain (vWFD). This domain has sequence similarity to human blood clotting factor, von Willebrand Factor D (Baker, 1988a).

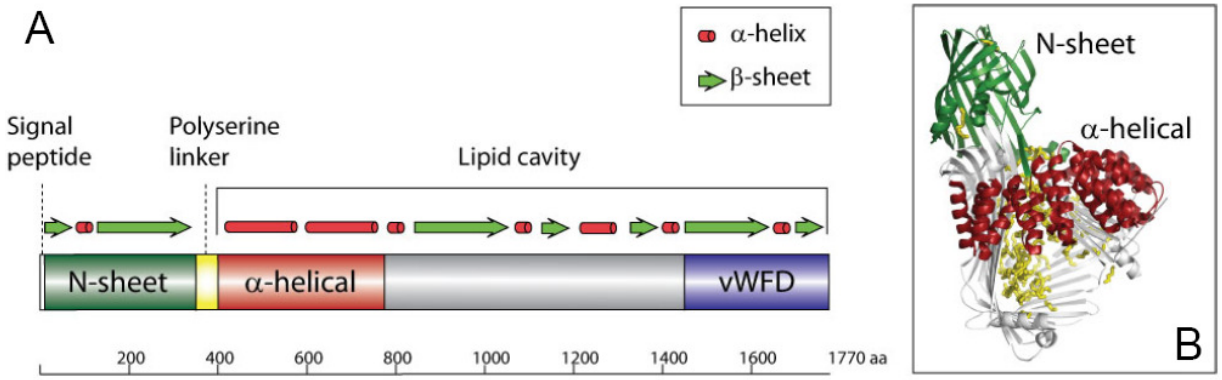


Figure 2. Structural prediction of honey bee Vg. A) The domains of Vg. The secondary structure prediction was prepared using PSIPRED (Jones, 1999). B) The crystal structure of lamprey lipovitellin – the only existing structure of the Vg family (Anderson et al., 1998). The N-sheet (green) and the α-helical domain (red) are highlighted. The lipids are shown as yellow sticks. The figure is from our Paper II, prepared by the author.

The structural understanding of Vgs largely relies on the only experimental structure of the Vg family studied this far: the X-ray structure of lamprey lipovitellin (Fig. 1 and 2B). This protein has a tertiary structure of a funnel filled with phospholipids. The lipids are thought crucial for the folding of Vg as the lipids are loaded cotranslationally (Anderson et al., 1998). The secondary structures of lamprey lipovitellin comprise large β-sheets, a half-belt of α-helices (α-helical domain) and several extended coils. The most conserved parts of all Vgs are the N-sheet and the α-helical domain (Babin, 2008; Baker, 1988b). As phylogenetically distant homologues as the human apoB and MTP proteins have been modeled using these most conserved domains of lamprey lipovitellin (Mann et al., 1999).

Two important parts of the honey bee Vg are not present in the lamprey lipovitellin structure. One is the polyserine linker (Fig. 2A), which is a likely multiple phosphorylation site and an insect-specific hypervariable segment of unknown function (Tufail and Takeda, 2008). Lamprey and other vertebrates lack this segment. The other is the vWFD that is present in lamprey lipovitellin sequence, but not in the structure, likely due to high flexibility at the connection of this C-terminal domain to the remaining part of the protein (Anderson et al., 1998).

Summarizing, honey bee Vg is a large lipid transporting protein, expected to be heavily post-translationally modified based on studies in other species (Tufail and Takeda, 2008). The X-ray structure of lamprey lipovitellin can be used as a template for structural considerations of Vgs, as done by Mann et al. (1999) with the human homologues. The protein contains interesting regions that have been little studied, such as the insect-specific polyserine tract. Honey bee Vg

offers the luxury and the frustration of studying a nearly uncharacterized protein. It is a fascinating target, but there are not many articles to absorb ideas or protocols from, and therefore basic characterization of the protein is necessary to get started with in-detail studies.

2. Methods

2.1 Overview of the methods

Escherichia coli expression of full-length Vg was deemed very difficult due to considerations of size and the lipid load of the protein. Furthermore, reasonable amounts of wild-type protein was available from natural sources. We collected Vg-rich *diutinus* bees that are plentiful in the hive compared to the queen. Vg is abundantly stored in their abdominal adipose tissue, the fatbody (Roma et al., 2010). Thus, we used not only hemolymph, which is time-consuming to sample micro liter after micro liter from the living bees, but also quick-to-collect abdomens, emptied of the gut and the ovaries. We invented a three-step chromatography purification protocol (size-exclusion, ion-exchange and Concanavalin A affinity) for the fatbody samples, based on the queen hemolymph purification protocol of Wheeler and Kawooya (1990), who used ultracentrifugation, ion-exchange and Concanavalin A affinity chromatography. When it comes to *diutinus* hemolymph, I was able to produce purified Vg with a carefully fractioned single ion-exchange step (Paper V; for the level of purity, see Paper I); the purity can be further improved with Concanavalin A affinity chromatography as used by Wheeler and Kawooya (1990), but there is a significant sample loss during this step. Therefore, the affinity-step was excluded in Paper V.

Vg is steadily getting cleaved into two pieces during purification and some of the protein is in two pieces in the raw protein extract of abdominal samples prior to the purification. We realized that this was due to tissue-specific Vg cleavage (Paper I), and mass-spectrometry was used to identify the cleavage fragments. Mass-spectrometry was in fact used in all the experimental papers in close collaboration with PROBE proteomic unit at University of Bergen and Protein Chemistry Research Group and Core Facility at the Institute of Biotechnology, University of Helsinki. Vg (or piece of Vg) was separated from other proteins using electrophoresis, the protein band on gel was excised and the protein was enzymatically digested into peptides. These polypeptides can be separated and measured based on their weight/charge using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) or liquid chromatography mass-spectrometry (LC-MS). The result is analyzed based on calculated peptide masses and can lead into peptide/protein identification.

Based on our Vg domain prediction (Fig. 2), as well as practical considerations, we designed and purchased (from a collaborative company GenScript, NY, USA) the α-helical domain and the vWFD domain, expressed in *E. coli*. Also, the polyserine linker of our studies was commercially synthesized with partial isotope labeling (CPC Scientific, CA, USA). The fragments were used in the thesis work to assign properties of Vg to the putative domains (Paper IV and V).

Many assays in this thesis are gel or blot based, since these approaches can visualize the specific fragmentation of Vg and because there was an established and sensitive Vg antibody available (Seehuus et al., 2007). The gel assays were combined with phosphorylation specific staining and deglycosylation for monitoring post-translational modifications (Paper I) or with limited proteolysis (Paper IV and V) in order to observe, which parts of Vg are readily cleavable (Sharp et al., 2006; Stroh et al., 2005).

Homology modeling was used as a tool for the characterization of Vg throughout the thesis. An X-ray or NMR structure of a protein can be used as a template for modeling the structure of another homologous protein (or the same but modified protein); for honey bee Vg, the template is the lamprey lipovitellin structure. A good model is based on a comprehensive sequence alignment, and hence, I included eight Vgs of various taxa in the sequence alignment prior to model building. Furthermore, I verified that the most conserved parts were correctly aligned with the assistance of a published multiple sequence alignment of 28 Vgs and homologous proteins including human MTP and apoB (Avarre et al., 2007). This is an important step, because the general sequence identity is low between any Vgs, but certain stretches are extremely conserved (Avarre et al., 2007). Then, I inspected manually the sequence and the secondary structure predictions to secure that the secondary structures to be modeled are feasible. For instance, proline residues would be unusual in the middle of an α-helix or a β-sheet, due to their backbone rigidity. The alignment was performed and the models were build using program Bodil (Lehtonen et al., 2004), and energetically minimized using Amber (Case D.A., 2008) or Discovery Studio (Accelrys, San Diego, CA, USA). This protocol was used for the modeling of both the N-sheet (Paper I) and the α-helical domain (Paper V). In these domains, very conserved amino acid residues are found at structurally important parts, such as in a hydrophobic helix inside the half-β-barrel of the N-sheet and in the helices of the α-helical domain.

Methods that I used to a large extent, but which did not end up that much in the papers are circular dichroism (CD) spectroscopy and dynamic light scattering (DLS). Both methods are useful for controlling the quality of a protein sample. I used CD spectroscopy (190-260 nm) together with NMR in Paper IV to verify the secondary structure content of our Vg polyserine peptides. This method uses circularly-polarized light, which has two wave components with a phase difference but the same amplitude. If an optically active sample is applied, the sample absorbs light and can change the amplitude of the two light components relative to each other, leading into elliptically polarized light. This ellipticity is measured against a wavelength spectrum, and since protein secondary structures contribute to the spectrum differently, the secondary structure composition of the sample can be estimated. DLS, on the other hand, is another type of optical method that can be used to measure the size distribution of the sample particles. DLS is based on measuring the intensity fluctuations of a scattered laser beam. These fluctuations are caused by the Brownian motion of the particles in the sample, and the particle sizes can be derived from this information. I used DLS prior to many experiments for testing the level of aggregation.

The most complicated methods used in the thesis were NMR and SPR, of which the former is introduced in detail in the next chapter. SPR is a method for real time quantification of biomolecular interactions, where one binding partner is immobilized on a sensor chip and the other is then applied over it at a constant flow rate. Binding causes a change in the angle of light reflected by the sensor chip, which is optically measured and reported in response units. The response units are proportional to the increase of the mass on the chip. We used SPR for measuring the binding of Vg on membrane-mimicking phospholipid bilayers of neutral or negative net charge (phosphatidylcholine or phosphatidylserine mixed with phosphatidylcholine, respectively). The liposomes were immobilized on a specific chip with hydrophilic groups (L1-chip, Biacore, GE Healthcare, Uppsala, Sweden) and Vg (purified from hemolymph and fatbody) and two recombinant domains (α-helical and vWFD) were applied as the analytes.

2.2 Nuclear magnetic resonance spectroscopy

For references to the general NMR information, see:
(Friebolin, 1993; Levitt, 2009; Rule and Hitchens, 2006)

NMR is used for various chemical, biological and medical studies (Fig. 3). An NMR experiment is essentially a manipulation of the quantum state of the sample by the constant magnetic field of the NMR instrument and by pulses of radio frequency radiation (hence: radio waves) applied by the experimentalist. Radio waves are sent to the sample, which has a smaller magnetic field induced by the large field of the instrument. This causes quantum transitions in the spin-states of susceptible nuclei, and the resulting oscillating magnetic field is recorded and interpreted. For me, this was an exiting, complicated new method to learn, used in Paper IV for studying the polyserine tract of Vg. Next, I will describe the basics of *nuclear magnetic resonance spectroscopy* word by word.

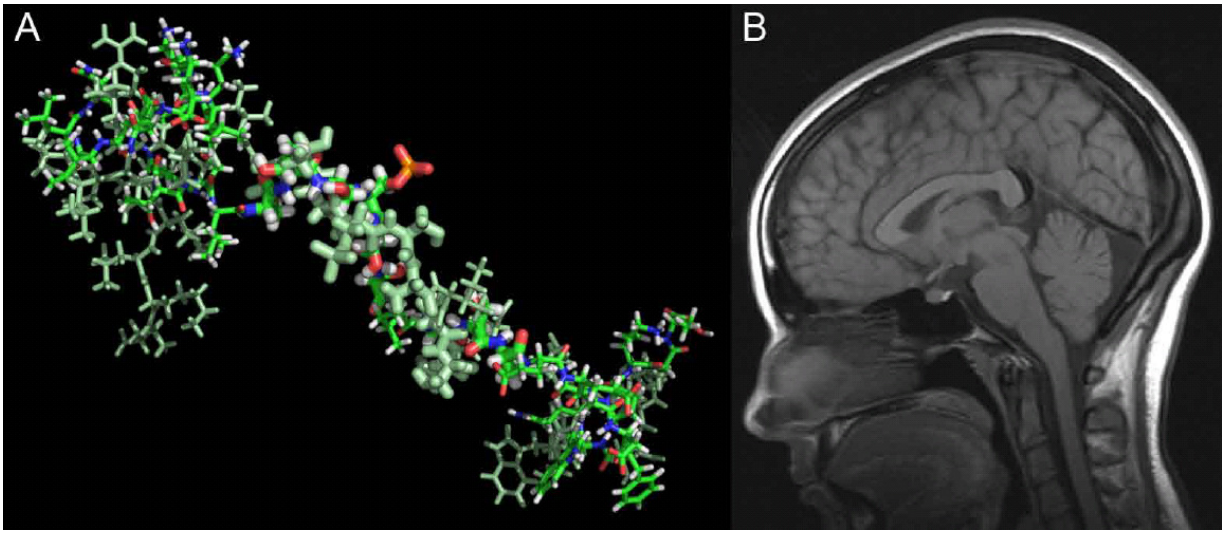


Figure 3. NMR is used for studying interatomic distances and relative bond orientations, and for inspecting internal tissues. A) Two superimposed structures of a honey bee Vg polyserine peptide calculated based on NMR constraints. B) The author's brain visualized by magnetic resonance imaging (MRI by Unilabs, Bergen).

Nuclear – NMR spectroscopy records signals derived from atomic nuclei. Not all nuclei can be measured by NMR, only those with nuclear angular momentum, i.e. spin, *S*, are measurable:

Equation 1

$$S = \sqrt{I(I+1)}\hbar$$

where I is the spin quantum number and $\hbar$ is the reduced Planck's constant (6.62×10^{-34} J s / 2 π). Nuclei having spin quantum number 0 can not be observed by NMR spectroscopy, since according to Equation 1, they have 0 spin. For instance, the most abundant carbon isotope, ^{12}C, has $I = 0$, and is invisible to NMR. The most relevant isotopes in biological NMR are ^{1}H, ^{13}C and ^{15}N, all having I = ½. ^{1}H is the most common hydrogen isotope, but the carbon and nitrogen isotopes are rare. This is why NMR samples are often labeled with ^{13}C and ^{15}N isotopes. The labeling can be done during peptide synthesis, or labeled proteins can be produced with recombinant techniques by supplying the bacteria with isotope-containing media.

Magnetic – All charged particles with intrinsic motion (i.e. spin) have a magnetic moment μ. The connection between the magnetic moment μ and the spin S is the following:

Equation 2

$$\mu = \gamma S$$

where γ is the gyromagnetic ratio [10^7 rad T^{-1} s^{-1}] specific for each nucleus type. ^{1}H has the highest gyromagnetic ratio of the most biologically relevant nuclei: 26.75 x 10^7 rad T^{-1} s^{-1}.

The magnetic moment μ is like a tiny magnetic field of the nucleus and can be presented as a vector. Similarly to a compass needle, which attempts to align to Earth's magnetic field, the magnetic moment of a nucleus aligns according to an external magnetic field. In the NMR instrument, thus, the magnetic moment of each nuclei align with the magnetic field B_0 of the instrument. However, the orientation is quantized, meaning that it can only have certain values:

Equation 3

$$\mu_z = m\gamma\hbar$$

where μ_z is the component of the magnetic moment orienting along the direction of the external magnetic field B_0 and m is the magnetic quantum number, also called the directional quantum number. The spin and m are connected so that the amount of m values is ($2I$ + 1). In the case of ^{1}H, ^{13}C and ^{15}N that have spin ½, there are two possibilities for m: -½ and ½, since m = (2 x ½ +

1) = 2. These two stages are often presented in the terms of classical physics as a double-cone, with vectors precessing on the surface of the upper (m = ½) and the lower (m = -½) half of the cone, with the surface of the cones at 54.44º for I = ½ relative to the x-y plane perpendicular to the external magnetic field, B_0. In an external magnetic field, the magnetic moments of each nucleus precess around the axis of B_0 at the Larmor frequency, ω, which in turn is related to resonance.

Resonance – The radio waves that are used to irradiate the nuclei have to be of certain frequency corresponding to the energy differences between two quantum spin levels. These waves are said to be "in resonance" with the frequency of the precession of the nuclear magnetic moment around B_0. The Larmor frequency, ω, of a nucleus, is calculated from the gyromagnetic ratio of the nucleus and of the strength of the magnetic field:

Equation 4

$$\omega = -\gamma B$$

Since the Larmor frequency differs for ^{1}H, ^{13}C and ^{15}N at a given magnetic field, radio waves of different frequency have to be used for each nucleus, if data of each of them is desired.

Spectroscopy – Spectroscopic methods, in principle, rely on measuring energy differences of molecules using electromagnetic waves; in the case of NMR, these are of radio frequency, corresponding to very small differences in energy. In detail, NMR is a spectroscopic method of nuclear energy levels called Zeeman levels. As mentioned earlier, the magnetic moment of a nucleus can be in one of its directional quantum states m. In Earth's weak magnetic field, the energy difference between these states is tiny. However, the powerful magnetic field of the NMR instrument causes the states m to have a greater energy difference (i.e. the magnetic field causes a Zeeman effect), according to:

Equation 5

$$\Delta E = \gamma \hbar B_0$$

Due to the Zeeman effect and the principle of energy minimum, a little excess of the nuclear magnetic moments of an NMR sample prefer the orientation with a lower energy state. In

the NMR experiment, nuclei on the lower energy Zeeman level are transited to the higher energy level using the energy of the radio waves. After the radio wave pulse, the nuclei return back to the relaxed state and the released energy is measured as a complex, decaying interferogram of frequencies. Figure 4A shows this interferogram, called the Free Induction Decay (FID). The greater the energy difference, ΔE, between the Zeeman levels, the greater the strength of the resulting NMR signal. Thus, Equation 5 explains the urge to build more powerful magnets (i.e. stronger B_0).

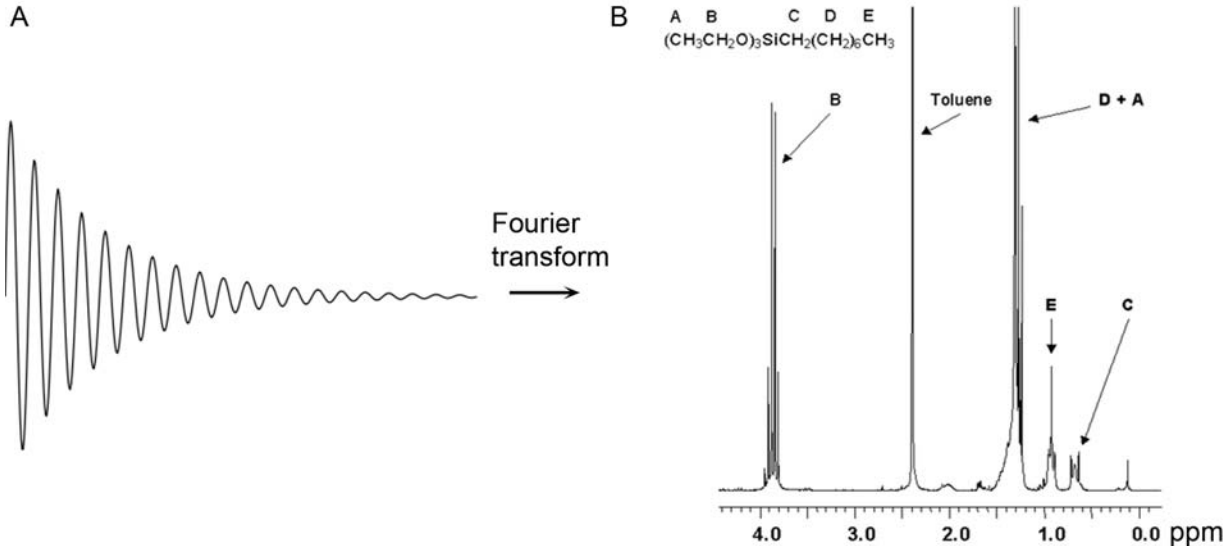


Figure 4. The NMR signal. A) FID, measured over time, is transformed into the frequency domain using Fourier transform. B) The proton NMR spectrum of an organosilane compound. The location of the hydrogen peak is dependent on the chemical environment, i.e. the neighboring atoms. The proximity of oxygen generates a peak left in the spectrum near 4 ppm. Oxygen attracts the electron density of the neighboring atoms, which reduces the shielding (presence of countercharge in the form of electrons) of the hydrogen protons, and this causes the shift. 4A-B figure source: Wikimedia Commons (slightly modified).

The FID is not normally analyzed directly. Figure 4B shows the result of the mathematical processing of the FID using an operation called Fourier transformation. Fourier transformation converts the signal from a time-domain into a frequency domain. The information content of the signal is the same in A and B, but in B, it is more readily readable for the human eye.

In the frequency domain (Fig. 4B), peaks of different frequency arise from nuclei that precess in a slightly (notice the parts per million scale) altered frequency compared to a reference substance (for example tetrasilane). This is due to the electron density of the neighboring atoms affecting the precession frequency of a nucleus. In principle, every molecule with different chemical composition provides an individual spectrum, and by recording the NMR spectrum of a

molecule it is possible to estimate its composition. The ability of NMR, in principle at least, to have one identifiable frequency for each individual chemical environment, makes NMR unique in spectroscopy.

In protein studies, the number of active nuclei is so great that more complicated measurements are employed. In our study (Paper IV), we used two-dimensional NMR spectroscopy. This means that the FID is recorded several hundreds of times with a slightly altered delay period in the NMR radio wave pulse sequence. This introduces an artificial time-dependency in the spectra recorded, which can be exploited to identify the nuclei that exchange magnetic information with each other, i.e. are "coupled." When Fourier transformed along the artificial dimension as well as the FID dimension, the result is a 2D map of the NMR peaks, on one axis the basic 1D NMR measurement and on the other axis the development of these 1D spectra over time. Depending on the particular pulse-sequence used, each amino acid residue yields a specific peak pattern, which helps the assignment of the 2D spectra or produce experimentally determined constraints used for structural calculations.

Different types of NMR spectra are needed in order to solve the structure of a polypeptide. NOESY (nuclear Overhauser effect spectroscopy) NMR spectrum shows NOE peaks that originate from nuclei of amino acid residues that are in proximity to each other. In brief, the NOE peaks arise from transfer of magnetization over space. The magnetization of the nuclei can, thus, be affected not only by neighbors connected by bonds, but also through space. NOE peaks are of particular interest in structural research, since they reveal how the protein is folded. For instance, amino acid stretches forming an α-helix have a typical pattern of NOE peaks between every third residue's proton of the backbone carbon and the backbone nitrogen-bound proton. Summarizing, NMR is an effective method for solving the atomic three-dimensional structure of molecules in solution.

3. Results

3.1 Overview of the results

The main finding of Paper I is the tissue specific cleavage of Vg. The two following invited papers (II-III) are a book chapter where I draft the molecular architecture of honey bee Vg (see Fig. 2) and a commentary article, which discusses the molecular evolutionary impact of *vg* sequencing results from different honey bee populations by Kent et al. (2011). In the fourth paper, we present the cleavage of Vg as a phosphorylation-regulated process taking place at a disordered polyserine linker. Paper V concentrates on describing a novel property of Vg, that is, membrane binding.

3.2 Paper I: Vitellogenin in two pieces

Here, we describe Vg purification protocols for hemolymph and fatbody samples. We report that Vg can get cleaved in two fragments (150 + 40 kDa) in the fatbody tissue. However, Vg is found intact in the hemolymph, suggesting that the cleavage serves tissue-specific purposes. Using mass-spectrometry, we show that the N-terminal domain (N-sheet; 40 kDa) is detached in the cleavage. We speculate that the cleavage might be connected to the putative receptor-binding role of the N-sheet (Li et al., 2003) or to the exposure of the lipophilic cavity present inside the N-sheet. I provide a homology model of the N-sheet; for the first time, lamprey lipovitellin is used for modeling honey bee Vg.

In this article, we also bring forward the characterization of the post-translational modifications of honey bee Vg. We show that the protein carries at least two N-linked glycans: one in the N-sheet and the other in the C-terminal fragment. We show that phosphorylation is present in both cleavage products and that Vg is the dominating phosphorylated protein in the hemolymph.

3.3 Paper II: Vitellogenin in honey bee behavior and lifespan

I got an opportunity to summarize the honey bee Vg knowledge acquired by 2010 in the book "Honey bee Neurobiology and Behavior". In our chapter, we describe Vg as a regulator of bee behavior and lifespan.

My task was to concentrate on the view of the Vg structure and the protein as a molecule. The molecular perspective is important, since much is known about the effects of Vg in the bee, but little is known about the molecular basis. I review what was known, compare honey bee Vg to the lamprey lipovitellin structure and speculate with potential sites for molecular interaction in Vg.

3.4 Paper III: Social pleiotropy and the molecular evolution of honey bee vitellogenin

We were invited to comment on a very recent work by Kent et al. (2011), who sequenced *vg* of several honey bee colonies and found high rates of ongoing selection. Vg mutations were found enriched in parts of the lipid-binding cavity. Citing our Paper I, Kent et al. conclude that the N-sheet is a well-conserved domain not only within species, but within honey bee colonies as well.

Kent et al. (2011) show a molecular model of whole honey bee Vg based on lamprey lipovitellin. I must criticize the modeling work, since Kent et al. (2011) has not taken into account the insect-specific insertion sequences of honey bee Vg such as the polyserine linker. This leads into a biased model, where the polyserine tract has been modeled as a part of the α-helical domain. Kent et al. (2011) uses the flawed model to measure interactions between amino acid residues and lipids at an Ångstrom scale. This mistake highlights the crucial part of molecular modeling: the sequence alignment has to be done using sequences of several species and considering what is structurally feasible. Despite the defects of the Kent et al. (2011) model, we do agree with their important main results: *vg* is a strikingly polymorphic gene subjected to ongoing evolution and the nucleotide polymorphism is concentrated at the major lipid-binding cavity.

3.5 Paper IV: The polyserine linker cleavage site

This submitted manuscript narrows down the cleavage site of Vg to the polyserine linker located at the domain boundary of the N-sheet and the major lipid-binding cavity (Fig. 2A). Based on *in vitro* phosphorylation of the polyserine tract in addition to limited proteolysis experiments and mass-spectrometric analysis, we reason that the Vg cleavage is likely to be regulated by phosphorylation. Our NMR structures (Fig. 3A) suggest that the polyserine linker is an intrinsically disordered segment in honey bee Vg.

The insect polyserine linker is a site of evolutionary smith work (for a review, see Tufail & Takeda 2002). The honey bee has 14 serine residues in the middle of the linker of total 33 residues, whereas the solitary parasitoid *Nasonia vitripennis* jewel wasp has a reduced serine content (seven serine residues at the corresponding sequence). Due to the evolutionary context, we also studied the structure of the tract region of *N. vitripennis* using NMR. We show that the honey bee polyserine tract was found to form an extended, disordered coil, whereas the corresponding *N. vitripennis* peptide had a collapsed ensemble with helical tendency. These differences might arise from the varying serine content. A site this variable in a protein that is strongly linked to social behavior in a Hymenoptera species (the honey bee) is of general interest for evolutionary insect research.

3.6 Paper V: The membrane-affinity of vitellogenin

In this manuscript, we report honey bee Vg membrane association. Our work describes for the first time the binding of a Vg protein family member to membranes of somatic cells. We localize Vg on biological membranes of the bee tissues *in vivo* using immunodetection and show *in vitro* that the protein binds to insect Sf9 cells with a western blot based assay. Moreover, we use membrane-mimicking system, liposomes, to measure the binding using SPR and we estimate the binding region with a limited proteolysis assay. Summarizing, we use a set of methods and test-systems to verify that honey bee Vg is interacting with membranes. These results assign the α-helical domain as an important region with regard the membrane binding.

Piscine Vg is antibacterial (reviewed by Zhang et al., 2011) and also the human homolog apoB can bind to bacteria (Bartolome et al., 2010). Hence, we thought the membrane binding could be part of an antibacterial response (for antibacterial mechanisms, see Takahashi et al., 2010), providing part of the heightened immunity experienced by Vg-rich bees. We tested the

bactericidal and membrane disruption potential of honey bee Vg, but these results were negative. However, honey bee Vg was able to interact with *E. coli* based on western blot results. In fish, Vg not only is directly bactericidal, but also facilitates macrophage-mediated phagocytosis by opsonizing bacteria (Li et al., 2008), and this option could be tested with honey bee Vg and hemocytes in the future. If Vg is not antibacterial in the honey bee, then what purpose does the membrane binding property serve? We speculate with a link between the membrane association of honey bee Vg and the antioxidant nature of the protein in this species.

Conclusions

This thesis research is the first set of structure-oriented studies on the honey bee protein known for its fascinating effects in the bee behavior and aging – Vg. We characterized its posttranslational modifications and created faster purification protocols. We described the Vg cleavage pattern and suggested that the cleavage is likely to be regulated by phosphorylation, taking place at a disordered linker site. Furthermore, we showed that this protein is able to bind to membranes and to *E. coli*. The cleavage of the N-sheet domain, the phosphorylation, the membrane binding by the α-helical domain and the interaction with bacteria constitute the main contribution of this thesis to the molecular understanding of honey bee Vg. In addition, our molecular models of the N-sheet and the α-helical domain, and the NMR structures of the polyserine linker provide, for the first time, a structural perspective on this protein in the bee.

Future perspectives

Initially, we wanted to attempt to crystallize the bee Vg in order to solve the X-ray structure, and spent considerable time and resources pursuing this goal. However, the aggregation and the cleavage tendency created great problems that were not untangled. In the future, it would be important to find an inhibitor for the Vg cleavage. This could increase the chances of crystallizing Vg, and also give indications of the enzyme that cleaves Vg. Expression of Vg in an insect cell system (such as Sf9 cells) could be a profitable approach in the future. I did crystallization trials with the recombinant vWFD (384 buffer conditions) at Norwegian Structural Biology Centre at University of Tromsø, but no crystals were observed. In the future, construct optimization and/or NMR could be further utilized with regard to the vWFD domain, which still is a mystery for science both structure- and function-wise in Vgs; it is not known whether this domain participates hemolymph clotting similarly to the blood clotting of the homologous human protein.

Together with α-helical and vWFD domain, we also purchased a recombinant N-sheet domain. This sample formed a large, soluble, structured (I used CD spectroscopy to test that the domain had a typical β-barrel CD curve) aggregate of approximately 19 monomeric units measured with size-exclusion chromatography and DLS. The soluble aggregate was resistant to monomerization in the presence several additives, including detergent Triton X-100 and dimethyl sulfoxide, and in a pH range of 4-9, so further studies with this construct were abandoned. However, it was used for preparing an antibody (in collaboration with Harlan antibodies, IN, USA). The antibody has not yet been used for research, but I have verified that this antibody recognizes the natural source Vg and the N-terminal fragment, but not the natural C-terminal fragment (the lipid binding cavity). In the future, this antibody can be useful for tracking the N-terminal fragment in combination with the Vg antibody used in these thesis papers. The latter recognizes the whole length Vg and the C-terminal fragment, but not the N-sheet.

Finally, I have the impression that most honey bee research is currently oriented towards behavioral studies, physiology, genetics or neuroscience, i.e. towards other fields than structure oriented protein research (see, for instance, the reviews by Denison and Raymond-Delpech, 2008; Miklos and Maleszka, 2011). Based on the discussions at honey bee meetings and the contacts I have got after the publishing of Paper I, I believe there is much interest in detailed molecular

level research on bee proteins such as Vg, which is fundamentally linked to honey bee behavior and aging. I have endeavored to build a foundation for this field with my thesis research.

References

Amdam, G. V., Hartfelder, K., Norberg, K., Hagen, A. and Omholt, S. W. (2004a). Altered physiology in worker honey bees (Hymenoptera: Apidae) infested with the mite Varroa destructor (Acari: Varroidae): a factor in colony loss during overwintering? *J Econ Entomol* **97**, 741-747.

Amdam, G. V., Norberg, K., Hagen, A. and Omholt, S. W. (2003). Social exploitation of vitellogenin. *Proc Natl Acad Sci U S A* **100**, 1799-1802.

Amdam, G. V. and Omholt, S. W. (2003). The hive bee to forager transition in honeybee colonies: the double repressor hypothesis. *Journal of Theoretical Biology* **223**, 451-464.

Amdam, G. V., Simoes, Z. L., Hagen, A., Norberg, K., Schroder, K., Mikkelsen, O., Kirkwood, T. B. and Omholt, S. W. (2004b). Hormonal control of the yolk precursor vitellogenin regulates immune function and longevity in honeybees. *Exp Gerontol* **39**, 767-773.

Amdam, G. V., Aase, A. L., Seehuus, S. C., Kim Fondrk, M., Norberg, K. and Hartfelder, K. (2005a). Social reversal of immunosenescence in honey bee workers. *Exp Gerontol* **40**, 939-947.

Amdam, G. V., Aase, A. L. T. O., Seehuus, S. C., Fondrk, M. K., Norberg, K. and Hartfelder, K. (2005b). Social reversal of immunosenescence in honey bee workers. *Experimental Gerontology* **40**, 939-947.

Ament, S. A., Wang, Y. and Robinson, G. E. (2010). Nutritional regulation of division of labor in honey bees: toward a systems biology perspective. *Wiley Interdiscip Rev Syst Biol Med* **2**, 566-576.

Anderson, T. A., Levitt, D. G. and Banaszak, L. J. (1998). The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. *Structure with Folding & Design* **6**, 895-909.

Avarre, J. C., Lubzens, E. and Babin, P. J. (2007). Apolipocrustacein, formerly vitellogenin, is the major egg yolk precursor protein in decapod crustaceans and is homologous to insect apolipophorin II/I and vertebrate apolipoprotein B. *BMC Evol Biol* **7**, 3.

Babin, P. J. (2008). Conservation of a vitellogenin gene cluster in oviparous vertebrates and identification of its traces in the platypus genome. *Gene* **413**, 76-82.

Baker, M. E. (1988a). Invertebrate vitellogenin is homologous to human von Willebrand factor. *Biochem J* **256**, 1059-1061.

Baker, M. E. (1988b). Is vitellogenin an ancestor of apolipoprotein B-100 of human low-density lipoprotein and human lipoprotein lipase? *Biochem J* **255**, 1057-1060.

Bartolome, N., Aspichueta, P., Martinez, M. J., Vazquez-Chantada, M., Martinez-Chantar, M. L., Ochoa, B. and Chico, Y. (2010). Biphasic adaptative responses in VLDL metabolism and lipoprotein homeostasis during Gram-negative endotoxemia. *Innate Immun*. **26**, 1-11

Behrends, A., Scheiner, R., Baker, N. and Amdam, G. V. (2007). Cognitive aging is linked to social role in honey bees (Apis mellifera). *Exp Gerontol* **42**, 1146-1153.

Bitondi, M. M. G. and Simoes, Z. L. P. (1996). The relationship between level of pollen in the diet, vitellogenin and juvenile hormone titres in Africanized Apis mellifera workers. *Journal of Apicultural Research* **35**, 27-36.

Case D.A., D. T. A., Cheatham T.E., III, Simmerling C.L., Wang J., Duke R.E., Luo R., Crowley M., Walker Ross C., Zhang W., Merz K.M.,Wang B., Hayik S., Roitberg A., Seabra G., Kolossváry I.,Wong K.F., Paesani F., Vanicek J.,Wu X., Brozell S.R., Steinbrecher T., Gohlke H., Yang L., Tan C., Mongan J., Hornak V., Cui G., Mathews D.H., Seetin M.G., Sagui C., Babin V., Kollman P.A. (2008). AMBER 10. *University of California, San Francisco*.

Crailsheim, K. (1991). Interadult feeding of jelly in honeybee (*Apis mellifera*I colonies. *Journal of Comparative Physiology* **B**, 55-60.

Davis, C. F. (2006). The Honey Bee Inside Out. *Bee Craft Limited*.

Denison, R. and Raymond-Delpech, V. (2008). Insights into the molecular basis of social behaviour from studies on the honeybee, Apis mellifera. *Invert Neurosci* **8**, 1-9.

Dukas, R. (2008). Mortality rates of honey bees in the wild. *Insectes Sociaux* **55**, 252-255.

Finn, R. N. (2007). Vertebrate yolk complexes and the functional implications of phosvitins and other subdomains in vitellogenins. *Biol Reprod* **76**, 926-935.

Friebolin, H. (1993). Basic One- and Two-Dimensional NMR Spectroscopy. *VCH Publishers, NY, USA*.

Guidugli-Lazzarini, K. R., do Nascimento, A. M., Tanaka, E. D., Piulachs, M. D., Hartfelder, K., Bitondi, M. G. and Simoes, Z. L. (2008). Expression analysis of putative vitellogenin and lipophorin receptors in honey bee (Apis mellifera L.) queens and workers. *J Insect Physiol* **54**, 1138-1147.

Guidugli, K. R., Nascimento, A. M., Amdam, G. V., Barchuk, A. R., Omholt, S., Simoes, Z. L. and Hartfelder, K. (2005). Vitellogenin regulates hormonal dynamics in the worker caste of a eusocial insect. *FEBS Lett* **579**, 4961-4965.

Havukainen, H., Halskau, Ø., Skjaerven, L., Smedal, B. and Amdam, G. V. (2011). Deconstructing honeybee vitellogenin: novel 40 kDa fragment assigned to its N-terminus. *Journal of Experimental Biology* **15**, 528-592.

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**, 195-202.

Lehtonen, J. V., Still, D. J., Rantanen, V. V., Ekholm, J., Bjorklund, D., Iftikhar, Z., Huhtala, M., Repo, S., Jussila, A., Jaakkola, J. et al. (2004). BODIL: a molecular modeling environment for structure-function analysis and drug design. *J Comput Aided Mol Des* **18**, 401-419.

Levitt, M. H. (2009). Spin Dynamics - Basics of Nuclear Magnetic Resonance. *Wiley*.

Li, A., Sadasivam, M. and Ding, J. L. (2003). Receptor-ligand interaction between vitellogenin receptor (VtgR) and vitellogenin (Vtg), implications on low density lipoprotein receptor and apolipoprotein B/E. The first three ligand-binding repeats of VtgR interact with the amino-terminal region of Vtg. *J Biol Chem* **278**, 2799-2806.

Li, Z., Zhang, S. and Liu, Q. (2008). Vitellogenin functions as a multivalent pattern recognition receptor with an opsonic activity. *PLoS One* **3**, e1940.

Mann, C. J., Anderson, T. A., Read, J., Chester, S. A., Harrison, G. B., Kochl, S., Ritchie, P. J., Bradbury, P., Hussain, F. S., Amey, J. et al. (1999). The structure of vitellogenin provides a molecular model for the assembly and secretion of atherogenic lipoproteins. *J Mol Biol* **285**, 391-408.

Miklos, G. L. G. and Maleszka, R. (2011). Epigenomic communication systems in humans and honey bees: From molecules to behavior. *Hormones and Behavior* **59**, 399-406.

Munch, D. and Amdam, G. V. (2010). The curious case of aging plasticity in honey bees. *FEBS Lett* **584**, 2496-2503.

Munch, D., Amdam, G. V. and Wolschin, F. (2008). Ageing in a eusocial insect: molecular and physiological characteristics of life span plasticity in the honey bee. *Functional Ecology* **22**, 407-421.

Nelson, C. M., Ihle, K. E., Fondrk, M. K., Page, R. E. and Amdam, G. V. (2007). The gene vitellogenin has multiple coordinating effects on social organization. *PLoS Biol* **5**, e62.

Page, R. E., Jr., Scheiner, R., Erber, J. and Amdam, G. V. (2006). The development and evolution of division of labor and foraging specialization in a social insect (Apis mellifera L.). *Curr Top Dev Biol* **74**, 253-286.

Pinto, L. Z., Bitondi, M. M. G. and Simoes, Z. L. P. (2000). Inhibition of vitellogenin synthesis in Apis mellifera workers by a juvenile hormone analogue, pyriproxyfen. *Journal of Insect Physiology* **46**, 153-160.

Piulachs, M. D., Guidugli, K. R., Barchuk, A. R., Cruz, J., Simoes, Z. L. and Belles, X. (2003). The vitellogenin of the honey bee, Apis mellifera: structural analysis of the cDNA and expression studies. *Insect Biochem Mol Biol* **33**, 459-465.

Richards, M. P. (1997). Trace mineral metabolism in the avian embryo. *Poult Sci* **76**, 152-164.

Robinson, G. E., Page, R. E., Strambi, C. and Strambi, A. (1992). Colony Integration in Honey-Bees - Mechanisms of Behavioral Reversion. *Ethology* **90**, 336-348.

Roma, G. C., Bueno, O. C. and Camargo-Mathias, M. I. (2010). Morpho-physiological analysis of the insect fat body: a review. *Micron* **41**, 395-401.

Rule, S. G. and Hitchens, T. K. (2006). Fundamentals of Protein NMR Spectroscopy. *Springer*.

Sappington, T. W. and Raikhel, A. S. (1998). Molecular characteristics of insect vitellogenins and vitellogenin receptors. *Insect Biochem Mol Biol* **28**, 277-300.

Seehuus, S. C., Amdam, G. V., Norberg, K., Gimsa, U. and Krekling, T. (2006a). Reproductive protein protects functionally sterile honey bee workers from oxidative stress. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 962-967.

Seehuus, S. C., Krekling, T. and Amdam, G. V. (2006b). Cellular senescence in honey bee brain is largely independent of chronological age. *Exp Gerontol* **41**, 1117-1125.

Seehuus, S. C., Norberg, K., Krekling, T., Fondrk, K. and Amdam, G. V. (2007). Immunogold localization of vitellogenin in the ovaries, hypopharyngeal glands and head fat bodies of honeybee workers, Apis mellifera. *J Insect Sci* **7**, 1-14.

Seeley, T. D. (2010). Honeybee Democracy. *Princeton University Press*.

Sharp, T. R., Morris, R., Horan, G. J., Pezzullo, L. H. and Stroh, J. G. (2006). Method for determining the average degree of substitution of o-vanillin derivatized porcine somatotropin. *J Pharm Biomed Anal* **40**, 185-189.

Smedal, B., Brynem, M., Kreibich, C. D. and Amdam, G. V. (2009). Brood pheromone suppresses physiology of extreme longevity in honeybees (Apis mellifera). *J Exp Biol* **212**, 3795-3801.

Stroh, J. G., Loulakis, P., Lanzetti, A. J. and Xie, J. (2005). LC-mass spectrometry analysis of N- and C-terminal boundary sequences of polypeptide fragments by limited proteolysis. *J Am Soc Mass Spectrom* **16**, 38-45.

Takahashi, D., Shukla, S. K., Prakash, O. and Zhang, G. (2010). Structural determinants of host defense peptides for antimicrobial activity and target cell selectivity. *Biochimie* **92**, 1236-1241.

Tautz, J. (2008). The Buzz about Bees - Biology of a Superorganism. *Springer*.

Trenczek, T. and Engels, W. (1986). Occurrence of Vitellogenin in Drone Honeybees (Apis-Mellifica). *International Journal of Invertebrate Reproduction and Development* **10**, 307-311.

Tufail, M. and Takeda, M. (2008). Molecular characteristics of insect vitellogenins. *J Insect Physiol* **54**, 1447-1458.

Weinstock, G. M., Robinson, G. E., Gibbs, R. A., Worley, K. C. and Evans, J. D. (2006). Insights into social insects from the genome of the honeybee Apis mellifera. *Nature* **443**, 931-949.

Wheeler, D. E. and Kawooya, J. K. (1990). Purification and characterization of honey bee vitellogenin. *Arch Insect Biochem Physiol* **14**, 253-267.

Wilson, B. (2004). The Hive - The Story of the Honeybee and Us. *John Murray (Publishers)*.

Zhang, S., Wang, S., Li, H. and Li, L. (2011). Vitellogenin, a multivalent sensor and an antimicrobial effector. *Int J Biochem Cell Biol* **43**, 303-305.

Paper I

Deconstructing honeybee vitellogenin: novel 40 kDa fragment assigned to its N terminus

Havukainen, H., Halskau, Ø., Skjaerven, L., Smedal, B., Amdam, G. V.
Journal of Experimental Biology **214**: 582-592 (2011)

Abstract

Vitellogenin, an egg-yolk protein precursor common to oviparous animals, is found abundantly in honeybee workers – a caste of helpers that do not usually lay eggs. Instead, honeybee vitellogenin (180 kDa) participates in processes other than reproduction: it influences hormone signaling, food-related behavior, immunity, stress resistance and longevity. The molecular basis of these functions is largely unknown. Here, we establish and compare molecular properties of vitellogenin from honeybee hemolymph (blood) and abdominal fat body, two compartments that are linked to vitellogenin functions. Our results reveal a novel 40 kDa vitellogenin fragment in abdominal fat body tissue, the main site for vitellogenin synthesis and storage. Using MALDI-TOF combined with MS/MS mass-spectroscopy, we assign the 40 kDa fragment to the N-terminus of vitellogenin, whereas a previously observed 150 kDa fragment corresponded to the remainder of the protein. We show that both protein units are N-glycosylated and phosphorylated. Focusing on the novel 40 kDa fragment, we present a homology model based on the structure of lamprey lipovitellin that includes a conserved β-barrel-like shape, with a lipophilic cavity in the interior and two insect-specific loops that have not been described before. Our data indicate that the honeybee fat body vitellogenin experiences cleavage unlike hemolymph vitellogenin, a pattern that can suggest a tissue-specific role. Our experiments advance the molecular understanding of vitellogenin, of which the multiple physiological and behavioral effects in honeybees are well established.

The Journal of Experimental Biology 214, 582-592

doi:10.1242/jeb.048314

RESEARCH ARTICLE

Deconstructing honeybee vitellogenin: novel 40 kDa fragment assigned to its N terminus

Heli Havukainen[1,2,*], Øyvind Halskau[1,2], Lars Skjaerven[2], Bente Smedal[1] and Gro V. Amdam[1,3]

[1]Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, P.O. Box 5003 1432 Aas, Norway, [2]Department of Biomedicine, University of Bergen, Jonas Lies vei 91, 5009 Bergen, Norway and [3]School of Life Sciences, Arizona State University, P.O. Box 874501, Tempe, AZ 85287-4701, USA

*Author for correspondence (heli.havukainen@umb.no)

Accepted 27 October 2010

SUMMARY

Vitellogenin, an egg-yolk protein precursor common to oviparous animals, is found abundantly in honeybee workers – a caste of helpers that do not usually lay eggs. Instead, honeybee vitellogenin (180 kDa) participates in processes other than reproduction: it influences hormone signaling, food-related behavior, immunity, stress resistance and longevity. The molecular basis of these functions is largely unknown. Here, we establish and compare the molecular properties of vitellogenin from honeybee hemolymph (blood) and abdominal fat body, two compartments that are linked to vitellogenin functions. Our results reveal a novel 40 kDa vitellogenin fragment in abdominal fat body tissue, the main site for vitellogenin synthesis and storage. Using MALDI-TOF combined with MS/MS mass-spectroscopy, we assign the 40 kDa fragment to the N terminus of vitellogenin, whereas a previously observed 150 kDa fragment corresponded to the remainder of the protein. We show that both protein units are N glycosylated and phosphorylated. Focusing on the novel 40 kDa fragment, we present a homology model based on the structure of lamprey lipovitellin that includes a conserved β-barrel-like shape, with a lipophilic cavity in the interior and two insect-specific loops that have not been described before. Our data indicate that the honeybee fat body vitellogenin experiences cleavage unlike hemolymph vitellogenin, a pattern that can suggest a tissue-specific role. Our experiments advance the molecular understanding of vitellogenin, of which the multiple physiological and behavioral effects in honeybees are well established.

Supplementary material available online at http://jeb.biologists.org/cgi/content/full/214/4/582/DC1

Key words: vitellogenin, hemolymph, fat body, functional protein fragments, homology modeling, glycosylation, phosphorylation, honeybee.

INTRODUCTION

Vitellogenins are female-specific egg-yolk precursors (Spieth et al., 1991), relatives of the vertebrate lipoproteins (Baker, 1988; Mann et al., 1999). These precursors are synthesized by most oviparous animals for transfer into oocytes, where they serve as nourishment for embryos. In honeybees (*Apis mellifera*), however, vitellogenin is also expressed by workers, a caste of essentially sterile female helper bees (Engels and Fahrenhorst, 1974). Vitellogenin circulates in hemolymph (blood) and is stored at its site of production, the fat body – an insect organ functionally homologous to the vertebrate liver and white fat (Arrese and Soulages, 2009).

During the life course of worker honeybees, vitellogenin levels in the hemolymph and fat body drop, and these reduced concentrations influence several aspects of the life history of the bee (Munch and Amdam, 2010). RNA interference-mediated knockdown of the *vitellogenin* gene shows that, as vitellogenin protein levels decline, titers of the life-shortening juvenile hormone (JH) increase, and workers show immune senescence, susceptibility to oxidative stress and reduced survival, in addition to a higher probability of abandoning nest-tasks in favor of foraging for nectar (rather than pollen) from flowering plants (Amdam et al., 2003b; Amdam et al., 2004b; Guidugli et al., 2005; Nelson et al., 2007; Seehuus et al., 2006).

The specific, pleiotropic effects of vitellogenin on honeybee physiology, longevity and food-related behavior suggest that this protein can suppress insulin/insulin-like signaling (IIS) in workers (Corona et al., 2007; Hunt et al., 2007; Seehuus et al., 2006). IIS is eukaryotic pathway for nutrient sensing that integrates responses in metabolism, growth, feeding, reproduction, immunity, stress tolerance and survival (Kenyon, 2010). This possibility has implications beyond honeybee biology, because of the conserved nature of IIS; IIS is also a focus of human biomedical research (Munch and Amdam, 2010). However, it is largely unknown how the honeybee vitellogenin molecule exerts its many functions.

In terms of biochemistry, honeybee vitellogenin is grossly described as a 180 kDa monomeric phospholipoglycoprotein (Wheeler and Kawooya, 1990). All insect vitellogenins, excluding those of the honeybee suborder Apocrita, are cleaved in vivo, typically close to the polyserine track(s) at an RXXR consensus-sequence motif by subtilisin-like endoproteases (Barr, 1991; Rouille et al., 1995). The proteins diverge in their specific location of the polyserine tracts and cleavage sites but, overall, cleavage results in one smaller N-terminal fragment and a larger C-terminal fragment (Tufail and Takeda, 2008). Honeybee vitellogenin lacks the RXXR consensus sequence in the vicinity of the polyserine region, as well as the typical presence of cleavage products in hemolymph. The protein has, nonetheless, been observed in two forms: the mature 180 kDa protein in the hemolymph and fat body, and a lighter, 150 kDa fragment in the ovaries of queens (Seehuus et al., 2007)

and in the hypopharyngeal head glands of workers, where vitellogenin constituents are processed into food secretions for larval feeding (Amdam et al., 2003a). The 150 kDa unit also appeared during purification of vitellogenin from queen hemolymph (Wheeler and Kawooya, 1990) and, therefore, was perceived as a degradation product (Wheeler and Kawooya, 1990).

Residue range 351–381 of honeybee vitellogenin comprises 13 serine residues, and comparative work from several taxa (Don-Wheeler and Engelmann, 1997; Tufail and Takeda, 2002) point to phosphorylation patterns in this region that can be important in protein-receptor interactions (Miller et al., 1982; Raikhel and Dhadialla, 1992). Glycosylation of vitellogenin has been connected to its secretion from the fat body of a cockroach (*Leucophaea maderae*) (Don-Wheeler and Engelmann, 1997), but this mechanism is not verified in bees. Overall, the placement and role of post-translational modifications on honeybee vitellogenin are unexplored topics in research.

Here, we first improve and develop purification protocols that allow us to compare vitellogenin derived from the hemolymph with that derived from the fat body of honeybee workers. We report a novel 40 kDa fragment that is abundant in sample purified from the fat body compartment. We assign the 40 kDa fragment to the vitellogenin N terminus using mass spectroscopy, and establish that the corresponding polypeptide is missing from the previously reported 150 kDa vitellogenin unit. Our interpretation is that the honeybee vitellogenin molecule is cleaved into a small N-terminal fragment (40 kDa) and a larger C-terminal fragment (150 kDa) and, thus, the reported 150 kDa unit does not correspond to a degradation product in fat body. We test the overall phosphorylation and glycosylation status of the 40 kDa and 150 kDa fragments, as well as that of the full-length 180 kDa protein. Finally, based on the known structure of lamprey (*Ichthyomyzon unicuspis*) lipovitellin (Raag et al., 1988; Thompson and Banaszak, 2002), we prepare a homology model for the 40 kDa unit, identifying a large positively charged groove that may be involved in ligand binding, as well as two insect-specific loops and a cavity with the potential to bind fatty acids. The functional significance of these structures is discussed. Our results represent the first major step forward in understanding the functional structure and composition of honeybee vitellogenin since the publication of its coding sequence eight years ago (Piulachs et al., 2003).

MATERIALS AND METHODS

Bees

To establish reliable sources of worker honeybees with ample levels of vitellogenin, four colonies were induced to develop *diutinus* bees (winter bees). This worker subcaste is naturally occurring in Northern Europe, and is characterized by a considerable accumulation of vitellogenin in the hemolymph and fat body (Fluri et al., 1977; Smedal et al., 2009). In brief, this development was achieved by caging the queen in each colony, which effectively blocks brood rearing. The technique was established and used before (Amdam et al., 2004a; Amdam et al., 2005; Maurizio, 1950).

Sample collection

Mature workers (minimum 2 weeks old) were collected from the nest into small cages for transfer to the laboratory. Bees were anesthetized in a refrigerator (4°C) until immobile. Next, hemolymph was collected through a small incision through the abdominal wall, as described previously (Amdam et al., 2004a). Hemolymph (10 µl) was dissolved in 100 µl ice-cold PBS (phosphate-buffered saline tablet in 200 ml ddH_2O; Sigma-Aldrich, St Louis, MO, USA) containing 1 mmol l^{-1} disodium EDTA and protease inhibitor cocktail (Mini tablets; Roche, Indianapolis, IN, USA). To obtain abdominal fat body tissue, the last body segment, including the stinger apparatus and adhering rectum, gut and ovaries, was removed from the anesthetized bees by gripping the segment with tweezers. The remaining abdominal carcass (the body wall lined with fat body) was disconnected from the thoracic body-segment, placed in a cryotube and immediately frozen in liquid N_2. All samples were stored at –80°C until use.

Immunofluorescence

Bee abdomens, dissected on ice, were fixed in formaldehyde and embedded in London Resin White (Electron Microscopy Science, Hatfield, USA), as described before (Seehuus et al., 2007), to provide resin blocks with embedded abdominal tissue ready for sectioning. Sections (1–2 µm) of resin-embedded material were cut with a diamond knife using a Reichert Jung ultra-microtome (Leica, Wetzlar, Germany). Fat body sections were dried onto SuperFrost/Plus slides (Thermo Scientific, Walldorf, Germany), rinsed and washed with PBS-T (PBS with 0.02% Triton X-100) and blocked with 2% BSA (Sigma-Aldrich) in PBS-T for 60 min at room temperature. After 3×15 min washes with PBS-T, sections were incubated overnight at 4°C with a polyclonal (rabbit) anti-vitellogenin antibody at 1:500 (raised against 180 kDa honeybee vitellogenin; Pacific Immunology, Ramona, CA, USA); specificity had been tested and confirmed previously (Seehuus et al., 2007). The negative control was incubated with PBS-T, as verified before (Seehuus et al., 2007). After washing in PBS-T, all samples were incubated with a polyclonal anti-rabbit antibody coupled to the fluorochrome Cy3 (AffiniPure Goat Anti-Rabbit IgG, Jackson ImmunoResearch Europe Ltd, Suffolk, UK) diluted 1:200, overnight at 4°C. Thus, for negative control staining, sections were incubated with secondary antibody only (Seehuus et al., 2007). Finally, samples were washed in PBS-T and subsequently mounted in 50% glycerol/PBS. Images were acquired with a confocal laser scanning microscope (TCS SP5, Leica).

Protein purification

Hemolymph

After quick thawing of the samples, debris was removed by centrifuging at 20,800 ***g*** at 4°C for 20 min, and thereafter filtering the supernatant with a syringe filter (Supor Membrane pore size 0.2 µm; Pall Life Sciences, Port Washington, NY, USA). Ion-exchange chromatography was performed as described previously for queen hemolymph using 50 mmol l^{-1} $NaPO_4$ buffer and a 0–0.45 mol l^{-1} NaCl gradient (Wheeler and Kawooya, 1990) in order to remove apolipophorins and hexamerins, which are the major hemolymph proteins in insects. Sample collection was restricted to those fractions containing no apolipophorins or hexamerins; this procedure makes it possible to exclude the density ultracentrifugation step of the purification protocol described by Wheeler and Kawooya (Wheeler and Kawooya, 1990). Purification was completed after the Concanavalin A (ConA; Sigma-Aldrich) affinity chromatography step, as described before (Wheeler and Kawooya, 1990); the buffer was changed to ConA buffer (Wheeler and Kawooya, 1990) containing EDTA-free inhibitor cocktail (Mini pills; Roche) and a PD-10 desalting column (Amersham Biosciences, Piscataway, NJ, USA) was used. The purification steps were monitored by running samples on 4–20% gradient SDS-PAGE gels (Bio-Rad Laboratories, Hercules, CA, USA) with a Precision Plus Protein Standard molecular weight marker (Bio-Rad). The gels were stained with Coomassie Blue (Bio-Rad).

Fat body

Abdominal carcasses stored at –80°C were ground directly under liquid N_2 without thawing and, thereafter, were transferred quickly to ice-cold 20 mmol l^{-1} Hepes buffer (pH 7) containing 0.2 M NaCl and protease inhibitor cocktail (Mini pills; Roche). Homogenates were prepared such that 1.5 ml buffer was added to materials from 7–10 abdominal carcasses. Next, samples were mixed by vortexing and sonicated (10 watts, 5 s, 10 times) with a Vibra Cell ultrasonic processor (Sonics & Materials Inc., Danbury, CT, USA) using a 2 mm probe on ice. Sonicated samples were centrifuged at 20,800 ***g***, 4°C, 20 min, and the supernatant filtered as described above. Thereafter, samples were frozen (–80°C), thawed, and centrifuged again, before 1 ml of sample with a total protein concentration of approximately 5 mg ml^{-1} as defined by Bio-Rad protein assay (Bradford, 1976) was applied to a size-exclusion Superdex 200 10/300 GL column (GE Healthcare, Piscataway, NJ, USA). A total of 33 ml Hepes buffer (same as the homogenization buffer above) were used as a solvent. The size-exclusion column separates proteins by their size. The fractions containing largely vitellogenin were collected and the buffer changed to ion-exchange buffer in a PD-10 desalting column (Amersham Biosciences). Ion-exchange and thereafter ConA affinity purification was performed as with the hemolymph samples, and the purification steps were similarly monitored by SDS-PAGE gels.

To test whether the mechanical processing of fat bodies, *per se*, could influence our results, hemolymph sample (non-fragmented vitellogenin) was vortexed and sonicated like the fat body extract as a control.

To estimate the relative amount of vitellogenin fragments in fat body raw extract, five independent observers without prior expectations of band intensities were asked to assess 150 and 40 kDa vitellogenin bands relative to standard protein bands that covered a similar range of intensities. The observers were presented with four distinct 4–20% gradient gels (Bio-Rad) with a total of six gel lanes of fat body extract run together with a dilution series of a standard with proteins of known concentrations (Precision Plus Protein Standard, Bio-Rad). Each gel contained a separately prepared fat body extract to avoid pseudo-replication; i.e. each extraction used different *diutinus* abdomens (16 bees were used in total). From this material, the observers estimated the ratio of 40/150 kDa vitellogenin for every gel independently from each other.

Vitellogenin identification by western blot

Aliquots from each purification step were run on a 4–20% gradient SDS-PAGE gel (Bio-Rad). Separated proteins were transferred to nitrocellulose membrane and incubated with polyclonal antibody against vitellogenin (1:100,000; Pacific Immunology) as described before (Amdam et al., 2003a). Immun-Star horseradish peroxidase-conjugated secondary antibody kit (Bio-Rad) and Gel doc imager (Bio-Rad) were used for visualization.

Vitellogenin identification by mass spectroscopy

Proteolytic preparation of the samples

Purified protein from fat body was run on a SDS gel and stained as outlined above. The bands corresponding to 180, 150 and 40 kDa vitellogenin and sections of gel between 180 and 150 kDa and between 150 and 40 kDa that did not contain visible bands of separated proteins (i.e. 'blank' controls) were excised and prepared for digestion following existing protocols (Nasrin et al., 2009). Porcine trypsin (Promega, Madison, WI, USA), 180 ng per gel band in 10 mmol l^{-1} ammoniumbicarbonate (pH 8) was added to digest the samples overnight at 37°C. Digestion was stopped with 1% trifluoroacetic acid. Samples of tryptic peptides were purified on nC18 StageTip microcolumns, as described by Rappsilber et al. (Rappsilber et al., 2003). For tandem matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF/TOF) mass spectrometry, samples were eluted using 1 µl of matrix [6 mg ml^{-1} α-cyano-4-hydroxycinnamic acid in 60% acetonitrile (ACN)], 15% methanol, 0.1% trifluoroacetic acid (TFA), onto a ground steel MALDI target plate. For liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis, peptides were eluted twice with 10 µl each of 80% ACN, 0.1% formic acid. ACN was removed by vacuum concentrating.

MALDI-TOF/TOF

Experiments were conducted on an Ultra-flex MALDI-TOF/TOF mass spectrometer (Bruker Daltonics, Bremen, Germany) at an acceleration voltage of 20 kV. A peptide calibration standard kit II (Bruker Daltonics) covering a mass range from 757–3147 Da was used for calibration of the mass scale. The spectra were processed with FlexAnalysis (Bruker Daltonics) and the annotated peaks were targeted for peptide mass fingerprint analyses with Mascot (Matrix Sciences, London, UK; in-house Mascot server) using the NCBI database (*Apis mellifera*). The seven most-intense peaks of the 40 kDa vitellogenin sample were also analyzed in the TOF/TOF mode for fragmentation analysis. These were: 70–77, VWNGQYAR; 152–168, VNSVQVPTDDEPYASFK; 218–225, NFDNCDQR; 226–237, INYHFGMTDNSR; 265–276, HFTIQSSVTTSK; 287–294, QNGLVLSR; and 524–533, SMEAANIMSK.

Vitellogenin peptide identification by LC-MS/MS

Peptides (see preparation of samples above) were separated by an Ultimate 3000 nano-HPLC (Dionex Corporation, Sunnyvale, CA, USA), using a nC18 reverse-phase column coupled to a 4000 Q-Trap mass spectrometer (Applied Biosystems/MDS SCIEX, Concord, ON, Canada). The nC18 enrichment column was a C18 Pepmap 100, 5 µm particle diameter, 100 Å pore size, 5000×300 µm i.d. column (Dionex) and the nC18 analytical column was a 15,000×75 µm i.d. fused silica capillary column packed in-house with ReproSil-Pur 5 µm C18 resin (Dr Maisch Gmbh, Ammerbuch-Entringen, Germany). The peptides were eluted from the column at a constant flow rate of 300 nl min^{-1} using a mobile phase gradient [solvent A, aqueous 2% in 0.1% fluoroacetic acid (FA); solvent B, aqueous 90% ACN in 0.1% FA] as follows: solvent B was increased rapidly from 5–12% (0–2 min), followed by a slow increase from 12–30% (2–30 min) and from 30–50% (30–43 min). Elution of very hydrophobic peptides and conditioning of the column were performed during a 10 min isocratic elution with 95% solvent B and a 16 min isocratic elution with 5% B, respectively. The Q-Trap was operated in 'Information Dependent Acquisition' mode using automatic MS to MS/MS switching to acquire data. The instrument settings for ion-spray voltage was 2800 V and the collision energy was automatically adjusted based upon the ion charge state and mass of the compound. Data were processed using Analyst (AB SCIEX, Concord, ON, Canada) and analyzed with Mascot (Matrix Sciences).

Phosphostaining

Centrifuged hemolymph and fat body extracts, as well as vitellogenin from fat body (purified by size-exclusion, ion-exchange and ConA chromatography) were run along with PeppermintStick phosphoprotein standard/control (Invitrogen, Carlsbad, CA, USA) on a 4–20% gradient SDS-PAGE gel (Bio-Rad). The gel was treated according to the manufacturer's instructions, i.e. with the Pro-Q Diamond (Invitrogen) phosphostaining protocol for specific

detection of phosphorylated proteins. The gel was visualized in trans-UV light with a Gel Doc imager (Bio-Rad).

Glycosylation

Fat body protein samples purified with size-exclusion and ion-exchange chromatography were deglycosylated in order to test the presence of glyco-groups in the 180, 150 and 40 kDa vitellogenin fractions. PNGaseF cleaves N-linked oligosaccharides from glycoproteins (Maley et al., 1989; Plummer and Tarentino, 1991). A control sample without the deglycosylation enzyme PNGaseF amidase (New England BioLabs, Inc., Beverly, MA, USA) was treated like the deglycosylated sample, as suggested by the manufacturer. The incubated (37°C, 1 h) samples were subjected to electrophoretic separation and vitellogenin bands were monitored for size shifts (conferring the presence *vs* absence of glycans) after Coomassie staining. For optimal protein separation, 4–20% SDS-PAGE gels were used for the 40 kDa band target, whereas 5% SDS-PAGE gels were used for the 180 and 150 kDa band targets.

Sequence alignment and modelling

The modeling software Bodil (Lehtonen et al., 2004) was used to align honeybee vitellogenin residues 22–323 with lamprey lipovitellin residues 20–281. N-terminal vitellogenin sequences from the following species were also aligned to guide the development of a consensus alignment: *Bombus ignitus* (Swiss-prot accession number B9VUV6), *Pimpla nipponica* (O17428), *Pteromalus puparum* (B2BD67), *Leucophaea maderae* (Q5TLA5), *Acipenser transmontanus* (Q90243) and *Xenopus laevis* (P18709). Also, a previously published multiple alignment of 20 vitellogenins (including honeybee vitellogenin) (Avarre et al., 2007) was used as a reference. The homology model was built with program HOMODGE in Bodil. Two putative loop sequences could not be properly modeled because lamprey does not contain these sequences; these residues were aligned with corresponding stretches of vitellogenin from 19 other insect species and nine vertebrate species using ClustalW (Thompson et al., 1994) and was manually adjusted to detect sequence conservation. An all-atom model was prepared for minimization and molecular dynamics (MD) simulation with Amber10 (Case et al., 2008) and the corresponding Amber03 (Duan et al., 2003; Lee and Duan, 2004) force field. The protein was solvated in a periodic truncated octahedron box with TIP3 water molecules (Jorgensen et al., 1983), providing 12 Å of water between the protein surface and the periodic box edge. The solute was minimized for 10,000 steps, followed by 10,000 steps of the system yielding an energetically refined homology model. The system was heated to 300 K in 300 ps, with restraints, followed by 2 ns of equilibration with constant pressure and temperature (NPT) in order to ensure correct water density. The production phase lasted for 10 ns and was performed with constant volume and energy (NVE) with a 1 fs time step, using SHAKE on hydrogen-heavy atom bonds. Analysis of the MD trajectories was performed with the Bio3d package in order to observe the relative stability of the secondary structure elements (Grant et al., 2006). The rotamer, geometry and Ramachandran plot location of single amino acid residues were adjusted with the validation tools of Coot (Emsley and Cowtan, 2004). An electrostatic map based on residue charges was created with PyMOL (Version 1.2r3pre, Schrödinger, LLC).

RESULTS

Vitellogenin immunolocalization, purification and western verification

Immunofluorescence staining confirmed that the hemolymph and fat body contained vitellogenin (Fig. 1A–C). Of the diverse tissues of the abdominal carcass (fat body, muscle, connective tissue, epithelium), vitellogenin immunoreactivity was exclusively detected in fat body trophocyte cells (Fig. 1A,B), which previously were identified as the sole source of the *vitellogenin* gene transcript in the abdominal carcass of honeybees (Corona et al., 2007).

From hemolymph, vitellogenin was purified in two steps: ion-exchange and ConA chromatography (Fig. 1D). From the abdominal fat body, successful purification required an additional size-exclusion chromatography step preceding the ion-exchange and ConA steps (Fig. 1E). The presence of full-length vitellogenin at 180 kDa in the hemolymph and abdominal fat body, as well as the previously detected 150 kDa fragment in the fat body, was verified by western blots with a vitellogenin-specific antibody (Fig. 1F,G). Partial fragmentation of the purified whole-length vitellogenin into 150 kDa was visible on the gels and the western blots (Fig. 1D,F).

A less intensely stained band of ~40 kDa was observed in purified vitellogenin fractions from fat body (Fig. 1E). A band of the same size was observed to be abundantly present in the unprocessed fat body protein extract and throughout vitellogenin purification (Fig. 1E). Although the 40 kDa band was not recognized as vitellogenin by western blots (Fig. 1F,G), the closeness of 40 kDa to the size difference between 180 and 150 kDa vitellogenin made the small protein product interesting, and we followed it up for further identification.

We initially tested whether the mechanical stress or buffer conditions that were used during protein extraction from fat body, but not from hemolymph, could induce cleavage in some of the 180 kDa vitellogenin to produce the 150 and 40 kDa pieces observed during gel separation of fat body proteins. For example, a harshly treated protein might be denatured and, thus, become more vulnerable to enzymatic cleavage. To test this, we vortexed and sonicated hemolymph proteins (180 kDa vitellogenin only) under the same conditions and buffer with protease inhibitors as the abdominal fat body extract (see Materials and methods), but untreated and sonicated hemolymph vitellogenin remained equally intact at 180 kDa (supplementary material Fig. S1A). Semiquantitive estimation of the 150 and 40 kDa vitellogenin bands (using standard proteins for comparison) suggested an excess of the 40 kDa fragment (Materials and methods; see also supplementary material Fig. S1B). The average observed ratio was 8.9-fold more 40 than 150 kDa vitellogenin (standard deviation = 3.9).

In summary, our vitellogenin purification protocols for honeybee hemolymph and fat body yielded a 180 kDa protein product in hemolymph with some fragmentation into a 150 kDa product, and two immunoreactive products in fat body (180 and 150 kDa). In addition, a 40 kDa protein present in all fat body vitellogenin purification steps was identified, and could not be ascribed to a technical artifact.

Vitellogenin identification by mass spectrometry

We initially used MALDI-TOF analysis to test amino acid sequence correspondence between the bands detected at 180 and 150, as well as at 40, kDa in extracts of honeybee abdominal fat body (Fig. 2). The peptides identified from the 180 kDa band (Fig. 2A) were spread along the entire sequence of vitellogenin (Fig. 2D), confirming correspondence to the full-length protein. The 150 kDa vitellogenin mostly lacked hits at the N terminus (Fig. 2B,E). Conversely, in the case of the 40 kDa band, the peptide hits were heavily concentrated at the vitellogenin N terminus (residues 54–294) (Fig. 2C,F). This pattern of hits and misses is consistent with vitellogenin that has fragmented into a smaller N-terminal (40 kDa) and a larger C-

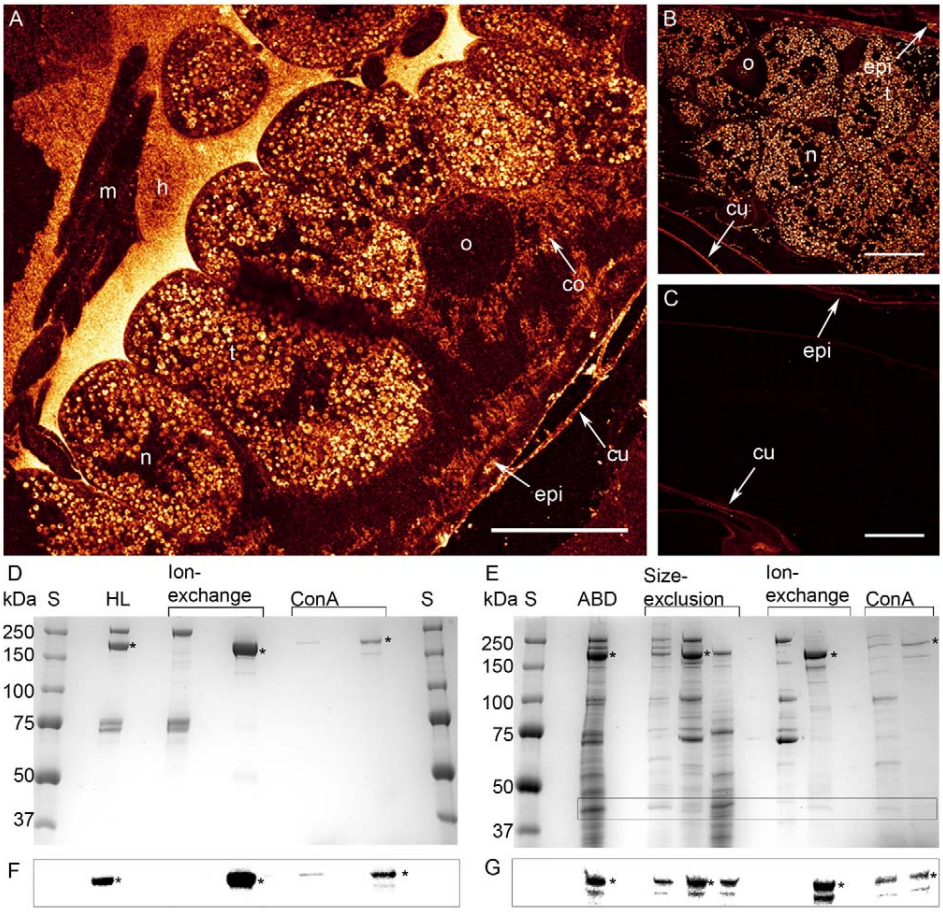


Fig. 1. Immunolocalization of vitellogenin in the abdominal carcass of *diutinus* workers and purification of vitellogenin by SDS-PAGE, as verified by western blots. (A,B) Vitellogenin immunoreactivity (white/yellow) is detected in hemolymph (h) and the cytoplasm of trophocyte cells (t), but is not present in the cell nuclei, muscle cells (m), or the oenocyte cells of the fat body (o). Epidermis (epi, epithelial tissue), connective tissue (co) and cuticle (cu) are autofluorescent with confocal imaging, and thus visible also in the control (C). (D-G) Asterisks indicate full-length vitellogenin in the vitellogenin purification steps. S, molecular weight standard. (D) Purification from hemolymph. HL: raw hemolymph (3 μg protein). Ion-exchange: first lane shows flow-through (2 μg); second lane shows high salt elute (2 μg). ConA: first lane is Concanavalin A affinity chromatography column flow through (1 μg); second lane is eluted pure vitellogenin (0.8 μg). (E) Purification from abdomen. The 40 kDa vitellogenin is framed. ABD: sonicated abdominal protein extract (10 μg). Size-exclusion: chromatography fractions eluted by their size (4 μg each lane). Ion-exchange: chromatography flow through and subsequently high salt elute (2 μg each). ConA: affinity chromatography column flow through and on the last lane the pure vitellogenin eluate (1.4 μg each). (F,G) Western blots confirming the presence of vitellogenin in each purification step. The lanes are identical to the lanes for D and E above. (F) There is 2.5 μg protein on the HL lane, 2 μg on both ion-exchange lanes and 0.8 μg on the ConA lanes. (G) There is 1.5 μg abdominal protein on the gel, 1 μg on the size-exclusion lanes, 1.1 μg on the ion-exchange lanes and 1 μg on the ConA lanes.

terminal (150 kDa) piece. No hits to honeybee proteins were verified for the control samples (gel sections picked between the respective bands; see Materials and methods).

The 40 kDa sample MALDI-TOF peptide hits were subjected to further fragmentation analysis to establish whether they were supported: 70–77, VWNGQYAR (not supported); 152–168, VNSVQVPTDDEPYASFK (confirmed); 218–225, NFDNCDQR (confirmed); 226–237, INYHFGMTDNSR (confirmed); 265–276, HFTIQSSVTTSK (confirmed); 287–294, QNGLVLSR (not supported); and 524–533, SMEAANIMSK (not supported) (see Fig. 2F). Finally, we analyzed the 40 kDa sample with LC-MS/MS and obtained a profile of 10 peptides (Fig. 2F), which were all confirmed to match exclusively to the N-terminal vitellogenin sequence from residue 53 to 294.

Thus, we observed a correspondence between three vitellogenin products in honeybee abdominal fat body: the 180 kDa band corresponds to full-length protein, the 150 kDa band corresponds to vitellogenin lacking the N terminus and the 40 kDa band corresponds to the N terminus of the protein.

Phosphorylation and glycosylation status of vitellogenin and its fragments

We found that full-length vitellogenin (180 kDa), and the 150 and 40 kDa fragments, are phosphorylated. This result was obtained from raw hemolymph and fat body proteins, as well as from purified abdominal vitellogenin samples (Fig. 3).

Full-length vitellogenin was previously found to be glycosylated, as the protein binds to ConA columns, in which lectins will crosslink to N-linked glyco groups (Ohyama et al., 1985; Wheeler and Kawooya, 1990). Likewise, we found that both the 150 and 40 kDa vitellogenin bind to the ConA column, as they were retained in the final step of the fat body purification protocol (Fig. 1E).

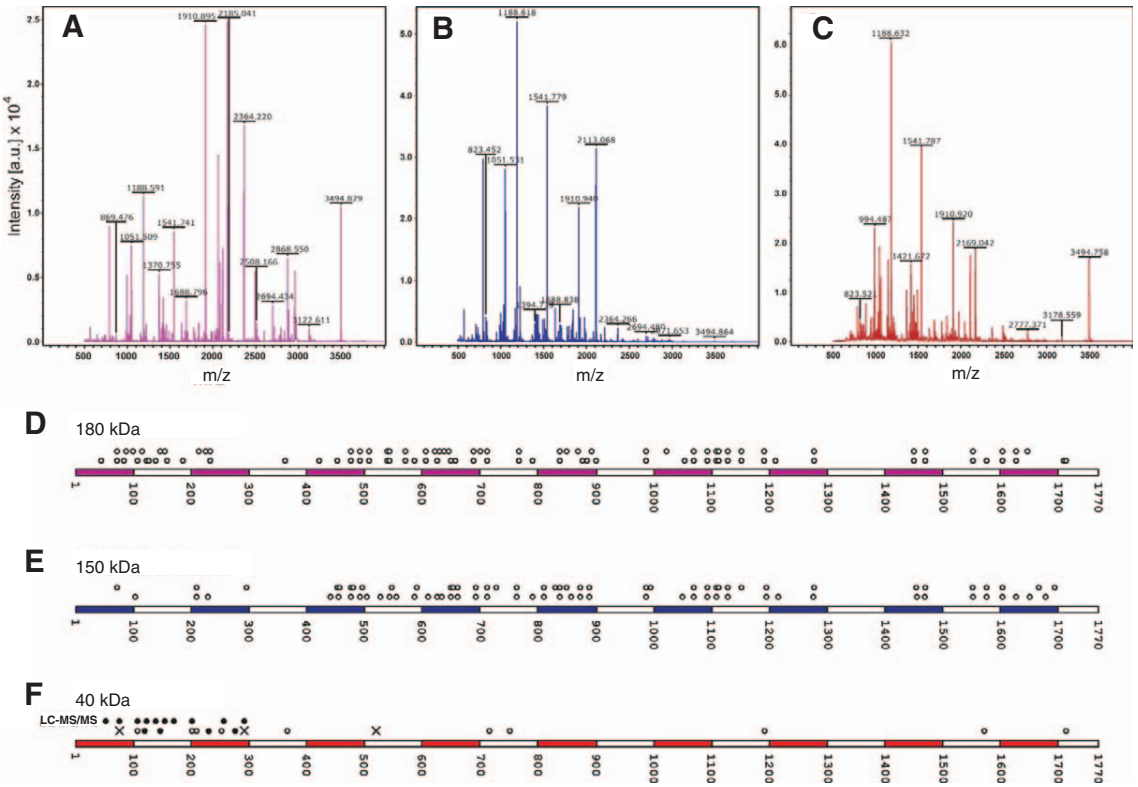


Fig. 2. Mass-spectrometry identification of vitellogenin-derived peptides from the 180 kDa, 150 kDa and 40 kDa SDS-PAGE bands. Representative spectra of the 180 kDa (A), 150 kDa (B) and 40 kDa (C) vitellogenin bands. Selected peaks denoted by their mass/charge ratio are indicated in the figure panels. (D–F) Mapping of the peptides onto the vitellogenin primary sequence identifies which regions of vitellogenin can be assigned to each fragment. Each circle above a sequence denotes a peptide matching vitellogenin fragment that can be found in this position along the primary sequence. In D and E, the two rows of circles above sequences are independent repeats of a MALDI-TOF experiment on 180 and 150 kDa vitellogenin, respectively. (F) Vitellogenin peptide matches, 40 kDa gel band. LC-MS/MS peptide hits on the first lane indicated with black circles. MALDI-TOF peptide hits on the second lane are marked by seven peptides that were either confirmed (black circle) or not supported (crosses) by fragmentation analysis.

Furthermore, in our deglycosylation test, the sizes of the 180, 150 and 40 kDa vitellogenin species were reduced (by ~4.2, ~2.4 and ~2.5 kDa, respectively) as a result of the loss of this sugar group (Fig. 4). Thus, both the ConA binding and the size shift after deglycosylation support the presence of two glyco groups in 180 kDa vitellogenin that will segregate to one group each in the 40 and 150 kDa fragments.

A positively charged patch, a lipophilic cavity and an indication of insect-specific coils in the honeybee N-terminal domain

We succeeded in building a β-barrel-like homology model of the N-terminal honeybee vitellogenin region (N-sheet) (Anderson et al., 1998; Raag et al., 1988) (Fig. 5A; Table 1). There were no Ramachandran outliers in the β-sheets or α-helices indicating an excellent protein back-bone chemistry (Ramachandran et al., 1963). Five outliers were located in coils (Table 1). Loops have more evolutional plasticity than β-sheets or α-helices (Panchenko et al., 2005), which makes these regions more difficult to model accurately by homology, as their sequence differs from the template structure. Our electrostatic map of the model shows a patch of positively charged residues on the β-sheets (Fig. 5B). The hydrophobic inner cavity is predominately positively charged – contrary to lamprey lipovitellin, in which the cavity is partly negatively charged (Fig. 5B). Simulation of the model was performed in order to test the stability of the predicted secondary structures, and of the whole protein fragment. A relative mobility of the loops of the model compared with its more stationary α-helixes and β-sheets was observed, as is expected for a stable protein (Fig. 5C; supplementary material Fig. S2 for RMSD).

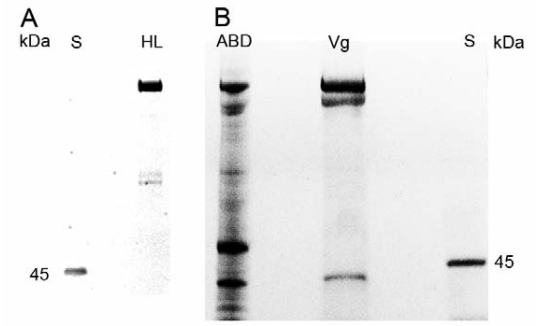


Fig. 3. Phosphorylated proteins in hemolymph (A), abdominal protein extract and purified vitellogenin (B). ProQ phosphostaining with PeppermintStick (S) standard that provides both a positive and negative control (showing here only phosphorylated ovalbumin at 45 kDa). (A) HL: the *diutinus* worker hemolymph sample (2.5 µg protein) is dominated by full-length vitellogenin. (B) ABD: the abdominal protein extract (14 µg protein) shows 180, 150 and 40 kDa vitellogenin. Additionally, other abdominal proteins appear as putatively phosphorylated. Vg: the vitellogenin purified from *diutinus* abdomens (4.7 µg) shows 180, 150 and 40 kDa vitellogenin bands to be phosphorylated.

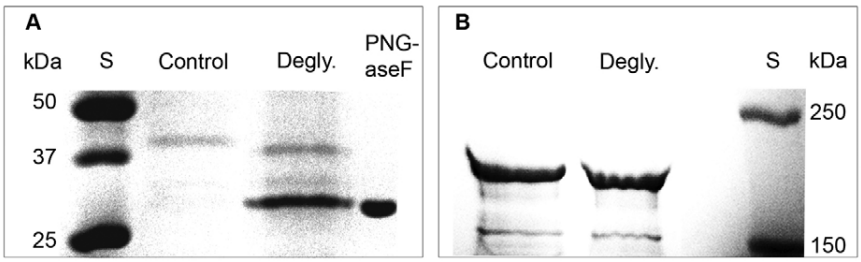


Fig. 4. Effects of deglycosylation by PNG-aseF on abdominal vitellogenin on SDS-PAGE. S, molecular weight standard. (A) Deglycosylation of the 40 kDa vitellogenin. The deglycosylated sample appears as a ~2.5 kDa lighter molecular weight band. PNGaseF, the deglycosylation enzyme. (B) Deglycosylation of the 180 and 150 kDa vitellogenin. The full-length vitellogenin shifts ~4.2 kDa and the 150 kDa vitellogenin shifts ~2.4 kDa after the removal of N-linked glycans.

We found two insect-specific loops in the 40 kDa vitellogenin fragment (location indicated with pink arrows in Fig. 5A). These two regions could not be properly modelled, as the vertebrate template structure lacks the corresponding sequences. The regions were considered as loops based on PSIPRED (Jones, 1999) and Gor4 (Garnier et al., 1996) secondary structure predictions (predictions not shown) and because of the observation of the conserved coil-indicator residues glycine and proline (Krieger et al., 2005) in these

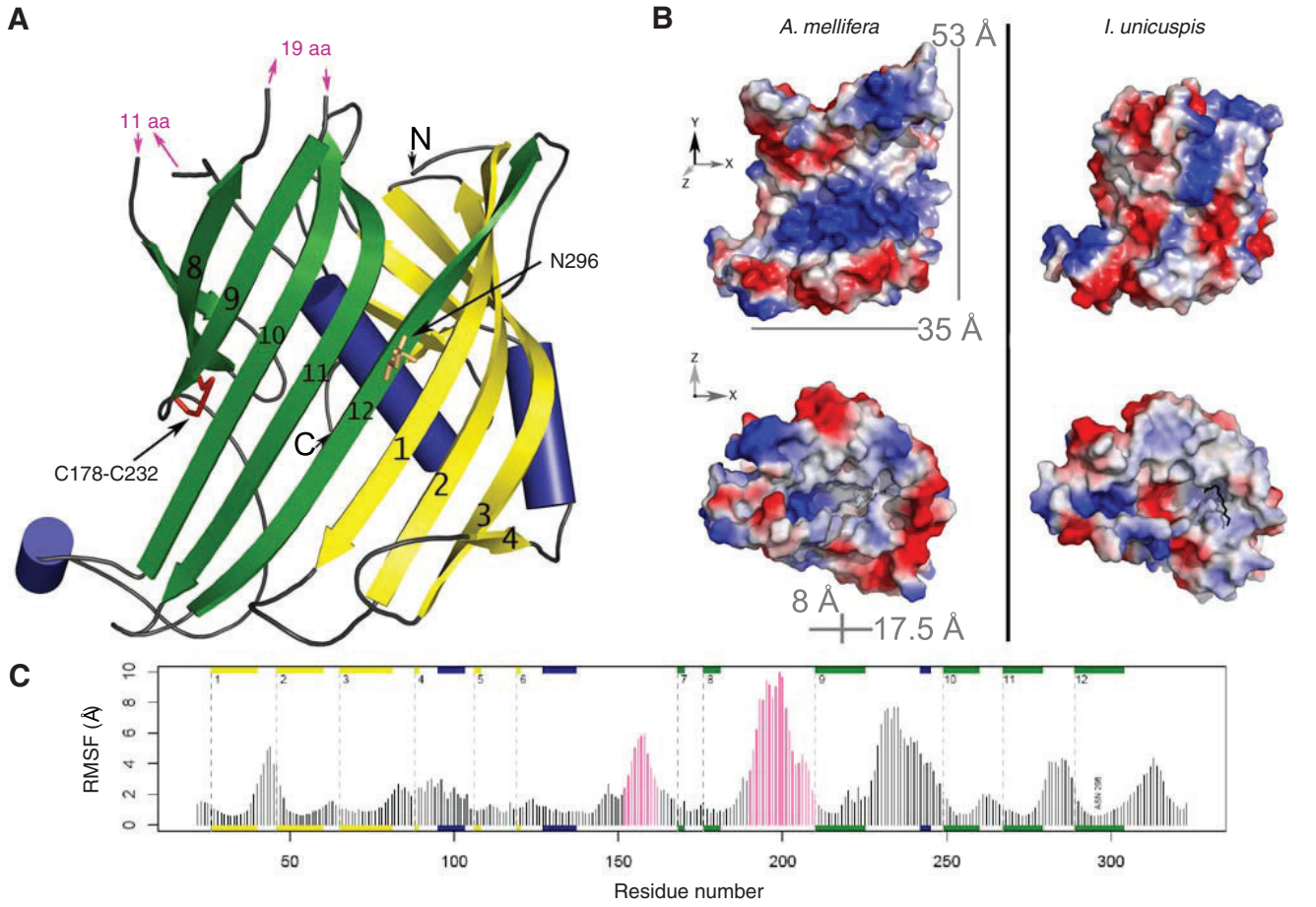


Fig. 5. Structural characterization of the honeybee vitellogenin N-sheet. (A) A homology model based on the X-ray structure of lamprey lipovitellin. Helices are shown in blue, β-sheets are yellow and green (green for those corresponding to the homodimerization site of lamprey lipovitellin), the disulphide bridge (C178–C232) is shown as a red stick and the putatively glycosylated asparagine 296 residue is shown as an orange stick. The N and C termini are indicated and the β-sheets are numbered starting at the N terminus. The missing loop residues of 11 amino acids (from central helix to β-sheet 7) and 19 amino acid residues (from β-sheets 8 to 9) are indicated by pink arrows. (B) Electrostatic comparison between the honeybee (*A. mellifera*) vitellogenin N-sheet model and the lamprey (*I. unicuspis*) lipovitellin N-sheet. The insect-specific loops, with energetically minimized but otherwise random structure, are included. Red color represents the negatively charged Asp and Glu residues; blue color represents the positively charged His, Arg and Lys residues. Honeybee N-sheet accommodates a large positively charged patch (upper left; residues R33, R35, K53, R256, K276 and R294 on β-sheets 1–2 and 10–12), whereas the β-sheets of lamprey show a more even charge distribution (upper right). The cavity inside the N-sheet also differs in charge: the honeybee vitellogenin cavity (bottom left) contains a positive charge (K205, K 207, R251), whereas the lamprey cavity (bottom right) contains a negative charge due to two glutamic acid residues not present in honeybee. The black rod inside the lamprey cavity is a lipid found in the crystal structure. The dimensions of the honeybee N-sheet inner cavity are circa 8×17.5 Å, and the whole half-β-barrel is approximately 53×35 Å in size. (C) The positional fluctuations as determined by a 10 ns long MD simulation. Secondary structure elements are indicated schematically with helices in dark blue and strands in yellow and green, as in A. Pink bars depict the 11- and 19-residue long loops.

Table 1. Details of the honeybee N-sheet model and its template structure

		Corresponding amino acids or percentage of total model residues
Number of amino acids in honeybee model	302	W22-Y323
Ichthyomyzon unicuspis template amino acids (PDB ID: 1LSH)	262	F18-Y281
Number of helices	3	L95-S103, T127-L140 (central helix), P240-N245
Number of sheets	12	(1) G25-T38, (2) T48-K61, (3) N64-Q79, (4) E87-E89, (5) K105-I109, (6) R117-V121, (7) F167-E171, (8) K177-P186, (9) H210-N218, (10) F248-E261, (11) F266-S280, (12) N288-L304
Internal missing residues	30	V152-E162 (11 amino acids in a loop), F190-G208 (19 amino acids in a loop)
Ramachandran plot, preferred regions	155	89.5%
Ramachandran outliers	5	1.9%; R110, K112, N235, H265, Q287

regions. A multiple alignment of 29 species shows that these loops are conserved in many insect species but are absent in vertebrates (supplementary material Fig. S3).

DISCUSSION

Vitellogenin influences hormone signaling, food-related behavior, immunity, stress resistance and longevity in honeybee workers (Amdam et al., 2004a; Amdam et al., 2004b; Fluri et al., 1982; Nelson et al., 2007; Seehuus et al., 2006). The molecular properties of the protein, however, are not well known. Here, we identify a previously uncharacterized 40 kDa vitellogenin fragment in the worker fat body, and reject the hypothesis that a complimentary 150 kDa unit represents a breakdown product. We describe the post-translational modification of the 40 kDa and 150 kDa vitellogenin units, as well as of the full-length vitellogenin, in two body compartments: hemolymph and fat body. Finally, we present the first homology model for the 40 kDa unit of this vitellogenin.

Initially, we established compartment-specific purification protocols for the target protein: from hemolymph and fat body. Immunofluorescence staining confirmed vitellogenin to be abundantly and specifically present in these compartments. Improved yields of vitellogenin were facilitated by our choice of source bees: workers that were induced to accumulate vitellogenin based on a natural process in which vitellogenin-rich worker bees develop in the absence of brood rearing (Maurizio, 1950). Our protocol for hemolymph vitellogenin purification is based on an extraction protocol optimized for honeybee queens (Wheeler and Kawooya, 1990). Building from this method, we were able to omit a time-consuming (16 h) density ultra-centrifugation step by pooling ion-exchange chromatography fractions that do not contain apolipophorin or hexamerins. Our protocol for hemolymph may not only be time efficient, but also yield more intact vitellogenin as the protein has less time to degrade during handling – although we did experience limited vitellogenin fragmentation. Our procedure for fat body was built on the same original protocol (Wheeler and Kawooya, 1990), to which we added size-exclusion chromatography as the first step.

After purification, we found three vitellogenin products in fat body that separated on SDS-PAGE gels as bands of 180, 150 and 40 kDa in size. Western blots confirmed the 180 and 150 kDa bands to be vitellogenin, as seen previously (Amdam et al., 2003a). The antibody, however, did not immunostain the 40 kDa band. The lack of recognition of this fragment, which we identified as vitellogenin by other means (see below), is likely to be due to a smaller fraction of the polyclonal antibody population being specific to an epitope or epitopes on the N terminus of vitellogenin relative to the remainder of the protein, as seen in vinculin antibody-binding pattern (Kilic and Ball, 1991).

We used MALDI-TOF mass-spectrometry to verify that the 40 kDa protein band was a vitellogenin fragment. Further LC-MS/MS experiment showed peptide hits exclusively at the N terminus of vitellogenin (residues 54–294). The MALDI-TOF peptide hits of the 150 kDa vitellogenin, reciprocally, were centered between residues 427–1700, supporting our proposition that the 40 and 150 kDa vitellogenin fragments correspond to non-overlapping products of the full-length 180 kDa vitellogenin. Note that some regions in the full-length and 150 kDa vitellogenin products seem to systematically lack MALDI-TOF peptide hits (see Fig. 2A,B). We believe these misses are caused by post-translational modification. For instance, phospho groups at phosphorylated peptides, particularly at the so-called polyserine linker between the residues 351–381 of vitellogenin, tend to complicate detection (Steen et al., 2006).

The 180, 150 and 40 kDa vitellogenin units were shown to be phosphorylated by phosphoprotein staining. Strong candidates for phosphorylation are the 13 serine residues situated between residues 351 and 381 (Don-Wheeler and Engelmann, 1997; Tufail and Takeda, 2002). This region approximately lies at the intersection of the 40 kDa and the 150 kDa vitellogenin. In some cases, protein cleavage is known to be regulated by phosphorylation (Alexandru et al., 2001; Walter et al., 1999), and this cannot be excluded in the case of honeybee vitellogenin. Where is the exact cleavage site located? Honeybee vitellogenin sequence lacks the RXXR consensus cleavage site near the polyserine linker present in many other insect vitellogenins (Barr, 1991; Rouille et al., 1995). The LC-MS/MS peptide hits end at residue 294, but this does not prove it to be the terminal amino acid of 40 kDa vitellogenin. According to a theoretical molecular mass calculation, a molecular mass of 40 kDa corresponds to the residues 17–370 (1–16 is a cleaved signal sequence (Piulachs et al., 2003). Amino acid 370 is in the middle of the polyserine motif: nslssseekl kq**d**ilnlrtd issssssisss (residue 370 bolded). Hypothetically, if vitellogenin were cut at residue 370, there would be multiple serine tracts in both the resulting fragments, 40 and 150 kDa, which is in accordance with the finding that both fragments are clearly phosphorylated. However, there are many possible phosphorylation sites along the whole vitellogenin sequence. Therefore, further studies are needed to trace the exact site, and to understand the regulation of the cleavage.

Another post-translational modification – glycosylation – of 180, 150 and 40 kDa vitellogenin is supported by binding to Concanavaline A that has affinity for N-glycosylated proteins (Ohyama et al., 1985). The deglycosylation experiment also showed size shifts of the fragments on a SDS-PAGE gel. The vitellogenin sequence contains three Asn-X-Ser/Thr sequences for possible N glycosylation (Marshall, 1974): one in the N terminus (Asn296; highlighted in Fig. 5A) and two towards the C terminus (Asn1067 and Asn1153). The sum of our size-shift estimates for 150 and 40 kDa vitellogenin (~2.4 and ~2.5 kDa, respectively) differs from the 180 kDa vitellogenin estimate (~4.2 kDa) by 0.7 kDa, which could be due to a minor inaccuracy on our SDS-PAGE analysis.

For example, decapod *Cherax quadricarinatus* vitellogenin contains a glycan of a similar size, 2067.7 Da Glc3Man9GlcNAc2, among other glycans (Khalaila et al., 2004).

Our homology model based on the lamprey lipovitellin crystal structure (Anderson et al., 1998; Mann et al., 1999; Raag et al., 1988; Thompson and Banaszak, 2002) of N-terminal vitellogenin, comprises residues 22–323 that are present in the 40 kDa fragment. Interestingly, the N-sheet accommodates two putative loop regions that are found in insects only, with the exception of the vitellogenin of the cockroaches *Blattella germanica* and *Leucophaea maderae* (supplementary material Fig. S3). These loops contain several conserved proline and glycine residues that usually hinder α-helix or β-sheet formation (Krieger et al., 2005), and there are also some conserved charged residues in the loops. The loops are located at the lamprey homodimerization region, and might thus be connected to possible dimerization differences between insect and vertebrate vitellogenins. The presence of many charged residues raises a question of a more catalytical role of these loops. There are many examples of loops with important biological functions – some loops even possess active site residues (Hedstrom et al., 1992; Joseph et al., 1990; Williams and McCammon, 2009; Williams and Essex, 2009; Xu et al., 2007). As they are clearly conserved, a structural or catalytic role of some kind could be hypothesized for these N-terminal loops in insects.

A comparison between the lamprey lipovitellin and the honeybee vitellogenin N-sheet reveals differences in their charge distribution. There are six Arg and His residues on honeybee N-sheet β-sheets that form a positively charged patch. In lamprey, the positive and negative charges are evenly distributed. A positively charged patch might attract negatively charged ligands (Cherstvy, 2009; Corsico et al., 2005; Grigg et al., 2010; Honig and Nicholls, 1995; Horii et al., 2004). The interior of the N-sheet is lined by a mainly hydrophobic cavity, in which a lipid molecule is present in the lamprey lipovitellin crystal structure (Thompson and Banaszak, 2002). It seems likely that the 40 kDa vitellogenin of honeybee also has the ability to bind a lipid, as the cavity is even more exposed to ligands when the rest of the protein is removed. The honeybee N-sheet cavity is positively charged at one side. This could offer an ideal site for a fatty acid with a negatively charged headgroup to bind, like in the case of intestinal fatty acid binding protein (Cistola et al., 1989; Jakoby et al., 1993; Sacchettini et al., 1989).

According to our study, the honeybee vitellogenin fragmentation pattern differs in fat body and hemolymph. Vitellogenin cleavage is abundant in abdominal (fat body containing) protein extract, whereas we only observed the intact 180 kDa protein in fresh worker hemolymph. However, the hemolymph vitellogenin also formed a 150 kDa unit after sampling and during purification (see Results) (see also Wheeler and Kawooya, 1990). In *Oreochromis aureus*, the N terminus of vitellogenin was identified as the vitellogenin receptor-binding site (Li et al., 2003). If this is also the case with honeybee vitellogenin, then 150 kDa vitellogenin formation in hemolymph would be undesirable, as, thereafter, the protein could not be taken up by tissues via receptor-mediated endocytosis. It is an open question as to how vitellogenin integrity remains intact in the hemolymph of a living bee.

Lack of 40 or 150 kDa units in fresh hemolymph might also suggest that cleaved vitellogenin is not secreted from the fat body. The intracellular role or roles of the 40 kDa unit are currently unknown, but it is tempting to speculate on a mechanism that links the putative lipid binding properties of 40 kDa vitellogenin to a biologically relevant context in which worker bees store increasing amounts of vitellogenin in their fat bodies. A fatty acid mixture from larvae, called brood pheromone (Le Conte et al., 1994), inhibits fat body vitellogenin accumulation. In accordance with this, larvae must be absent from the colony milieu before vitellogenin-rich bees develop (Smedal et al., 2009). In theory, the lipophilic cavity of the 40 kDa vitellogenin might bind such pheromone fatty acids. When larval brood is present, this interaction could stabilize full-length vitellogenin and perhaps facilitate vitellogenin secretion. Reciprocally, more vitellogenin would remain in fat body when brood and pheromone is absent. We also observed an excess of the 40 kDa vitellogenin in our abdominal extracts when compared with the 150 kDa vitellogenin (Fig. 1E; see also supplementary material Fig. S1B). The intensity of any one protein band separating on a one-dimentional gel can be cofounded by overlapping or partly overlapping bands. However, we did not observe other 40 kDa proteins during the various steps of vitellogenin purification, and mass-spectrometry analysis detected only 40 kDa vitellogenin in the same range. We therefore conclude that the 40 kDa band might represent an almost pure fraction of vitellogenin. A bias towards the 40 kDa fragment, relative to the 150 kDa fragment, could indicate that the 150 kDa fragment is consumed in the fat body, whereas the remaining 40 kDa vitellogenin is involved in an additional mechanism for preserving vitellogenin stores; for instance, by repressing exocytosis via membrane binding.

The findings presented here represent a step forward in the understanding of honeybee vitellogenin. Knowledge about this protein could, eventually, provide unique illustrations on how insect social organization can be achieved at the level of protein structure. Moreover, insights into the molecular properties of vitellogenins are of broader general interest, as alternative functions of this family of yolk precursor proteins are now surfacing in diverse taxa: including in the immune defense against viruses and bacteria in fish (Garcia et al., 2010; Li et al., 2008; Shi et al., 2006), and in aging in *Caenorhabditis elegans* (Nakamura et al., 1999).

LIST OF ABBREVIATIONS

ABD	abdominal protein extract
ACN	acetonitrile
ConA	concanavalin A
Degly.	deglycosylated
HL	hemolymph
IIS	insulin/insulin-like signaling
JH	juvenile hormone
LC-MS/MS	liquid chromatography tandem mass spectrometry
MALDI-TOF/TOF	tandem matrix-assisted laser desorption/ionization time-of-flight
MD	molecular dynamics
S	molecular weight standard
TFA	trifluoroacetic acid
Vg	vitellogenin

ACKNOWLEDGEMENTS

The work was performed at the Department of Biochemistry at University of Bergen in the laboratory headed by professor Aurora Martinez. H.H., Ø.H. and G.V.A. conceived and designed the experiments. H.H., Ø.H. and B.S. collected the samples. H.H. performed the experiments with the exception of mass-spectroscopy and immunohistochemistry. Ø.H., H.H. and G.V.A. analyzed the mass-spectroscopy data. B.S. performed the immunohistochemistry. L.S. performed model minimization and simulation. H.H., Ø.H. and G.V.A. wrote the paper. L.S. wrote the model simulation and minimization section of the paper, and B.S. wrote the immunohistochemistry section. We thank Professor Martinez for kindly supporting and assisting the project, and for access to instruments and reagents. We also thank Randi Svebak for technical support. We thank C. Kreibich, M. Speth and A. H. Søyland for providing bees and assisting the sampling. The mass-spectroscopy experiments – from sample extraction from gel pieces until data collection and protein identification – and assistance with data interpretation and writing were performed by the PROBE Proteomic Unit at University of Bergen. We thank Dr Anne Døskeland for expert technical support. The PROBE work was partly supported by the National Program for Research in

Functional Genomics (FUGE) funded by the Norwegian Research Council. Ø.H. was supported by the Research council (#185306) and Norwegian Cancer Society (#58240001). G.V.A. was supported by the Research Council of Norway (#180504, 185306 and 191699), the National Institute on Aging (NIA P01 AG22500) and the PEW Charitable Trust.

REFERENCES

Alexandru, G., Uhlmann, F., Mechtler, K., Poupart, M. A. and Nasmyth, K. (2001). Phosphorylation of the cohesin subunit Scc1 by Polo/Cdc5 kinase regulates sister chromatid separation in yeast. *Cell* **105**, 459-472.

Amdam, G. V., Norberg, K., Hagen, A. and Omholt, S. W. (2003a). Social exploitation of vitellogenin. *Proc. Natl. Acad. Sci. USA* **100**, 1799-1802.

Amdam, G. V., Simoes, Z. L., Guidugli, K. R., Norberg, K. and Omholt, S. W. (2003b). Disruption of vitellogenin gene function in adult honeybees by intra-abdominal injection of double-stranded RNA. *BMC Biotechnol.* **3**, 1.

Amdam, G. V., Hartfelder, K., Norberg, K., Hagen, A. and Omholt, S. W. (2004a). Altered physiology in worker honey bees (Hymenoptera: Apidae) infested with the mite Varroa destructor (Acari: Varroidae): a factor in colony loss during overwintering? *J. Econ. Entomol.* **97**, 741-747.

Amdam, G. V., Simoes, Z. L. P., Hagen, A., Norberg, K., Schroder, K., Mikkelsen, O., Kirkwood, T. B. L. and Omholt, S. W. (2004b). Hormonal control of the yolk precursor vitellogenin regulates immune function and longevity in honeybees. *Exp. Gerontol.* **39**, 767-773.

Amdam, G. V., Norberg, K., Omholt, S. W., Kryger, P., Lourenco, A. P., Bitondi, M. M. G. and Simoes, Z. L. P. (2005). Higher vitellogenin concentrations in honey bee workers may be an adaptation to life in temperate climates. *Insectes Soc.* **52**, 316-319.

Anderson, T. A., Levitt, D. G. and Banaszak, L. J. (1998). The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. *Structure* **6**, 895-909.

Arrese, E. L. and Soulages, J. L. (2009). Insect fat body: energy, metabolism, and regulation. *Annu. Rev. Entomol.* **55**, 207-225.

Avarre, J. C., Lubzens, E. and Babin, P. J. (2007). Apolipocrustacein, formerly vitellogenin, is the major egg yolk precursor protein in decapod crustaceans and is homologous to insect apolipophorin II/I and vertebrate apolipoprotein B. *BMC Evol. Biol.* **7**, 3.

Baker, M. E. (1988). Is Vitellogenin an ancestor of Apolipoprotein-B-100 of human low-density lipoprotein and human lipoprotein-lipase. *Biochem. J.* **255**, 1057-1060.

Barr, P. J. (1991). Mammalian subtilisins: the long-sought dibasic processing endoproteases. *Cell* **66**, 1-3.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-254.

Case, D. A. D. T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowley, M., Walker Ross, C., Zhang, W., Merz, K. M. et al. (2008). AMBER 10. San Francisco, CA: University of California, San Francisco.

Cherstvy, A. G. (2009). Positively charged residues in DNA-binding domains of structural proteins follow sequence-specific positions of DNA phosphate groups. *J. Phys. Chem. B* **113**, 4242-4247.

Cistola, D. P., Sacchettini, J. C., Banaszak, L. J., Walsh, M. T. and Gordon, J. I. (1989). Fatty acid interactions with rat intestinal and liver fatty acid-binding proteins expressed in Escherichia coli. A comparative 13C NMR study. *J. Biol. Chem.* **264**, 2700-2710.

Corona, M., Velarde, R. A., Remolina, S., Moran-Lauter, A., Wang, Y., Hughes, K. A. and Robinson, G. E. (2007). Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. *Proc. Natl. Acad. Sci. USA* **104**, 7128-7133.

Corsico, B., Franchini, G. R., Hsu, K. T. and Storch, J. (2005). Fatty acid transfer from intestinal fatty acid binding protein to membranes: electrostatic and hydrophobic interactions. *J. Lipid Res.* **46**, 1765-1772.

Don-Wheeler, G. and Engelmann, F. (1997). The biosynthesis and processing of vitellogenin in the fat bodies of females and males of the cockroach Leucophaea maderae. *Insect Biochem. Mol. Biol.* **27**, 901-918.

Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T. et al. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999-2012.

Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132.

Engels, W. and Fahrenhorst, H. (1974). Alters- und kastenspezifische Veränderungen der Haemolymph-protein-spectren bei Apis mellifera. *Wilhelm Roux Arch. Entwickl. Mech. Org.* **174**, 285-296.

Fluri, P., Wille, H., Gerig, L. and Luscher, M. (1977). Juvenile hormone, vitellogenin and hemocyte composition in winter worker honeybees (Apis-Mellifera). *Experientia* **33**, 1240-1241.

Fluri, P., Luscher, M., Wille, H. and Gerig, L. (1982). Changes in weight of the pharyngeal gland and hemolymph titers of juvenile-hormone, protein and vitellogenin in worker honey bees. *J. Insect Physiol.* **28**, 61-68.

Garcia, J., Munro, E. S., Monte, M. M., Fourrier, M. C., Whitelaw, J., Smail, D. A. and Ellis, A. E. (2010). Atlantic salmon (Salmo salar L.) serum vitellogenin neutralises infectivity of infectious pancreatic necrosis virus (IPNV). *Fish Shellfish Immunol.* **29**, 293-297.

Garnier, J., Gibrat, J. F. and Robson, B. (1996). GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol.* **266**, 540-553.

Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A. and Caves, L. S. (2006). Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695-2696.

Grigg, J. C., Cooper, J. D., Cheung, J., Heinrichs, D. E. and Murphy, M. E. (2010). The Staphylococcus aureus siderophore receptor HtsA undergoes localized conformational changes to enclose staphyloferrin A in an arginine-rich binding pocket. *J. Biol. Chem.* **285**, 11162-11171.

Guidugli, K. R., Nascimento, A. M., Amdam, G. V., Barchuk, A. R., Omholt, S., Simoes, Z. L. and Hartfelder, K. (2005). Vitellogenin regulates hormonal dynamics in the worker caste of a eusocial insect. *FEBS Lett.* **579**, 4961-4965.

Hedstrom, L., Szilagyi, L. and Rutter, W. J. (1992). Converting trypsin to chymotrypsin: the role of surface loops. *Science* **255**, 1249-1253.

Honig, B. and Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science* **268**, 1144-1149.

Horii, K., Okuda, D., Morita, T. and Mizuno, H. (2004). Crystal structure of EMS16 in complex with the integrin alpha2-I domain. *J. Mol. Biol.* **341**, 519-527.

Hunt, G. J., Amdam, G. V., Schlipalius, D., Emore, C., Sardesai, N., Williams, C. E., Rueppell, O., Guzman-Novoa, E., Arechavaleta-Velasco, M., Chandra, S. et al. (2007). Behavioral genomics of honeybee foraging and nest defense. *Naturwissenschaften* **94**, 247-267.

Jakoby, M. G., Miller, K. R., Toner, J. J., Bauman, A., Cheng, L., Li, E. and Cistola, D. P. (1993). Ligand-protein electrostatic interactions govern the specificity of retinol- and fatty acid-binding proteins. *Biochemistry* **32**, 872-878.

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195-202.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935.

Joseph, D., Petsko, G. A. and Karplus, M. (1990). Anatomy of a conformational change: hinged "lid" motion of the triosephosphate isomerase loop. *Science* **249**, 1425-1428.

Kenyon, C. J. (2010). The genetics of ageing. *Nature* **464**, 504-512.

Khalaila, I., Peter-Katalinic, J., Tsang, C., Radcliffe, C. M., Aflalo, E. D., Harvey, D. J., Dwek, R. A., Rudd, P. M. and Sagi, A. (2004). Structural characterization of the N-glycan moiety and site of glycosylation in vitellogenin from the decapod crustacean Cherax quadricarinatus. *Glycobiology* **14**, 767-774.

Kilic, F. and Ball, E. H. (1991). Partial cleavage mapping of the cytoskeletal protein vinculin. Antibody and talin binding sites. *J. Biol. Chem.* **266**, 8734-8740.

Krieger, F., Moglich, A. and Kiefhaber, T. (2005). Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *J. Am. Chem. Soc.* **127**, 3346-3352.

Le Conte, Y., Streng, L. and Trouiller, J. (1994). The recognition of larvae by worker honeybees. *Naturwissenschaften* **81**, 462-465.

Lee, M. C. and Duan, Y. (2004). Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model. *Proteins* **55**, 620-634.

Lehtonen, J. V., Still, D. J., Rantanen, V. V., Ekholm, J., Bjorklund, D., Iftikhar, Z., Huhtala, M., Repo, S., Jussila, A., Jaakkola, J. et al. (2004). BODIL: a molecular modeling environment for structure-function analysis and drug design. *J. Comput. Aided. Mol. Des.* **18**, 401-419.

Li, A., Sadasivam, M. and Ding, J. L. (2003). Receptor-ligand interaction between vitellogenin receptor (VtgR) and vitellogenin (Vtg), implications on low density lipoprotein receptor and apolipoprotein B/E. The first three ligand-binding repeats of VtgR interact with the amino-terminal region of Vtg. *J. Biol. Chem.* **278**, 2799-2806.

Li, Z., Zhang, S. and Liu, Q. (2008). Vitellogenin functions as a multivalent pattern recognition receptor with an opsonic activity. *PloS One* **3**, e1940.

Maley, F., Trimble, R. B., Tarentino, A. L. and Plummer, T. H., Jr (1989). Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Anal. Biochem.* **180**, 195-204.

Mann, C. J., Anderson, T. A., Read, J., Chester, S. A., Harrison, G. B., Kochl, S., Ritchie, P. J., Bradbury, P., Hussain, F. S., Amey, J. et al. (1999). The structure of vitellogenin provides a molecular model for the assembly and secretion of atherogenic lipoproteins. *J. Mol. Biol.* **285**, 391-408.

Marshall, R. D. (1974). The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins. *Biochem. Soc. Symp.* **1974**, 17-26.

Maurizio, A. (1950). The influence of pollen feeding and brood rearing on the length of life and physiological conditions of the honeybee. *Bee World* **31**, 9-12.

Miller, M. S., Benore-Parsons, M. and White, H. B., III (1982). Dephosphorylation of chicken riboflavin-binding protein and phosvitin decreases their uptake by oocytes. *J. Biol. Chem.* **257**, 6818-6824.

Munch, D. and Amdam, G. V. (2010). The curious case of aging plasticity in honey bees. *FEBS Lett.* **584**, 2496-2503.

Nakamura, A., Yasuda, K., Adachi, H., Sakurai, Y., Ishii, N. and Goto, S. (1999). Vitellogenin-6 is a major carbonylated protein in aged nematode, Caenorhabditis elegans. *Biochem. Biophys. Res. Commun.* **264**, 580-583.

Nasrin, N., Kaushik, V. K., Fortier, E., Wall, D., Pearson, K. J., de Cabo, R. and Bordone, L. (2009). JNK1 phosphorylates SIRT1 and promotes its enzymatic activity. *PloS One* **4**, e8414.

Nelson, C. M., Ihle, K. E., Fondrk, M. K., Page, R. E. and Amdam, G. V. (2007). The gene vitellogenin has multiple coordinating effects on social organization. *PLoS Biol.* **5**, e62.

Ohyama, Y., Kasai, K., Nomoto, H. and Inoue, Y. (1985). Frontal affinity chromatography of ovalbumin glycoasparagines on a concanavalin A-sepharose column. A quantitative study of the binding specificity of the lectin. *J. Biol. Chem.* **260**, 6882-6887.

Panchenko, A. R., Wolf, Y. I., Panchenko, L. A. and Madej, T. (2005). Evolutionary plasticity of protein families: coupling between sequence and structure variation. *Proteins* **61**, 535-544.

Piulachs, M. D., Guidugli, K. R., Barchuk, A. R., Cruz, J., Simoes, Z. L. and Belles, X. (2003). The vitellogenin of the honey bee, Apis mellifera: structural analysis of the cDNA and expression studies. *Insect Biochem. Mol. Biol.* **33**, 459-465.

Plummer, T. H., Jr. and Tarentino, A. L. (1991). Purification of the oligosaccharide-cleaving enzymes of Flavobacterium meningosepticum. *Glycobiology* **1**, 257-263.

Raag, R., Appelt, K., Xuong, N. H. and Banaszak, L. (1988). Structure of the lamprey yolk lipid-protein complex lipovitellin-phosvitin at 2.8 A resolution. *J. Mol. Biol.* **200**, 553-569.

Raikhel, A. S. and Dhadialla, T. S. (1992). Accumulation of yolk proteins in insect oocytes. *Annu. Rev. Entomol.* **37**, 217-251.
Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95-99.
Rappsilber, J., Ishihama, Y. and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663-670.
Rouille, Y., Duguay, S. J., Lund, K., Furuta, M., Gong, Q., Lipkind, G., Oliva, A. A., Jr, Chan, S. J. and Steiner, D. F. (1995). Proteolytic processing mechanisms in the biosynthesis of neuroendocrine peptides: the subtilisin-like proprotein convertases. *Front. Neuroendocrinol.* **16**, 322-361.
Sacchettini, J. C., Gordon, J. I. and Banaszak, L. J. (1989). Refined apoprotein structure of rat intestinal fatty acid binding protein produced in Escherichia coli. *Proc. Natl. Acad. Sci. USA* **86**, 7736-7740.
Seehuus, S. C., Norberg, K., Gimsa, U., Krekling, T. and Amdam, G. V. (2006). Reproductive protein protects functionally sterile honey bee workers from oxidative stress. *Proc. Natl. Acad. Sci. USA* **103**, 962-967.
Seehuus, S. C., Norberg, K., Krekling, T., Fondrk, K. and Amdam, G. V. (2007). Immunogold localization of vitellogenin in the ovaries, hypopharyngeal glands and head fat bodies of honeybee workers, Apis mellifera. *J. Insect Sci.* **7**, 1-14.
Shi, X., Zhang, S. and Pang, Q. (2006). Vitellogenin is a novel player in defense reactions. *Fish Shellfish Immunol.* **20**, 769-772.
Smedal, B., Brynem, M., Kreibich, C. D. and Amdam, G. V. (2009). Brood pheromone suppresses physiology of extreme longevity in honeybees (Apis mellifera). *J. Exp. Biol.* **212**, 3795-3801.
Spieth, J., Nettleton, M., Zucker-Aprison, E., Lea, K. and Blumenthal, T. (1991). Vitellogenin motifs conserved in nematodes and vertebrates. *J. Mol. Evol.* **32**, 429-438.
Steen, H., Jebanathirajah, J. A., Rush, J., Morrice, N. and Kirschner, M. W. (2006). Phosphorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements. *Mol. Cell. Proteomics* **5**, 172-181.
Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.
Thompson, J. R. and Banaszak, L. J. (2002). Lipid-protein interactions in lipovitellin. *Biochemistry* **41**, 9398-9409.
Tufail, M. and Takeda, M. (2002). Vitellogenin of the cockroach, Leucophaea maderae: nucleotide sequence, structure and analysis of processing in the fat body and oocytes. *Insect Biochem. Mol. Biol.* **32**, 1469-1476.
Tufail, M. and Takeda, M. (2008). Molecular characteristics of insect vitellogenins. *J. Insect. Physiol.* **54**, 1447-1458.
Walter, J., Schindzielorz, A., Grunberg, J. and Haass, C. (1999). Phosphorylation of presenilin-2 regulates its cleavage by caspases and retards progression of apoptosis. *Proc. Natl. Acad. Sci. USA* **96**, 1391-1396.
Wheeler, D. E. and Kawooya, J. K. (1990). Purification and characterization of honey bee vitellogenin. *Arch. Insect Biochem. Physiol.* **14**, 253-267.
Williams, S. L. and McCammon, J. A. (2009). Conformational dynamics of the flexible catalytic loop in Mycobacterium tuberculosis 1-deoxy-D-xylulose 5-phosphate reductoisomerase. *Chem. Biol. Drug. Des.* **73**, 26-38.
Williams, S. L. and Essex, J. W. (2009). Study of the conformational dynamics of the catalytic loop of WT and G140A/G149A HIV-1 integrase core domain using reversible digitally filtered molecular dynamics. *J. Chem. Theory Comput.* **5**, 411-421.
Xu, W., Yuan, X., Beebe, K., Xiang, Z. and Neckers, L. (2007). Loss of Hsp90 association up-regulates Src-dependent ErbB2 activity. *Mol. Cell. Biol.* **27**, 220-228.

Supporting information

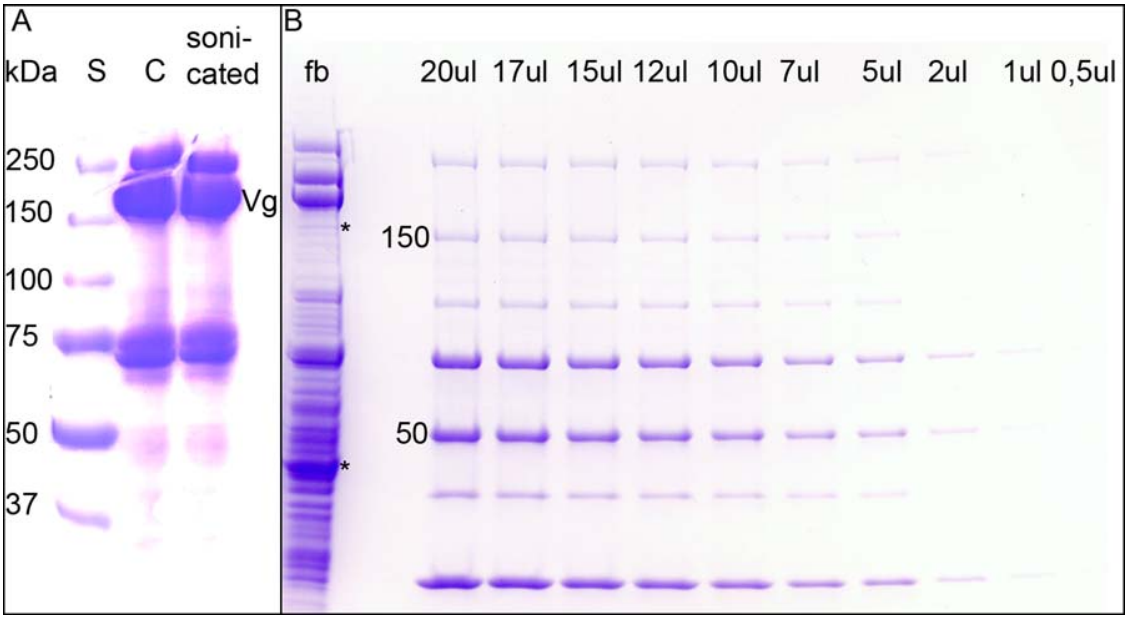


Figure S1: Effect of physical stress on vitellogenin (A) and indication of excess of 40 kDa vitellogenin compared to 150 vitellogenin fragment (B). (A) The effect of physical stress, ie sonication and vortexing, on whole length honey bee vitellogenin (180 kDa) was tested for investigating the possibility that the abdominal 40 kDa vitellogenin is a by-product of the sample treatment. The test was done by vortexing and sonicating a hemolymph sample – containing 180 kDa vitellogenin only – similarly to abdominal extract: 30 µl TBS buffer containing 0.3 µl haemolymph was added to 1.5 ml 20 mM HEPES buffer (pH 7) with 0.2 M NaCl and protease inhibitors (minipills; Roche). The sample was strongly vortexed and then sonicated (10 watts, 5 s, 10 times) with Sonics Vibra Cell ultrasonic processor and a 2 mm probe on ice. The control sample (C) was not vortexed or sonicated, but was stored on ice during the treatments. No difference between the treated sample and the control can be seen. There is no indication of physical stress induced decomposition of honey bee vitellogenin. (B) An estimation of higher intensity of the 40 kDa gel band is based on SDS-PAGE gels with fatbody extracts and standard proteins of known concentration (see material and methods). 150 and 40 kDa vitellogenin bands are indicated with an asterisk in the fatbody protein extract (fb), and the standard bands of 150 and 50 kDa are indicated. The Precision Plus Protein Standard (Bio-Rad) contains 0.21 µg/µl and 0.75 µg/µl of 150 kDa and 50 kDa standard proteins, respectively. The concentration of the other standard proteins is: 0.36 µg/µl (250 kDa), 0.14 µg/µl (100 kDa), 0.66 µg/µl (75 kDa), 0.22 µg/µl (37 kDa) and 0.8 µl/µl (25 kDa).

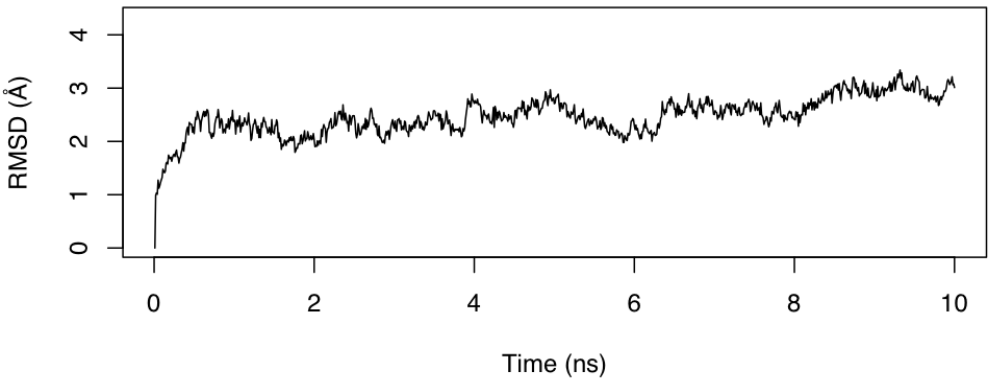


Figure S2: Structural dynamics of honeybee vitellogenin model at the N-terminus. Root mean squared deviations of the major secondary structure elements in honey bee vitellogenin relative to inital minimized model (based on lamprey lipovitellin, PDB-ID: 1lsh) against time are shown. The calculations are based upon 10 ns long all-atom molecular dynamics simulation.

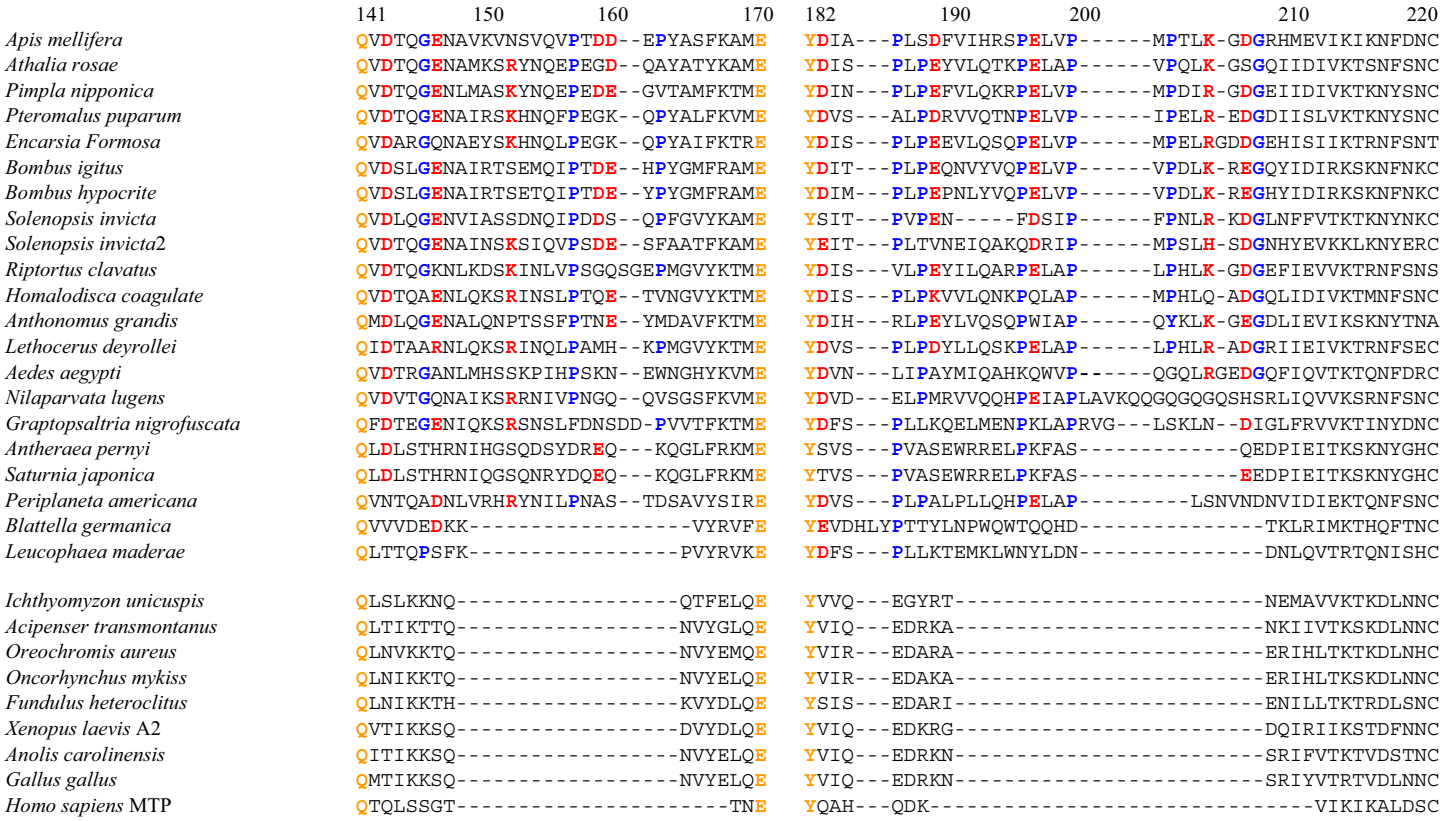


Figure S3: Alignment of insect specific loops of vitellogenin. The alignment comprises the loop sequences missing in the honey bee vitellogenin N-sheet model (V152-E161 and F190-G208). The uppermost line of numbers corresponds to the honey bee vitellogenin sequence. Orange color refers to ultra-conserved amino acid residues. Conserved glycine and proline residues of the loops are shown as blue, and conserved charged residues are shown as red. If a species has multiple vitellogenin genes, only vitellogenin-1 is presented, if not stated otherwise after the species name. The first 21 species are insects and the last 9 species are vertebrates.

Paper II

Vitellogenin in honey bee behavior and lifespan

Amdam, G. V., Fennern, E., Havukainen, H.
Honey bee Neurobiology and Behavior, Springer. Chapter 1.2. (2011)

Abstract

Vitellogenin is a phospholipoglycoprotein that affects multiple aspects of honey bee life-history. Across the vast majority of oviparous taxa, vitellogenins are female-specific egg yolk proteins, with their essential function tied to oogenesis. In honey bees, however, vitellogenin is also expressed by female helpers, called workers, which are largely sterile. Here, vitellogenin influences behavior and stress resilience, and is believed to be important to honey bee social organization. Together with longtime collaborators, we have discovered roles of vitellogenin in worker behavioral traits such as nursing, foraging onset and foraging bias, and in survival traits such as oxidative stress resilience, cell-based immunity, and longevity. We have also identified a mutually inhibitory interaction between vitellogenin and the systemic endocrine factor juvenile hormone (JH), which is central to insect reproduction and stress response. This regulatory feedback loop has spurred hypotheses on how vitellogenin and JH together have become key life-history regulators in honey bees. A current research focus is on how this feedback loop is tied to nutrient-sensing insulin/insulin-like signaling that can govern expression of phenotypic plasticity. Here, we summarize this body of work in the context of new structural speculations that can lead to a modern understanding of vitellogenin function.

Chapter 1.2
Vitellogenin in Honey Bee Behavior and Lifespan

Gro V. Amdam, Erin Fennern, and Heli Havukainen

Abstract Vitellogenin is a phospholipoglycoprotein that affects multiple aspects of honey bee life-history. Across the vast majority of oviparous taxa, vitellogenins are female-specific egg yolk proteins, with their essential function tied to oogenesis. In honey bees, however, vitellogenin is also expressed by female helpers, called workers, which are largely sterile. Here, vitellogenin influences behavior and stress resilience, and is believed to be important to honey bee social organization. Together with longtime collaborators, we have discovered roles of vitellogenin in worker behavioral traits such as nursing, foraging onset and foraging bias, and in survival traits such as oxidative stress resilience, cell-based immunity, and longevity. We have also identified a mutually inhibitory interaction between vitellogenin and the systemic endocrine factor juvenile hormone (JH), which is central to insect reproduction and stress response. This regulatory feedback loop has spurred hypotheses on how vitellogenin and JH together have become key life-history regulators in honey bees. A current research focus is on how this feedback loop is tied to nutrient-sensing insulin/insulin-like signaling that can govern expression of phenotypic plasticity. Here, we summarize this body of work in the context of new structural speculations that can lead to a modern understanding of vitellogenin function.

G.V. Amdam (✉)
School of Life Sciences, Arizona State University, PO Box 874501, Tempe, AZ 85287, USA

Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, PO Box 5002, N-1432 Aas, Norway
e-mail: Gro.Amdam@asu.edu

E. Fennern
School of Life Sciences, Arizona State University, PO Box 874501, Tempe, AZ 85287, USA

H. Havukainen
Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, PO Box 5002, N-1432, Aas, Norway

G.C. Galizia et al. (eds.), *Honeybee Neurobiology and Behavior: A Tribute to Randolf Menzel*, DOI 10.1007/978-94-007-2099-2_1.2,

Abbreviations

DRH	Double repressor hypothesis
JH	Juvenile hormone
RGPH	Reproductive ground plan hypothesis
Vg	Vitellogenin
vWFD	von Willebrand factor D-type

1.2.1 Introduction

In colonies of social insects, groups of individuals perform distinct tasks. What factors determine this differentiation of behavior? How can the regulation of social behavior be studied and understood? Here we look at the influence of vitellogenin, a multifunctional yolk precursor protein, on honey bee (*Apis mellifera*) social organization. The honey bee is one of the most important and well-researched models for the study of social behavior [31, 45]. Honey bees allow us to ask how complex social organization emerges without centralized control – through the summation of the behavioral interactions between individuals. Within each colony of insects, these interactions produce a structured division of labor that is correlated with age and associated with changes in physiology and lifespan.

Most honey bee eggs develop into essentially sterile helper females called 'workers'. Colonies have about 10,000–40,000 workers that show complex division of labor between several behavioral groups, the most important being nurse bees that take care of brood (eggs, larvae, and pupae) and foragers that collect resources from the external environment: nectar, pollen, propolis (resin) or water [37]. Worker division of labor emerges as an association between age and behavior: young bees are most often nurses while older workers forage. However, ontogeny is flexible. Foragers can revert to within-nest tasks while nurses can accelerate behavioral development and forage precociously [26]. Once foragers, workers usually die within 7–10 days, but foraging efforts decline if no brood is present requiring care such as feeding. During periods without brood rearing, workers develop into stress-resistant '*diutinus* bees' with lifespans of 280 days or more. This phenotypic plasticity produces different life-histories with worker lifespans ranging from a few weeks to nearly a year (review: [6]).

Variation in worker behavior and lifespan correlates with vitellogenin (review: [6]). In a highly regulated manner, this protein is expressed during development and adult life by both sexes and all behavioral groups of honey bees. The primary site of synthesis is the abdominal fat body (functional homologue of vertebrate liver and adipose tissue), which lines the body-wall as a single cell-layer composed of trophocytes and oenocytes. From fat body, vitellogenin is secreted into the hemolymph (blood). This circulating vitellogenin titer is highest in queens, and lowest in males [21]. In workers, vitellogenin synthesis and hemolymph levels vary with social role and longevity: the short-lived foragers produce substantially less vitellogenin and

have significantly lower titers than nurse bees, while the stress resilient *diutinus* bees have the highest levels of all the worker groups (see [3] for a recent review).

The expression of the *vitellogenin* gene is nutrient sensitive, and responds positively to the level of amino acids in hemolymph. Along with many genes and endocrine factors, such as protein-kinase G [13], *malvolio* that encodes a manganese transporter [12], and the systemic JH (review: [16]), *vitellogenin* has been linked to food-related worker behavior – specifically to the onset of foraging. In addition, vitellogenin takes part in brood-food synthesis in nurse bees, and it also affects the bees' propensity to collect pollen rather than nectar during foraging trips [32] (see also Chap. 1.1).

Added to the food-related effects, vitellogenin has an independent and positive influence on worker lifespan [32]. This effect may be explained by a nutritive role of the vitellogenin protein [21] and its positive influence on oxidative stress resilience and cell-based immunity [36]. Thereby, vitellogenin is among the most multifunctional life-history regulators known in honey bees, and is likely to be instrumental for colony social organization.

With this chapter, we summarize what is currently known about honey bee vitellogenin and its effects on worker life-histories. We outline mechanisms that may allow vitellogenin to influence worker phenotypic outcomes, and discuss how this protein has become such a central regulatory element during honey bee social evolution.

1.2.2 Vitellogenin Properties

Some effects of vitellogenin in honey bees are surprising (Fig. 1.2.1). Vitellogenins are female-specific yolk proteins that are central to egg development in most oviparous invertebrate and vertebrate animals. Worker honey bees, however, do not normally lay eggs. They have lost the ability to mate, their reproductive organs are greatly reduced, and further oocyte development is blocked by pheromonal inhibition (see [38] for a review).

1.2.2.1 Gene Sequence and Protein Structure

The honey bee *vitellogenin* gene (GenBank accession number: CAD56944.1) is 5,440 bp long and yields a polypeptide of 1,770 amino acids [33]. Many taxa, including *Xenopus* and *Caenorhabditis*, have multiple *vitellogenin* genes [35, 43], but only one is found in the honey bee genome. The vitellogenin amino acid sequence, as a whole, is only weakly conserved between species, e.g., the sequence similarity between honey bee and bumblebee (*Bombus ignitus*) vitellogenin is only 51%. However, vitellogenins contain conserved elements, including a GL/ICG (Glysine-Leucine/Isoleucine-Cysteine-Glycine) amino acid motif followed by nine cysteine residues near the C-terminus in all available insects sequences so far [33].

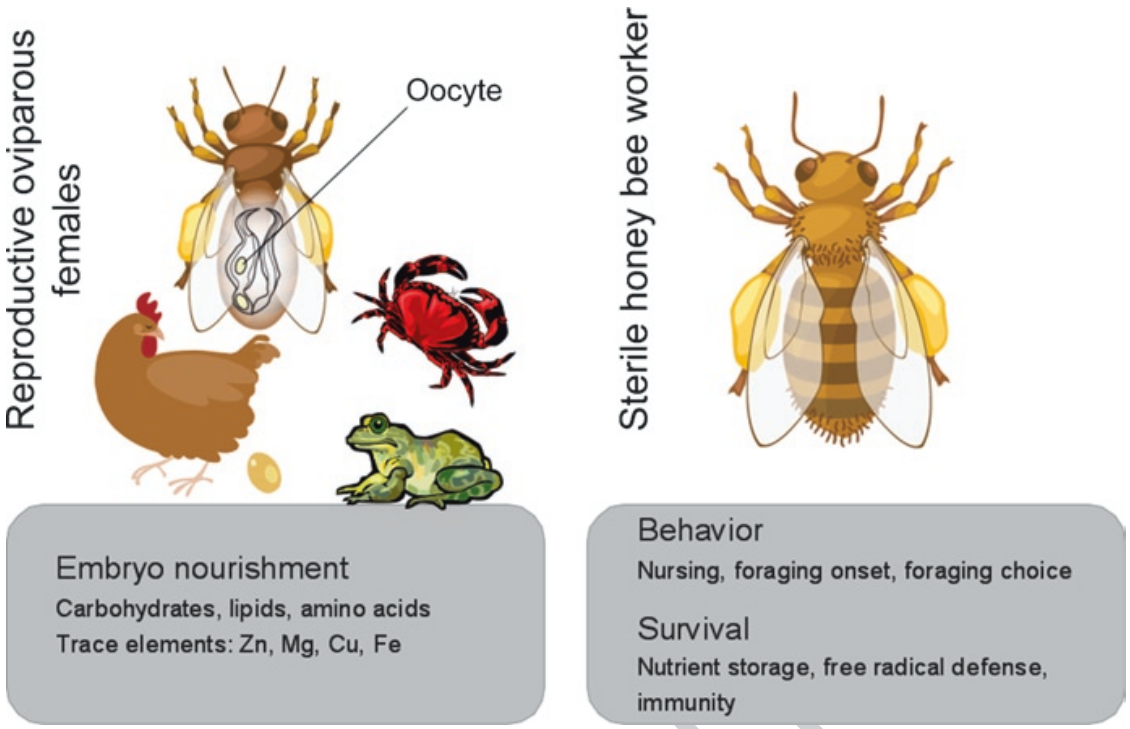


Fig. 1.2.1 The general roles of vitellogenin proteins in oviparous females of a variety of taxa versus honey bee vitellogenin functions in workers, which are largely sterile. In most oviparous taxa (some examples given in the *left panel*), vitellogenins are yolk precursors that are synthesized in liver (vertebrates) or fat body (invertebrates) for secretion into the blood. From circulation, vitellogenin protein is taken up by the ovary for transfer to the developing oocytes where it nourishes the embryo with macronutrients as well as zinc (Zn), magnesium (Mg), copper (Cu) or iron (Fe, as noted in the *gray box*, *lower left*). A bee (other than honey bee) is indicated with transparent abdomen to point out vitellogenic oocytes in the ovary. This process of vitellogenesis occurs in honey bee queens and in worker bees that are not inhibited from egg-laying. In worker bees, furthermore, vitellogenin influences behavior and survival (*right panel*, details on affected behavioral traits and mechanism of somatic maintenance in the corresponding *gray box*, *lower right*). Behavioral effects have been traced to workers, but the effect on longevity and survival is shared with queens [16, 36]

Honey bee vitellogenin is synthesized as a monomeric 180 kDa phospholipoglycoprotein [46]. The protein consists of several parts: N-sheet, polyserine linker and a lipid cavity that includes an α-helical region and a vWFD (von Willebrand factor, type D) domain (Fig. 1.2.2a). A secondary structure can be tentatively estimated based on *in silico* prediction (PSIPRED; [27]). For the N-sheet and α-helical domain, furthermore, the crystal structure of lamprey lipovitellin can be used as a guide [34]. The lamprey lipovitellin crystal structure (Fig. 1.2.2b) is currently the only solved structure from the vitellogenin protein family.

A region of the N-sheet domain may confer binding between vitellogenin and its receptor in *Oreochromis aureus* [29], but this domain-association is not confirmed in other species. The N-sheet is connected to the rest of vitellogenin by a putative phosphorylated linker. This polyserine linker is not present in vertebrates and shows diversity in insects for location and length [14]. The role of the α-helical domain is unidentified, but may bind zinc in lamprey [10]. The vWFD domain has cysteine-residues conserved in insects [42] but its structure and function is unknown.

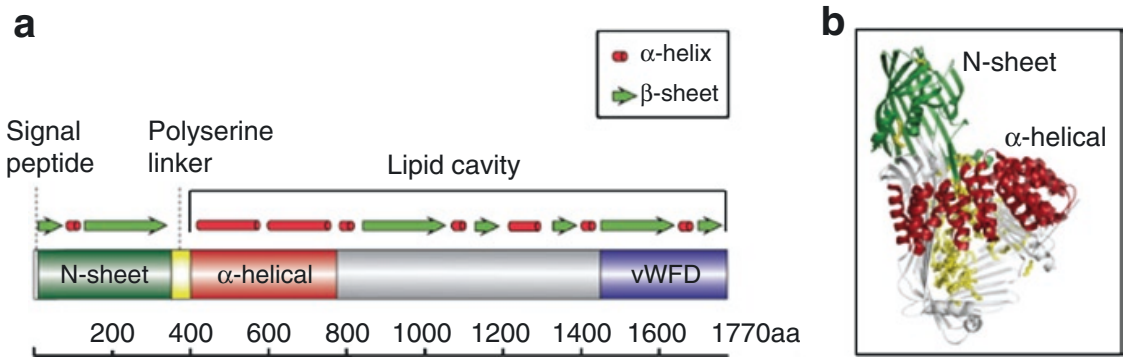


Fig. 1.2.2 Approximation of honey bee vitellogenin structure. (**a**) The amino terminal N-sheet domain starts after a 16 amino acid long signal sequence. This sheet is connected to the rest of vitellogenin by a polyserine linker. Based on the primary structure, the serine-rich linker is a large and flexible loop. The C-terminal region subsequent to the polyserine region forms a big lipid-carrying cavity, which can be approximated from a representation of the lamprey lipovitellin crystal structure. The C-terminal region of honey bee vitellogenin has two clear domains: an α-helical region (estimated to consist of total 16 helixes) and the vWFD-type domain. (**b**) Lamprey lipovitellin, shown as a monomer, is prepared with Pymol [19] based on its Protein Data Bank entry (ID: 1lsh). The β-barrel-like N-sheet (*green*) is estimated to resemble the honey bee N-sheet. The α-helical domain (*red*) forms an arch on the large β-sheets of the rest of the funnel-like lipid cavity. Lipids found in the crystal structure are depicted as yellow sticks. The solved structure lacks the vWFD domain, and the polyserine linker is missing

1.2.2.2 Vitellogenin Expression, Accumulation, Behavioral Correlation

In honey bees, most *vitellogenin*-expressing fat body cells are localized to the abdomen with some cells dispersed in the thorax and head [16, 39]. Vitellogenin synthesis is largely specific to trophocytes, which constitute one of two major cell types in fat body. Some transcript is also localized to queen ovarian tissue [24]. In most species, the vitellogenin pre-protein is cleaved, typically into a small N-terminal and a large C-terminal piece [42]. Honey bee vitellogenin lacks the conserved cleavage site found near the polyserine region of other insects, but appears as at least three forms in patterns that are tissue/compartment specific. The protein is detected as 180, 150 and 40 kDa in worker fatbody [25] as 180 kDa in hemolymph; and as 180 and 150 kDa in hypopharyngeal glands, which are the food-producing head glands of workers [5]. The 40 kDa fragment corresponds to the N-sheet of vitellogenin, and the 150 kDa fragment corresponds to the lipid cavity (Fig. 1.2.2a, b; [25]). The site of cleavage and the biological significance of these fragment patterns are unidentified.

The levels of mature vitellogenin protein in fat body and hemolymph change as a function of worker life-history progression. The rate of synthesis is enhanced soon after adult emergence from the pupal stage, and blood levels increase. During ambient conditions with active foraging by colonies, the vitellogenin level peaks in 7–10 day old nurse bees, and subsequently drops at foraging onset (see [3] and references therein). In nurse bees, vitellogenin provides amino acids and perhaps other nutrient building blocks for food synthesis or 'jelly production' by the hypopharyngeal head

glands [5]. Jelly is central to the nourishment of larvae and the queen, and is also fed to other adult workers including foragers [17]. During unfavorable periods, larval rearing and foraging decline, and nurse bees accumulate very large amounts of vitellogenin in the fat body and hemolymph. This physiology is the hallmark of the long-lived and stress-resilient *diutinus* worker bees [36].

1.2.2.3 Experimental Manipulation, Behavior, and Frailty

The *vitellogenin* gene can be experimentally suppressed by RNA interference (RNAi) [8]. This *vitellogenin* gene silencing causes a drop in the circulating level of vitellogenin protein, resulting in cessation of nursing behavior followed by precocious foraging onset relative to controls [30, 32]. These results demonstrate that vitellogenin can slow worker behavioral progression, and may explain the functional significance of reduced vitellogenin levels that co-occur with the natural nurse-to-forager transition of worker bees [6, 32]. Moreover, knockdowns bias their foraging efforts toward nectar (carbohydrate source) rather than pollen (protein and lipid source) [32]. In other words, vitellogenin also influences worker food-choice behavior and presumably biases a bee's foraging effort toward pollen (see Chap. 1.1).

In addition to these effects on behavior, the RNAi-mediated repression of vitellogenin shortens worker lifespan [32, 36]. This outcome is in part conferred by the precocious foraging onset of knockdowns, since foragers rapidly perish [32]. Longevity, however, is also shortened independent of behavior. This influence is attributed to a positive effect of vitellogenin on resilience to stressors such as oxidative metabolic damage, starvation, and immune challenge [32, 36].

In summary, the drop in vitellogenin levels that generally occurs during worker adult ontogeny is associated with increased frailty as well as behavioral change. These relationships are directly supported by experimental repression of *vitellogenin* activity.

1.2.3 Vitellogenin Functions – Hypotheses and Molecular Mechanisms

How can vitellogenin influence behavior and lifespan? The traditional view of vitellogenins as yolk precursors contrasts with the modern understanding of honey bee vitellogenin as a protein that broadly affects social life-histories [3].

1.2.3.1 Proximate and Ultimate Explanations

To explain the effect of vitellogenin on foraging onset, Amdam and Omholt [6] proposed a double repressor model. This model entails a mutually inhibitory feedback loop between vitellogenin and the life-shortening systemic endocrine

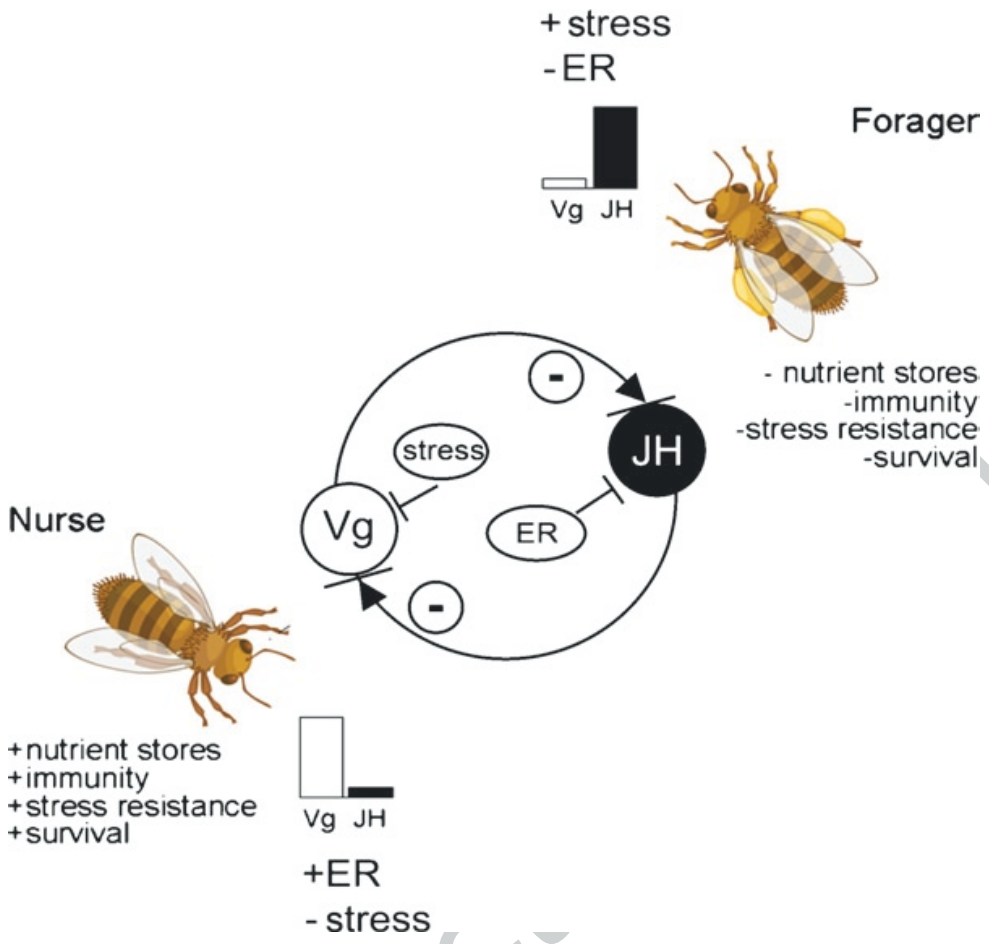


Fig. 1.2.3 The double repressor model (DRH). In nurse bees (*left*), vitellogenin (Vg, level indicated with white bars) and the presence (+) of an external repressor pheromone (+ER) (|— ER) inhibit juvenile hormone (JH, level indicated with black bars) and foraging activity in the absence of stress (show as minus (–) stress). If the ER signal is reduced (minus ER, –ER), workers with low Vg titers are at greater risk of JH signaling and foraging. Stress (+Stress) reduces Vg (|— Stress) and elicits foraging through JH signaling that is a conserved element of the insect stress response. Vg synthesis is suppressed by JH, closing a feedback loop between Vg and JH that locks the worker into the forager state (*right*). The feedback between Vg and JH is indicated with mutual repressor arrows (*gray*) and correspondence with increased vs. reduced levels of stress and ER is shown alongside these *arrows*

factor JH, which affects physiology and behavior in many insects via regulatory roles in reproduction and stress response. According to this hypothesis, honey bee nurse stage is governed by vitellogenin while JH is high in foragers and has pleiotropic effects on behavioral physiology that confers the depletion of nutrient stores and reduced somatic maintenances (Fig. 1.2.3). The model proposed that foraging onset initiates when a decline in vitellogenin (reducing nutrient concentration and stress resilience) allows JH to increase. JH feeds back to suppress vitellogenin further – in effect locking the bee into the forager state [6].

In forming the double repressor hypothesis (DRH), Amdam and Omholt proposed vitellogenin as an *internal* repressor that could slow the nurse bee-to-forager transition of bees. In reference to 'double' repression, the effect of vitellogenin joined an existing explanation where foraging onset was delayed by an *external* repressor

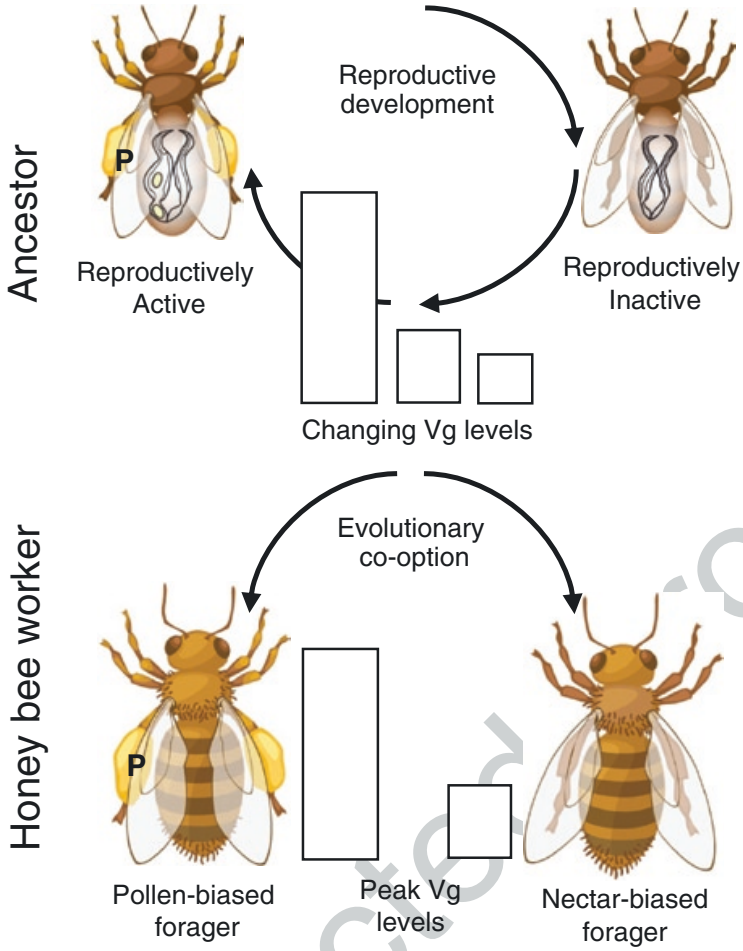


Fig. 1.2.4 The reproductive ground plan hypothesis (RGPH) outlines how the female reproductive biology could have been co-opted in the division of labour between foraging worker honey bees. From the *top*: In ancestors of honey bees, the sequential progression of the reproductive cycle was linked to changes in female food-related behavior. During periods of no active reproduction (*top*, *right*), the ovary was undeveloped (not enlarged with yolk) and the Vg titre (*white bars*) was low. During reproduction (*top left*), *vg* gene expression increased prior to yolk deposition, and pollen (P in bee, *upper left*) was required for nest provisioning. *Below*: Vg and foraging behavior is similarly associated in worker honey bees, where the Vg peak titer has a positive and seemingly direct influence on pollen collection [32]

contact pheromone. This repressor contains ethyl oleate, and appears to be produced by foragers to inhibit recruitment of nurses to foraging tasks [28]. The DRH also provided reasoning for adaptive division of labor between nurse bees and foragers. An internal repressor like vitellogenin conferred a selective advantage for honey bee societies, because it retained nutrient rich and healthy nurse bees in the nest while vitellogenin-poor or strained workers (less useful in nursing) would be first to respond with a shift to foraging, should levels of external repressor drop [6].

The role of vitellogenin in foragers' preference for pollen or nectar was addressed by Amdam and Page's reproductive ground plan hypothesis, or RGPH [4]. While the DRH describes the regulation of foraging onset, the RGPH is a broader hypothesis that seeks to explain what gene networks natural selection might have acted upon during honey bee social evolution (Fig. 1.2.4). This framework views the

relationship between vitellogenin and pollen foraging preference in honey bees as a 'footprint' of ancestral trait-associations. The RGPH was inspired by correlative relationships between yolk production, egg development, and division of labor in several species of wasps, ants, and bees (see [4] and references therein). Most solitary bees, moreover, feed on nectar and pollen while pollen-collecting (hoarding) behavior is specific to the stage of larval nest-provisioning that is evident during vitellogenesis (references in [4]). Similarly, high levels of vitellogenin are associated with pollen foraging in worker honey bees that – like solitary females – collect pollen for storage in the brood nest. The RGPH suggests that natural selection exploited the reproductive genetics of female bees to facilitate advanced division of labour between workers with different foraging biases.

The frailty of foragers compared to nurse worker bees, at least in part resulting from declining vitellogenin levels at foraging onset, could be advantageous at a colony-level [7]. By reducing vitellogenin, workers attain a physiology of high JH, reduced cell-based immunity, and increased oxidative stress susceptibility [32, 36]. Foragers can be exposed to a heavy load of pathogens in the field, and will return to the same source of food over the course of days. By being susceptible to substances that could harm the colony, foragers die before bringing much contamination to the nest. Likewise, stress by handling, injury, and disease causes JH to increase and bees to initiate foraging (references in [2]). The regulatory connection between vitellogenin, JH, and stress thereby ensures that bees in poor condition transition to a state where frailty and extrinsic mortality pretty much guarantee rapid death at a safe distance from the colony [7].

1.2.3.2 Putative Molecular Action

The pleiotropic effects of vitellogenin on behavior and stress resilience make sense in theory; however, the structural and regulatory aspects of the underlying processes are not well understood. To date, one structural mechanism has been proposed for the role of vitellogenin in oxidative stress resistance and cell-based immunity – effects that might emerge from zinc-binding properties [9, 36]. Also, a regulatory association between vitellogenin and insulin/insulin-like signaling has been hypothesized to explain the pleiotropic effect of vitellogenin on JH, behavior and lifespan [16, 36]. Insulin/insulin-like signaling governs responses to food-intake in eukaryotes, including eating behavior and metabolic physiology related to lifespan [11, 18]. In insect model systems, inhibition of insulin/insulin-like signaling (reducing nutrient sensing) will thereby lower JH and boost survival capacity [22, 41]. In honey bees, on the other hand, this relationship can be reversed, since increased nutrient status (vitellogenin) can confer low JH and increased survival [6, 32, 36]. It is envisioned that this inversion results from an ability of vitellogenin to suppress insulin/insulin like signals (review: [2]).This inhibitory function, however, is not yet validated experimentally.

Current knowledge on the basic sequence and structure of vitellogenin (Fig. 1.2.2) suggests that phosphorylation, cleavage, or ligand binding can format vitellogenin

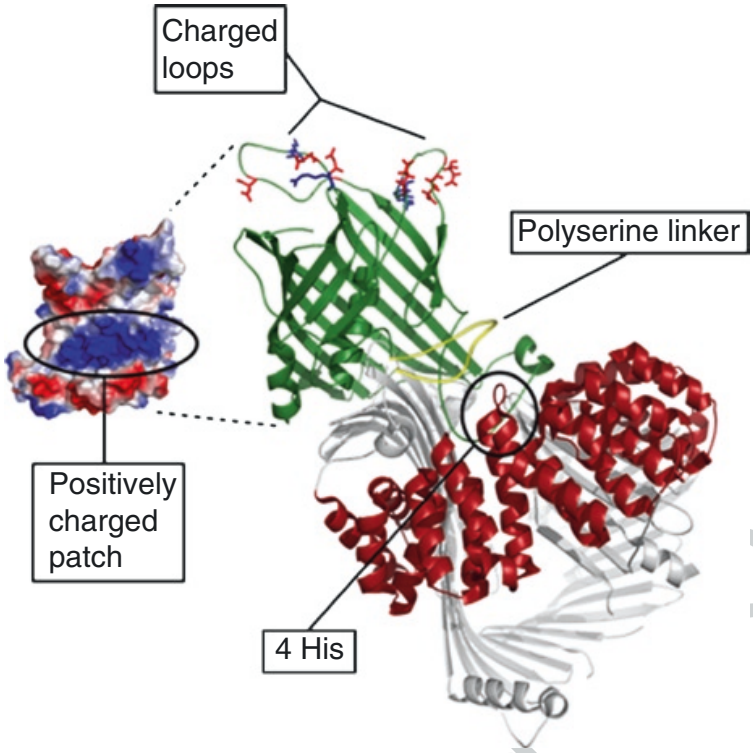


Fig. 1.2.5 A preliminary view of structures of honey bee vitellogenin. The model combines the lamprey crystal structure with a homology model of the vitellogenin N-sheet (*green*) and by adding a sketch of the polyserine linker (*yellow*) between the N-sheet and the lipid cavity. The polyserine loop was oriented randomly. Unlike the lamprey N-sheet, honey bee vitellogenin has two loops (Charged loops) whose charged residues are visualized as *red* (negative) and *blue* (positive) sticks. An electrostatic surface potential map (*left side* of the main structure) of the honey bee vitellogenin N-sheet model was prepared with Pymol [19]. The map visualizes a positively charged patch not present in lamprey lipovitellin, formed by residues of the β-sheets. In the middle of the α-helical region, four residues of histidine – a zinc coordinator – are found close to each other in sequence (approximation location *circled*)

for different functions. Building on this information and exploiting the crystal structure of lipovitellin [34], we can identify motifs with potential to serve as regulatory sites (Fig. 1.2.5).

One of the most striking structural elements in honey bee vitellogenin is the polyserine linker that offers multiple potential phosphoryl-acceptor sites (Fig. 1.2.2a and 1.2.5). The location of the polyserine linker is ideal for structural speculation: the linker separates the loosely attached N-sheet from the rest of the molecule. It is unknown how cleavage of the N-sheet is achieved, but this can be a phosphorylation-regulated process, such as in the case of presenilin-2 (cleavage inhibited by phosphorylation [44]) or cohesin (cleavage triggered by phosphorylation, [1]). It is well-documented that phosphorylation-induced structural changes can have effects on signaling, activation or deactivation of enzymes, and degradation, localization and binding (review: [15]).

The N-sheet contains several structurally interesting features. There are two insect-specific loops with a considerable number of conserved charged residues (Fig. 1.2.5; 'Charged loops'). A loop is not necessarily a passive linker, but can play

an intrinsic structural role, e.g. [20]. In some cases, even active site residues are located in a loop, e.g. [47]. Also, there is a positively charged patch on the β-sheets of the N-sheet of honey bee vitellogenin (Fig. 1.2.5). Charged patches typically attract negatively charged binding partners (for recent data, see [23]). The N-sheet and the polyserine linker of honey bee vitellogenin offer possibilities for regulation: phosphorylation, cleavage, flexible regions for conformational changes, and an attractive charge for binding. The vWFD and α-helical domains provide opportunity for metal binding. A proximate understanding of vitellogenin function will require targeted studies of structure and dynamic action at these sites.

1.2.4 Outlook

It is an open question how vitellogenin, and its associated molecular processes, can impact the development or function of the central nervous system to influence honey bee behavior. Vitellogenin and JH are intimately connected, and there is considerable evidence linking JH to insect brain development and more recently to the circadian cycle of insects (see [40] and references therein). Yet it is unclear, for example, whether vitellogenin levels vary during the circadian cycle of the bee, as well as whether and how receptor molecules are required for vitellogenin function.

In future, we need to ask how a modern understanding of vitellogenin can apply to processes of the honey bee brain. Perhaps molecular developmental queues associated with JH and vitellogenin, that are essential for normal holometabolic and reproductive development, were fine-tuned to accommodate more subtle development shifts required by the adult brain. Part of the answer may be that cells/neurons require development queues for a burst of neurite outgrowth during the first 2 weeks of adult brain development, and may be part of a program of worker behavioral progression.

References

1. Alexandru G, Uhlmann F, Mechtler K, Poupart MA, Nasmyth K (2001) Phosphorylation of the cohesin subunit Scc1 by Polo/Cdc5 kinase regulates sister chromatid separation in yeast. Cell 105(4):459–472
[AU1] 2. Amdam GV (2011) Social context, stress, and plasticity of aging. Aging Cell 10(1):18–27
3. Amdam GV, Ihle KE, Page RE (2009) Regulation of honey bee (*Apis mellifera*) life histories by vitellogenin. In: Pfaff D, Arnold A, Etgen A, Fahrbach S, Rubin R (eds) Hormones, brain and behavior, 2nd edn. Elsevier Academic Press, San Diego
4. Amdam GV, Norberg K, Fondrk MK, Page RE (2004) Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. Proc Natl Acad Sci USA 101:11350–11355
5. Amdam GV, Norberg K, Hagen A, Omholt SW (2003) Social exploitation of vitellogenin. Proc Natl Acad Sci USA 100:1799–1802
6. Amdam GV, Omholt SW (2003) The hive bee to forager transition in honeybee colonies: the double repressor hypothesis. J Theor Biol 223:451–464

7. Amdam GV, Seehuus SC (2006) Order, disorder, death: lessons from a superorganism. Adv Cancer Res 95:31–60
8. Amdam GV, Simões ZLP, Guidugli KR, Norberg K, Omholt SW (2003) Disruption of vitellogenin gene function in adult honeybees by intra-abdominal injection of double-stranded RNA. BMC Biotechnol 3:1–8
9. Amdam GV, Simões ZLP, Hagen A, Norberg K, Schrøder K et al (2004) Hormonal control of the yolk precursor vitellogenin regulates immune function and longevity in honeybees. Exp Gerontol 39:767–773
10. Anderson TA, Levitt DG, Banaszak LJ (1998) The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. Struct Fold Des 6(7):895–909
11. Bartke A (2005) Minireview: role of the growth hormone/insulin-like growth factor system in mammalian aging. Endocrinology 146(9):3718–3723
12. Ben-Shahar Y, Dudek NL, Robinson GE (2004) Phenotypic deconstruction reveals involvement of manganese transporter malvolio in honey bee division of labor. J Exp Biol 207:3281–3288
13. Ben-Shahar Y, Robichon A, Sokolowski MB, Robinson GE (2002) Influence of gene action across different time scales on behavior. Science 296:741–744
14. Chen JS, Sappington TW, Raikhel AS (1997) Extensive sequence conservation among insect, nematode, and vertebrate vitellogenins reveals ancient common ancestry. J Mol Evol 44(4):440–451
15. Cohen P (2000) The regulation of protein function by multisite phosphorylation – a 25 year update. Trends Biochem Sci 25(12):596–601
16. Corona M, Velarde RA, Remolina S, Moran-Lauter A, Wang Y et al (2007) Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. Proc Natl Acad Sci USA 104:7128–7133
17. Crailsheim K (1990) The protein balance of the honey bee worker. Apidologie 21:417–429
18. Dallman MF, Warne JP, Foster MT, Pecoraro NC (2007) Glucocorticoids and insulin both modulate caloric intake through actions on the brain. J Physiol 583:431–436
19. DeLano WL (2002) The PyMOL molecular graphics system. De Lano Scientific, Palo Alto
20. Doucet N, Watt ED, Loria JP (2009) The flexibility of a distant loop modulates active site motion and product release in ribonuclease A. Biochemistry 48(30):7160–7168
21. Engels W, Kaatz H, Zillikens A, Simões ZLP, Truve A et al (1990) Honey bee reproduction: vitellogenin and caste-specific regulation of fertility. In: Hoshi M, Yamashita O (eds) Advances in invertebrate reproduction, vol 5. Elsevier Science Publishers B.V., Amsterdam, pp 495–502
22. Flatt T, Min KJ, D'Alterio C, Villa-Cuesta E, Cumbers J et al (2008) *Drosophila* germ-line modulation of insulin signaling and lifespan. Proc Natl Acad Sci USA 105(17):6368–6373
23. Grigg JC, Cooper JD, Cheung J, Heinrichs DE, Murphy ME (2010) The *Staphylococcus aureus* siderophore receptor HtsA undergoes localized conformational changes to enclose staphyloferrin A in an arginine-rich binding pocket. J Biol Chem 285(15):11162–11171
24. Guidugli KR, Piulachs MD, Belles X, Lourenco AP, Simões ZLP (2005) Vitellogenin expression in queen ovaries and in larvae of both sexes of *Apis mellifera*. Arch Insect Biochem Physiol 59:211–218
25. Havukainen H, Halskau Ø, Sjærven L, Amdam GV (2011) Deconstructing honeybee vitellogenin: novel 40 kDa fragment assigned to its N-terminus. J Exp Biol 214:582–592
26. Huang Z-Y, Robinson GE (1996) Regulation of honey bee division of labor by colony age demography. Behav Ecol Sociobiol 39:147–158
27. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292(2):195–202
28. Leoncini I, Le Conte Y, Costagliola G, Plettner E, Toth AL et al (2004) Regulation of behavioral maturation by a primer pheromone produced by adult worker honey bees. Proc Natl Acad Sci USA 101(50):17559–17564
29. Li A, Sadasivam M, Ding JL (2003) Receptor-ligand interaction between vitellogenin receptor (VtgR) and vitellogenin (Vtg), implications on low density lipoprotein receptor

and apolipoprotein B/E. The first three ligand-binding repeats of VtgR interact with the amino-terminal region of Vtg. J Biol Chem 278(5):2799–2806
30. Marco Antonio DS, Guidugli-Lazzarini KR, Nascimento AM, Simões ZLP, Hartfelder K (2008) RNAi-mediated silencing of *vitellogenin* gene function turns honeybee (*Apis mellifera*) workers into extremely precocious foragers. Naturwissenschaften 95:953–961
31. Menzel R, Leboulle G, Eisenhardt D (2006) Small brains, bright minds. Cell 124:237–239
32. Nelson CM, Ihle K, Amdam GV, Fondrk MK, Page RE (2007) The gene *vitellogenin* has multiple coordinating effects on social organization. PLoS Biol 5:673–677
33. Piulachs MD, Guidugli KR, Barchuk AR, Cruz J, Simões ZLP et al (2003) The vitellogenin of the honey bee, *Apis mellifera*: structural analysis of the cDNA and expression studies. Insect Biochem Mol Biol 33:459–465
34. Raag R, Appelt K, Xuong NH, Banaszak L (1988) Structure of the lamprey yolk lipid-protein complex lipovitellin-phosvitin at 2.8 A resolution. J Mol Biol 200(3):553–569
35. Rina M, Savakis C (1991) A cluster of vitellogenin genes in the Mediterranean fruit fly *Ceratitis capitata*: sequence and structural conservation in dipteran yolk proteins and their genes. Genetics 127(4):769–780
36. Seehuus SC, Norberg K, Gimsa U, Krekling T, Amdam GV (2006) Reproductive protein protects sterile honey bee workers from oxidative stress. Proc Natl Acad Sci USA 103:962–967
37. Seeley TD (1995) The wisdom of the hive. Harvard University Press, Cambridge
38. Slessor KN, Winston ML, Le Conte Y (2005) Pheromone communication in the honeybee (Apis mellifera L.). J Chem Ecol 31(11):2731–2745
39. Snodgrass RE (1956) Anatomy of the honey bee. Comstock, New York
40. Stay B, Zera AJ (2010) Morph-specific diurnal variation in allatostatin immunostaining in the corpora allata of Gryllus firmus: implications for the regulation of a morph-specific circadian rhythm for JH biosynthetic rate. J Insect Physiol 56(3):266–270
41. Tatar M, Bartke A, Antebi A (2003) The endocrine regulation of aging by insulin-like signals. Science 299:1346–1350
42. Tufail M, Takeda M (2008) Molecular characteristics of insect vitellogenins. J Insect Physiol 54(12):1447–1458
43. Wahli W, Dawid IB (1980) Isolation of two closely related vitellogenin genes, including their flanking regions, from a *Xenopus laevis* gene library. Proc Natl Acad Sci USA 77(3):1437–1441
44. Walter J, Schindzielorz A, Grunberg J, Haass C (1999) Phosphorylation of presenilin-2 regulates its cleavage by caspases and retards progression of apoptosis. Proc Natl Acad Sci USA 96(4):1391–1396
45. Weinstock GM, Robinson GE, Gibbs RA, Weinstock GM, Weinstock GM et al (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 443(7114):931–949
46. Wheeler DE, Kawooya JK (1990) Purification and characterization of honey bee vitellogenin. Arch Insect Biochem Physiol 14:253–267
47. Williams SL, Essex JW (2009) Study of the conformational dynamics of the catalytic loop of WT and G140A/G149A HIV-1 Integrase core domain using reversible digitally filtered molecular dynamics. J Chem Theor Comp 5:411–421

Paper III

Social pleiotropy and the molecular evolution of honey bee vitellogenin

Havukainen, H., Halskau, Ø., Amdam, G. V.
Mol. Ecol. Accepted manuscript, in press. (2011)

Abstract

In this issue of Molecular Ecology, Kent et al. (2011) describe the adaptive evolution of honey bee vitellogenin that belongs to a phylogenetically conserved group of egg yolk precursors. This glyco-lipoprotein leads a double life: it is central to egg production in the reproductive queen caste, and a regulator of social behavior in the sterile worker caste. Does such social pleiotropy constrain molecular evolution? To the contrary; Kent and coworkers show that the *vitellogenin* gene is under strong positive selection in honey bees. Rapid change has taken place in specific protein regions, shedding light on the evolution of novel vitellogenin functions.

Heli Havukainen[a,b], Øyvind Halskau[a,c] and Gro V. Amdam[a,d,*]

[a] Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, P.O. Box 5003 1432 Aas, Norway
[b] Department of Biomedicine, University of Bergen, Jonas Lies vei 91, 5009 Bergen, Norway
[c] Department of Molecular Biology, University of Bergen, Thormøhlensgate 55, 5020 Bergen, Norway
[d] School of Life Sciences, Arizona State University, P.O. Box 874501, Tempe, AZ 85287-4701, USA
* Corresponding author. Gro.Amdam@asu.edu

Vitellogenin was identified in honey bees about 40 years ago and its reproductive role in queens was instantly recognized (Engels, 1974). Queen bees are responsible for egg-laying, but curiously, vitellogenin was also found in considerable amounts in worker bees, which are helper females that normally are sterile. The workers' expression of vitellogenin was initially seen as evolutionary baggage; an unavoidable consequence of selection for extreme rates of vitellogenin synthesis in queens. But later studies revealed pleiotropic functions of the protein (Fig. 1). Vitellogenin can coordinate social behavior in workers, which expend the protein in larval food, and it enhances stress resistance, immunity and survival in both workers and queens (see (Nelson *et al.*, 2007; Seehuus *et al.*, 2006) and references therein). While some species have several *vitellogenin* gene copies that may accommodate caste-specific functions, the honey bee has only one *vitellogenin*. Thus, this gene is socially pleiotropic through its influence on several fitness traits that are partly separated in queen and worker castes (Kent et al., 2011).

It may be difficult to accumulate selectable mutations in a pleiotropic gene while retaining or increasing its fitness contribution to two or more traits. Kent et al. (2011) therefore set out to test whether honey bee *vitellogenin* experiences this constraint. They sequenced *vitellogenin* and seven other genes in 41 honey bee workers from Africa, East- and West Europe, where honey bee population sizes, structures and histories are known. With this knowledge base, it is unlikely that changes or differences in population sizes are misinterpreted as signatures of selection in the data. Moreover, the seven additional genes served as reference points that would be equally affected by issues of population dynamics and genetic drift. These included *Erk7* and *for*, which are genes that, similar to *vitellogenin*, are expressed in both workers and queens and implicated in worker behavior (see (Kent et al. 2011) for details and references). Sequence information from one worker from each of three other bee species added to the dataset.

The high average pair-wise nucleotide differences, linkage disequilibrium and skew in the allele frequency spectrum that Kent et al. (2011) estimated from the bees documented allelic variation for *vitellogenin* and suggested recurrent positive selection acting on this gene. Among several statistical approaches, the authors employed McDonald-Kreitman tests to compare the ratio of nucleotide substitutions that have no effect on the protein sequence (synonymous) to those that change the amino acid

sequence (non-synonymous), and flagged relative rates of non-synonymous versus synonymous divergence as positive selection. The *vitellogenin* gene showed this high relative rate in comparisons between the four bee species of the dataset; the honey bee *Apis mellifera*, Asian honey bee *A. cerana,* giant honey bee *A. dorsata*, and dwarf honey bee *A. florea*. The specific analyses of honey bee *vitellogenin*, however, revealed significant linkage disequilibrium only for European populations, and not African. We believe this new finding supports the hypothesis that changes in *vitellogenin* occurred to accommodate colony survival in colder climates during and after the prehistoric migrations of *A. mellifera* from Africa to Europe (Amdam *et al.*, 2005; Seehuus *et al.*, 2006). Overall, Kent et al. (2011) could assign a higher recent and ongoing rate of adaptive protein evolution to *vitellogenin* than the seven reference genes, and concluded that social pleiotropy does not constrain honey bee vitellogenin adaptation by selection.

This positive selection on honey bee vitellogenin begs the question of how genetic changes translate to the level of protein structure and function. Vitellogenin is large (180 kDa) and has a complex domain structure. Are there degrees of freedom to sculpt parts of this molecule for novel functions while retaining its conserved role in yolk production? There are examples of much smaller proteins evolving several new features under considerable constraints, e.g. alpha-lactabumin (a milk protein) evolved from lysozyme into a substrate modifier and secretory nutritional protein while retaining its overall structure (Qasba, Kumar, 1997). Kent et al. (2011) discuss 64 single nucleotide polymorphisms (SNPs) that are unequally distributed in the *vitellogenin* sequence. Roughly dividing the protein in two parts, the N-terminal domain (N-sheet) is resistant to change, while the major lipid-binding cavity is sprinkled with SNP hotspots (Fig. 2a,b).

The N-sheet contains the phylogenetically conserved, putative receptor binding domain of vitellogenin that presumably is important for uptake to the ovary (Li et al., 2003). This domain can be cleaved off the mature protein (Fig. 2a) in a process that appears to be regulated tissue-specifically in honey bee workers (Havukainen et al., 2011). The biological significance of this dynamic is unknown, but current evidence supports the conclusion of Kent et al. (2011); that the N-sheet is central to vitellogenin's conserved function in reproduction, while the remaining protein has been freer to evolve.

This remaining part of vitellogenin is dominated by a lipid-binding cavity (Fig. 2a,b) where any structural analysis must proceed with caution. This is because the only available vitellogenin X-ray structure is from a vertebrate (Anderson et al., 1998), and insect vitellogenins have features that are missing from this structural template: The most significant is a linker coil (~80 amino acids long) with a tract of serine residues between the N-sheet and the subsequent α-helical domain of the lipid-binding cavity (Fig. 2a) (Tufail, Takeda, 2008). If such insect-specific sequence elements are forced into the cavity region of Vg during structural modeling, the positions of amino acids are displaced relative to the conserved domain fold in the cavity. These displacements occur in the homology model by Kent and coworkers. SNP locations are accordingly shifted from their true positions in the tertiary Vg structure, and the model's predictions about specific, physical connections (at 4-5 Å resolution) between individual SNPs and lipids become very difficult to interpret.

This discussion, however, does not detract from the overall conclusion that polymorphisms in the lipid-binding cavity of vitellogenin can alter ligand-binding properties (Kent et al., 2011). These properties may explain how vitellogenin regulates worker behavior, and SNPs that alter vitellogenin binding- and transport-efficacies could translate into adaptive phenotypic differences. For example, a proposed ligand of vitellogenin is juvenile hormone, which can promote flight behavior and foraging in worker bees. The binding-affinity of vitellogenin to circulating juvenile hormone appears to be weak (Nilsen et al., 2011), but may be biologically relevant since vitellogenin can be present in large amounts (up to 100 μg/μl blood). Other suggested ligands are fatty acids from brood pheromone, which might partly explain why workers with different vitellogenin levels, and thus different ligand binding capacities, have separate behavioral responses to the pheromone: Workers with high vitellogenin levels respond with larval care-behavior, while bees with low vitellogenin levels respond with increased protein (pollen) collection (see (Smedal et al., 2009) and references therein). Alternatively, the general lipid load of vitellogenin might vary in size or composition based on structural features of the cavity and influence several organism features.

In closing, Kent et al. (2011) suggest that genes involved in social pleiotropy, like *vitellogenin,* may be less constrained in molecular evolution than genes involved in other forms of pleiotropy. This can occur if socially pleiotropic alleles usually have neutral or positive effects across castes. But, can socially pleiotropic loci be actively protected from antagonistic alleles? For honey bee *vitellogenin*, we believe the answer might be yes: This gene is not only expressed by workers and queens; it is expressed by larvae before their caste fate is set (Guidugli et al., 2005). Such young larval bees are cannibalized at low but measurable rates by workers (see (Aase et al., 2005) and references therein). This *vitellogenin*, thereby, is expressed in a bipotent larval form that can be selectively discarded by the colony. The functions of *vitellogenin* in bee larvae are not well understood, but we have shown that *vitellogenin* knockdown larvae are viable but more frequently cannibalized than controls (Aase et al., 2005). We speculate that *vitellogenin* affects larval quality traits that can be assessed by the adult workers. Thereby, if bipotent larvae are negatively affected by antagonistic social pleiotropy, then the frequency of antagonistic *vitellogenin* alleles in honey bee populations could be suppressed and allow vitellogenin to be more effectively sculpted to the needs of the colony.

References

Aase ALTO, Amdam GV, Hagen A, Omholt SW (2005) A new method for rearing genetically manipulated honey bee workers. *Apidologie* **36**, 293-299.

Amdam GV, Norberg K, Omholt SW*, et al.* (2005) Higher vitellogenin concentrations in honey bee workers may be an adaptation to life in temperate climates. *Insectes Soc* **52**, 316-319.

Anderson TA, Levitt DG, Banaszak LJ (1998) The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. *Struct Folding & Design* **6**, 895-909.

Engels W (1974) Occurrence and significance of vitellogenins in female castes of social hymenoptera. *Am Zool* **14**, 1229-1237.

Guidugli KR, Piulachs MD, Belles X, Lourenco AP, Simões ZLP (2005) Vitellogenin expression in queen ovaries and in larvae of both sexes of *Apis mellifera*. *Arch Insect Biochem Physiol* **59**, 211-218.

Havukainen H, Halskau O, Skjaerven L, Smedal B, Amdam GV (2011) Deconstructing honeybee vitellogenin: novel 40 kDa fragment assigned to its N terminus. *J Exp Biol* **214**, 582-592.

Kent CF, ISSA A, Bunting AC, Zayed A (2011) Adaptive evolution of a key gene affecting queen and worker traits in the honey bee, *Apis mellifera*. *Molec Ecol* **20**, In press.

Li A, Sadasivam M, Ding JL (2003) Receptor-ligand interaction between vitellogenin receptor (VtgR) and vitellogenin (Vtg), implications on low density lipoprotein receptor and apolipoprotein B/E. *J Biol Chem* **278**, 2799-2806.

Nelson CM, Ihle K, Amdam GV, Fondrk MK, Page RE (2007) The gene *vitellogenin* has multiple coordinating effects on social organization. *PLoS Biol* **5**, 673-677.

Nilsen KA, Ihle KE, Frederick K*, et al.* (2011) Insulin-like peptide genes in honey bee fat body respond differently to manipulation of social behavioral physiology. *J Exp Biol* **214**, 1488-1497.

Qasba PK, Kumar S (1997) Molecular divergence of lysozymes and alpha-lactalbumin. *Crit Rev Biochem Mol Biol* **32**, 255-306.

Seehuus SC, Norberg K, Gimsa U, Krekling T, Amdam GV (2006) Reproductive protein protects sterile honey bee workers from oxidative stress. *Proc Natl Acad Sci USA* **103**, 962-967.

Smedal B, Brynem M, Kreibich CD, Amdam GV (2009) Brood pheromone suppresses physiology of extreme longevity in honeybees (*Apis mellifera*). *J Exp Biol* **212**, 3795-3801.

Tufail M, Takeda M (2008) Molecular characteristics of insect vitellogenins. *J Insect Physiol* **54**, 1447-1458.

Figure legends

Figure 1. Socially pleiotropic functions of vitellogenin in the honey bee *Apis mellifera*. In queens (a-b), vitellogenin (orange) is an egg-yolk protein precursor that may also enhance longevity through effects on metabolism and oxidative stress resistance. Queens can survive several years and they maintain constant high blood levels of vitellogenin, even during periods when they do not lay eggs (see (Seehuus et al., 2006) and references therein). In workers (c-d), vitellogenin influences social behavior: Nest workers that care for larvae have high vitellogenin levels and use vitellogenin-derived products in larval food. When the bees' blood levels of vitellogenin decline 2-3 weeks later, the workers switch from caregiving to risky foraging activities. The result is a division of labor between nest and field activities (Nelson et al., 2007). The high vitellogenin levels of nest workers have been functionally connected to increased oxidative stress resistance and immunity (Seehuus et al., 2006).

Figure 2. An illustration of the honey bee vitellogenin protein based on the lamprey lipovitellin X-ray structure (Anderson et al., 1998). (a) Vitellogenin's conserved N-sheet can get cleaved from the major lipid binding cavity, forming two independent fragments of unknown function in worker tissues (Havukainen et al. 2011). (b) Linear presentation of the 1770 amino acid residues of vitellogenin. The approximate locations of regions under positive or negative selection are adapted from Kent et al. (2011, Fig. 1, Table S4).

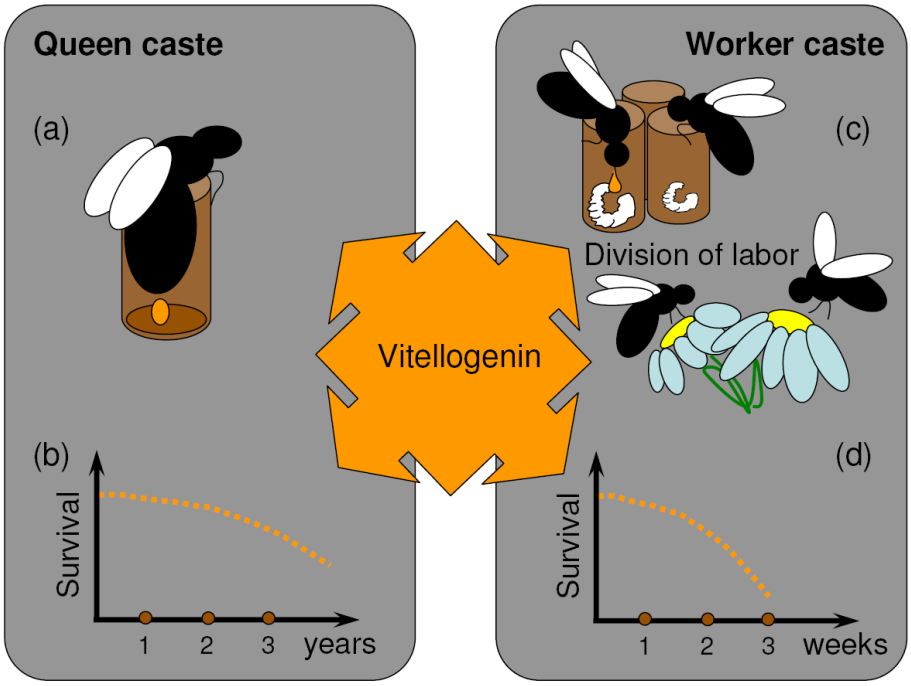


Figure 1.

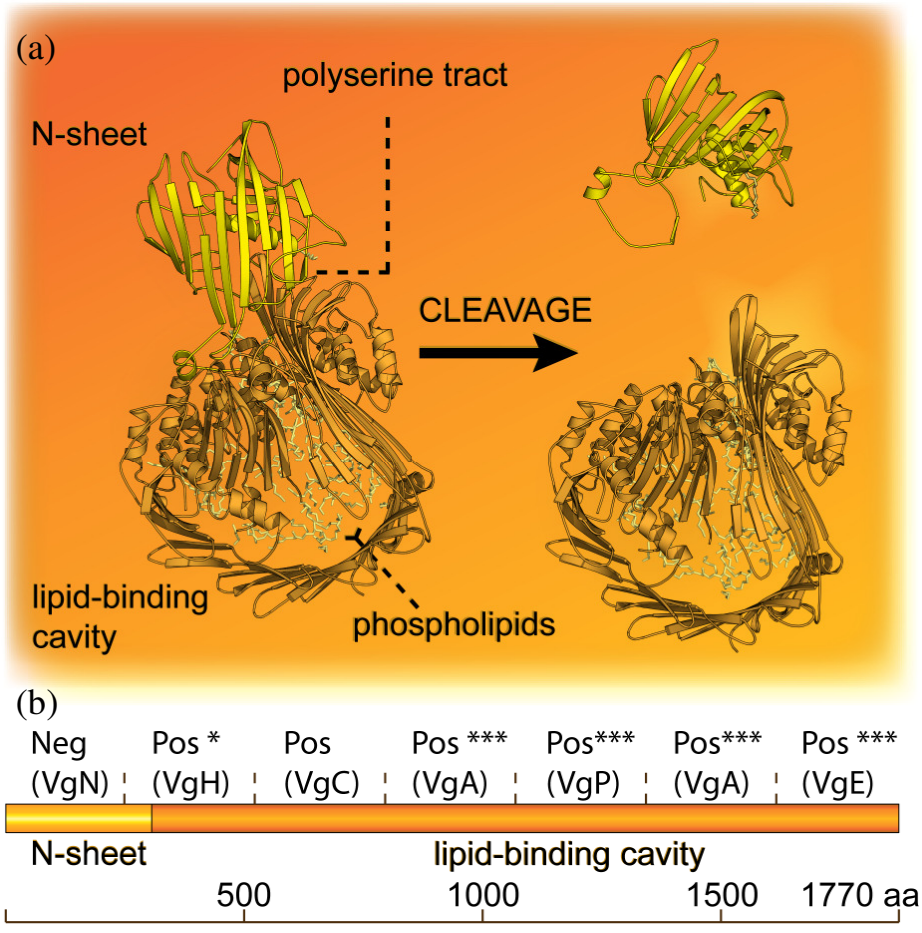


Figure 2.

Paper IV

A vitellogenin polyserine cleavage site: highly disordered conformation protected from proteolysis by phosphorylation

Havukainen, H., Underhaug, J., Wolschin, F., Amdam, G. V., Halskau, Ø.
Submitted manuscript.

Abstract

Vitellogenin (Vg) is an egg-yolk precursor protein in oviparous species. In honeybee (*Apis mellifera*), the protein (AmVg) also affects social behavior and life-span plasticity. Despite its manifold functions, the AmVg molecule remains poorly understood. The object of our structure-oriented AmVg study is its polyserine tract – a little investigated repetitive protein segment. We previously reported that AmVg is tissue-specifically cleaved in the vicinity of this tract. Here, we show that AmVg is unexpectedly resistant to trypsin/chymotrypsin digestion at the tract, given its potential for an open, disordered structure. Our findings suggest that multiple phosphorylations play a role in this resilience. Some Vgs that are phylogenetically close to AmVg have, intriguingly, lower serine content at the tract, among them the Vg of the solitary jewel wasp *Nasonia vitripennis*. We demonstrate that this sequence difference can lead to structural variation, as NMR and circular dichroism evidence assign different conformational propensities to polyserine peptides from the bee and the wasp; the first is extended and disordered and the latter more compact and helical. The spectroscopic results strengthen our model of the AmVg polyserine tract as a flexible domain linker shielded by phosphorylation.

Heli Havukainen[1,2,*], Jarl Underhaug[2], Florian Wolschin[1,3], Gro Amdam[1,3], Øyvind Halskau[1,2,a]

[1] Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, P.O. Box 5003 1432 Aas, Norway
[2] Department of Biomedicine, University of Bergen, Jonas Lies vei 91, 5009 Bergen, Norway
[3] School of Life Sciences, Arizona State University, P.O. Box 874501, Tempe, AZ 85287-4701, USA
[a] Present address: Department of Molecular Biology, University of Bergen, Thormøhlensgate 55, 5020 Bergen, Norway

Abbreviations:

Vg	Vitellogenin
AmVg	*Apis mellifera* Vitellogenin
CD	Circular Dichroism
NvVg	*Nasonia Vitripennis* Vitellogenin
NMR	Nuclear Magnetic Resonance (spectroscopy)
NOE	Nuclear Overhauser Effect

Introduction

Vitellogenins (Vgs) are large (usually >200 kDa), typically female-specific proteins involved in lipid and ion transport during oogenesis. Vg is produced in the insect fatbody as well as in the liver of oviparous vertebrates (Ansari et al., 1971; Dolphin et al., 1971), and is much-studied in many species due to its central function in reproduction. In *Apis mellifera* (honeybee), Vg (AmVg) is not only expressed in female reproductive individuals, but also in sterile worker bees (a helper caste) and in males (Piulachs et al., 2003; Trenczek and Engels, 1986). In the workers, Vg affects social behavior, oxidative stress resistance, and life-span plasticity (Amdam et al., 2003; Nelson et al., 2007; Seehuus et al., 2006). These functions interlink the evolution of AmVg with the development of the bee as a social insect and benefit the colony (Amdam and Omholt, 2003; Nelson et al., 2007). Little is known about the biochemical and structural properties that mediate AmVg's effects, which include hormone-like (Amdam et al., 2007; Amdam et al., 2004) and antioxidant functions (Seehuus et al., 2006). However, it appears that glycosylation, proteolytic cleavage, and phosphorylation could play a role in mediating AmVg's effects (Havukainen et al., 2011).

Vgs show considerable amino acid sequence similarity between species at an N-terminal and at an α-helical domain (Smolenaars et al., 2007; Tufail and Takeda, 2008). Between these two domains, however, there is a hypervariable linker. Insect Vgs appear to contain one to three polyserine tracts of various lengths embedded in this linker. Apocrita, the suborder of *A. mellifera*, may lack extensive polyserine tracts with honeybee and bumblebee as notable exceptions (Tufail and Takeda, 2008). For instance, the Apocrita wasp *Nasonia vitripennis* Vg (NvVg) has only seven serine residues in the corresponding

sequence (NvVg residues 345-367) compared to 14 in the honeybee (AmVg residues 349-381). The crystal structure of lamprey (*Ichthyomyzon unicuspis*) lipovitellin (Anderson et al., 1998; Raag et al., 1988) is widely used as a template for the Vg protein family (Havukainen et al., 2011; Kristoffersen et al., 2009; Li et al.) and even for homologous human proteins (Mann et al., 1999). The vertebrate Vgs do not contain a polyserine region, so structural models for the insect polyserine tracts can not be derived from the existing structural data. In general, very little is known about what functional role polyserine tracts might have.

The polyserine tract is an example of low complexity polypeptide regions. Low complexity regions are semi-repetitive (Dyson and Wright, 2005), and predicted to be disordered (Dunker et al., 2002; Huntley and Golding, 2006). Polyserine tracts are examples of such low-complexity regions and lag behind in biochemical characterization relative to other disordered regions of proteins (for review of those, see (Uversky et al., 2008)). This is presumably because polyserine sequences have traditionally been dismissed as the non-functional result of transcription slippage (Huntley and Golding, 2004; Huntley and Golding, 2006), and because of difficulties in establishing suitable expression systems (Chu et al., 2011) as well as in the chemical synthesis of serine repeats (Barlos et al., 1998). Yet, there are many reports of polyserine regions being functionally important, e.g. in molecular recognition (Huntley and Golding, 2006), bacterial carbohydrate metabolism (Howard et al., 2004) and the pathogenesis of viruses (Bates and DeLuca, 1998).

The putative structural disorder of the polyserine tract might build an attractive site for proteases, and the polyserine tracts are, indeed, flanking a consensus cleavage site

(RXXR) in many insect Vgs (Tufail and Takeda, 2008). However, Apocrita species provide an exception, in that they lack the consensus sequence (Tufail and Takeda, 2008). Despite this, AmVg can get cleaved, tissue-specifically, in the vicinity of the polyserine tract by a protease, or group of proteases, that have yet to be identified (Havukainen et al., 2011). All known Vgs are phosphorylated, and those of insects likely at the polyserine region (Finn, 2007; Tufail and Takeda, 2008). Multiple phosphorylation of proteins can regulate many protein activities including cleavage (Cohen, 2000), but the role of Vg phosphorylation is unclear.

Our aim is to structurally study the AmVg polyserine tract and its relation to the protein as whole. Initially, we narrowed down the AmVg cleavage site, and used limited proteolysis to a) see if the natural AmVg cleavage pattern can be triggered with trypsin/chymotrypsin and b) to see if the digestion pattern matches the expectations based on lamprey lipovitellin structure (Anderson et al., 1998; Raag et al., 1988). The latter expectation was met, but no indication of digestion in the vicinity of the polyserine tract was observed, except with a dephosphorylated AmVg sample. Therefore, we speculate that phosphorylation might shield the polyserine tract against unspecific cleavage. We noted that the polyserine tract contains a casein kinase II recognition motif and experimentally verified that the polyserine tract can be multiply phosphorylated by the respective kinase. In order to explore the predicted propensity for disorder of the polyserine tract, we studied it using NMR and far-UV circular dichroism (CD) methods. We chose a representative comparison among the serine-poor Apocrita species, *N. vitripennis*, in order to determine whether the varying level of the serine content has structural effects. The results unveil the AmVg polyserine tract as an extended, flexible

coil, whereas the peptide of *N. vitripennis* has a more collapsed conformational ensemble with helical tendency.

Results

Sequence alignment of Hymenoptera polyserine tracts and their predicted disorder: To highlight the variability of Vg polyserine tracts, we performed a sequence alignment with thirteen Hymenoptera Vgs. In general, the bees' Vgs (*A. mellifera*, *Bombus ignitus* and *Bombus hypocrita*) have more serine residues at the polyserine region than the solitary wasps (*Nasonia vitripennis*, *Pteromalus puparum*, *Encarsia formosa* and *Pimpla nipponica*) (Fig. 1A). Ants have multiple Vgs that vary in their serine content. All three Vgs of *Solenopsis invicta* are shown, since *S. invicta* has one Vg with very high serine proportion (Vg-2), but only one example of Vgs of *Harpegnathos saltator* and *Camponotus floridanus* is presented, because the latter two ants are not particularly rich in Vg serine content. Separated from the other species, the extremely serine-rich tract of *Athalie rosae*, a member of distinct suborder (Symphyta), is shown. Hymenoptera Vgs are predicted to have a long stretch of disorder at the linker site independent of the serine content (Fig. 1B). Figure 1C visualizes the polyserine linker location in between the N-terminal β-barrel-like domain and the main lipid-binding cavity, based on the crystal structure of lamprey lipovitellin (figure prepared based on the Protein Data Bank file **1LSH** (Anderson et al., 1998; Raag et al., 1988)). The red linker in this case does not contain a polyserine tract, since vertebrate Vgs such as lamprey lipovitellin do not contain the tract. We conclude that the prevalence of the serine residues at this disordered domain-connector site varies greatly in the Hymenoptera. We have endeavored to include

all the currently available Apocrita sequences in our alignment, but stress that these sequences are greatly outnumbered by the total Apocrita species.

Limited proteolysis and identification of fragments: The naturally occurring, tissue-dependent cleavage pattern of AmVg to 40 and 150 kDa fragments suggests that this is a specific occurrence, rather than a result of general degradation (Havukainen et al., 2011). The predicted disorder in the vicinity of the cleavage site should, in principle, make this segment prone to proteolysis by a range of proteases. To investigate whether the observed cleavage pattern can be mirrored *in vitro*, we proceeded to explore the effect of limited proteolysis on AmVg purified from bees as previously described (Havukainen et al., 2011). We tested proteolysis of AmVg by trypsin and chymotrypsin (Fig. 2A), both of which can cut at sites in the polyserine region (see the triangles in Fig. 2C). The polyserine tract, however, proved resistant to trypsin and chymotrypsin, since the enzymes preferentially cleaved at other locations: a C-terminal fragment is detached at low protease concentrations from both full-length and 150 kDa fragments, corresponding to approximately 100 C-terminal residues and several other bands of unspecific fragmentation emerged (Fig. 2A). Thus, neither of these proteases assists fragmentation to the 40 and 150 kDa units observed *in vivo*. LC-MS/MS experiments on the most trypsin- and chymotrypsin-resistant Vg bands (of Fig. 2A) identified them as the lipid-binding cavity of Vg, with approximately > 490 residue loss at the N-terminus and < 130 residue loss at the C-terminus (Fig. 2C). In order to determine the exact cleavage sites in the AmVg polypeptide both the 150 kDa and 40 kDa Vg fragments were subjected to N-terminal sequencing. Edman degradation of the 40 kDa fragment for 10 cycles gave the sequence DFQHNWQVGN, which shows that this fragment is at the N-terminal part of

the polypeptide (the 16 amino acid signal sequence has been removed as predicted). The 150 kDa fragment gave no signals (two repeated experiments) during 10 cycles of Edman degradation indicating that its N-terminus may be blocked. In order to more accurately focus on the location of the cleavage site between the 40 kDa and 150 kDa fragments, both fragments were subjected to MALDI-TOF peptide mass fingerprint analysis. The peptides from the 40 kDa Vg could be localized between AmVg residue 17 (D) (signal sequence cleavage site) and residue 335 (R). The peptide masses from the 150 kDa fragment could be localized from amino acid 427 (E) to amino acid 1732 (K). According to this experiment, the cleavage site between the 40 kDa and 150 kDa fragments is localized in the polyserine tract between amino acids 335 and 427 (Fig. 2C, Supplement 1). In an ExPASy PeptideCutter (Wilkins et al., 1999) search, caspase 1 was found as the only relevant, i.e. non-bacterial and non-gut, enzyme having a cleavage site at AmVg 336-427, at residue 370 and additionally at residue 1604.

Dephosphorylation of full-length AmVg and *in vitro* phosphorylation of polyserine region: The high resistance of the polyserine region towards proteolysis could be explained by the target site being structured, by interference from compact neighboring domains, or by post-translational modifications. Both the N-terminal (40 kDa) and the C-terminal (150 kDa) fragment of AmVg are known to be phosphorylated (Havukainen et al., 2011). However, the exact location of the modifications is uncertain. We detected a CKII (casein kinase II) recognition motif inside the polyserine region. A short synthetic peptide from this region was readily phosphorylated on multiple serine residues by casein kinase II *in vitro* (Fig. 3). We proceeded to test the role of phosphorylation in the protease resistance of the polyserine tract. AmVg was dephosphorylated and digested using

trypsin/chymotrypsin as above. The cleavage pattern gains characteristics of the natural 40 + 150 kDa pattern (Fig. 2B): both the 150 kDa and the 40 kDa band are stronger at higher protease concentrations for the dephosphorylated protein (Fig. 2A and 2B). The change in the cleavage pattern supports the hypothesis that phosphorylation is involved in conveying resistance to proteolysis.

NMR structural propensity of monophosphorylated polyserine region: To determine whether this region has a structure or structural propensity that can further help explain its behavior, both with respect to protease accessibility and the linker's role in AmVg, we performed an NMR study (Figs 4-6, Table 1, Supplement 2). In order to limit the complexity of the synthesis and to provide a measure of solubility to the polypeptide, we opted for a monophosphorylated version of the tract (See also Material and Methods – Peptide design and synthesis). In this study we attempted to alleviate the problems related to exploring a polypeptide segment detached from the protein as a whole to some extent by *i*) determining the structure for an AmVg peptide that is tractable with respect to solubility, synthesis, and spectral overlap, and *ii*) use the corresponding *Nasonia vitripennis* Vg (NvVg) segments as reference when interpreting structural data.

AmVg(358-392) and NvVg(351-385) produced TOCSY spectra of similar, good quality (Fig. 4A,C). Both peptides were assignable, despite some ambiguity in the polyserine interval in AmVg(358-392). This ambiguity was somewhat alleviated due to peptide design and phosphorylation (see Material and Methods – Peptide design and synthesis). The UV-spectrum of the peptides show a considerable difference, suggesting random coil for AmVg(358-392) and a more structured nature of NvVg(351-385) (Fig. 4B,D). Chemical Shift Indexing (CSI) (Wishart et al., 1992) indicated that the polypeptide

backbone propensities for secondary structure were different for the two peptides (Fig. 5B,D). While NvVg(351-385) was dominated by CSI values indicative of α-helical conformations, AmVg(358-392) was almost devoid of any propensity with respect to secondary structure, despite *in silico* prediction of an α-helix (Fig. 5A). The lack of NOEs in the spectra of AmVg(358-392) (Fig. 5A, Table 1) could not be ascribed to incomplete assignment or sample issues, since in its *N. vitripennis* counterpart of similar quality, NOEs indicative of structure were readily found.

Solving the structure of the two peptides resulted in an ensemble of outstretched coil structures in the case of AmVg(358-392) (Fig. 6A), whereas that for NvVg(351-385) contains a helix between residues 371-377 (Fig. 5A,D; Fig. 6C; Table 1). Figures 6B and 6D show the surface electrostatics for the serine regions, dominated by negative charge even in their monophosphorylated state. These structures are stored in Protein Data Bank and Biological Magnetic Resonance Bank with the following ID numbers for AmVg(358-392): **2lic** and **17888**, respectively, and for NvVg(351-385): **2lid** and **17889**.

We then proceeded to solve the longer peptides, which partially overlap with the peptide AmVg(358-392) and NvVg(351-385). We did this to extend the results along the Vg polypeptide chain, and to assess the effect different peptide endpoints have on the overall structure of the shorter peptides. The AmVg(226-385) peptide suffered from lower solubility and purity than its *N. vitripennis* counterpart; still, much of the backbone was assigned, and CSI values of both peptides are presented in Fig. 5C and E. AmVg(336-385) retains its random values and lack of NOEs indicative of any fixed structure, a result also supported by the CD-spectrum (S2 Fig. 1A,B); the data were not processed into a structure. NvVg(335-372), with its C-terminal in the helix occurring in the shorter

peptide, shows marked changes, with an attenuation of its NOE pattern (S2 Fig. 2, S2 Table 1). Moreover, the changes are visible in the spectrum as a collapse of the signals (S2 Fig. 1C), relative to Fig. 4C. The corresponding CD-spectra also indicate a random coil (S2 Fig. 1D).

Discussion

This study concentrates on a hypervariable domain connector, the polyserine tract, of AmVg. Our earlier work (Havukainen et al., 2011), found that AmVg was cleaved in a tissue-specific manner in the vicinity of this tract, and proposed that this might have relevance for the effects AmVg has on bee behavior and life-span plasticity. Although hypervariable regions may be nonfunctional in many cases, the differences seen at the polyserine tract between bees (Apoidea) and wasps (Chalcidoidea and Ichneumonoidea) and elsewhere in the Hymenoptera order are intriguing. For instance, are wasps such as *N. vitripennis* exposed to an evolutionary pressure to keep the serine content low? And why does a more distantly related hymenopteran sawfly, *Athalia rosae*, have extremely many serine residues, similarly to mosquito (Diptera: *Aedes aegypti*) (Romans et al., 1995)? As honeybees have harnessed the ancient yolk-precursor Vg for the organization of a social hierarchy (reviewed, among others, by (Denison and Raymond-Delpech, 2008)), other species might have evolved novel functions for their Vg(s) as well. The serine tracts of the sawfly and silkworm (*Bombyx mori*) have been speculated to be connected to silk formation as silk fibroin is rich in serine (Chen et al., 1997). Of the few functionally and structurally well-characterized poly-serine proteins, spider silk protein

may exemplify how transcriptional slippage has resulted in a tremendously successful and versatile adaption (Ayoub et al., 2007; Hu et al., 2005).

The polyserine linkers are an example of protein regions that are predicted to be essentially disordered *in silico* (Ward et al., 2004). Proteases like trypsin and chymotrypsin prefer easy-access, i.e. flexible, disordered regions as their primary targets (Fontana et al., 2004; Stroh et al., 2005). Thus, we initially expected to see instant cleavage at the polyserine linker by trypsin/chymotrypsin. When this was not observed, we speculated that lack of the expected effect was due to phosphorylation at the polyserine linker (Benore-Parsons et al., 1989). Indeed, instead of strong unspecific cleavage, the digestion of dephosphorylated Vg produces a fragmentation pattern of 40 + 150 kDa. We further confirmed by mass-spectrometry that an AmVg polyserine peptide can be multiphosphorylated *in vitro* by CKII. We used CKII since the polyserine tract harbors a CKII recognition motif (SXXE/D) and CKII has been previously shown to phosphorylate *Rhodnius prolixus* Vg (Silva-Neto et al., 2002). Thus, our results support prior speculations about the polyserine linker harboring phosphorylation sites (Chen et al., 1994; Tufail and Takeda, 2008).

The cleavage in the vicinity of the polyserine region occurs naturally in the bee fatbody tissue and during AmVg purification (Havukainen et al., 2011; Wheeler and Kawooya, 1990). In contrast to many species expressing Vg, AmVg does not have the consensus cleavage site (Barr, 1991; Rouille et al., 1995). We propose that an alternative way of cleavage may have evolved. For instance, auto-cleavage or activity of an un-characterized enzyme can be involved. So far, we have managed to narrow the location of the cleavage site to residues 335-427. Further efforts will be needed to pinpoint the exact

site and to test protease candidates. Protease database search against the cleavage site (336-427) results in one candidate enzyme: caspase 1. Insect caspases have been studied mostly in *Drosophila*, and they have similar Asp-X cleavage preference as the mammalian caspases (Cooper et al., 2009; Hawkins et al., 2000); however, this is not verified in honeybees. In mammals, the phosphorylation of residues adjacent to the cleavage site is known to block caspase cleavage (Hoon Kim et al., 2003). Intriguingly, the existence of a phosphorylation-mediated regulation mechanism has been previously speculated for mosquito Vg (Don-Wheeler and Engelmann, 1997). Other examples of phosphorylation-regulated cleavage are presenilin-2 cleavage that is inhibited by phosphorylation near the cleavage site (Walter et al., 1999) and cohesion cleavage that is triggered by nearby phosphorylation (Alexandru et al., 2001).

To gain a better understanding of the overall domain organization of AmVg, we inspected the result of limited proteolysis, as this is a useful technique in assessing compactness and disorder (Sharp et al., 2006; Stroh et al., 2005). A highly protease resistant region that contains a compact lipid-binding cavity at the corresponding region in lamprey structure was identified. Moreover, the observation that there is an easily cleavable C-terminal piece consisting of ~ 100 residue is consistent with the lamprey X-ray structure that lacks electron density corresponding to this region (Anderson et al., 1998; Raag et al., 1988). Thus, the general domains and disordered regions of AmVg as probed by proteolysis seem to correspond to the known Vg structure of lamprey.

A vertebrate Vg structure cannot be used for inspection of the insect specific polyserine tract. Our comparative study by NMR of the two segments from homologous proteins, AmVg and NvVg, predicted to be disordered, presents an opportunity to comment on

differences in the degree of disorder. Secondary structure prediction suggests some helical tendency at the polyserine tract for both species. However, only in the case of *N. vitripennis* linker was this prediction supported by experimental evidence. The extended conformations of the *A. mellifera* linker and the more collapsed ensemble of structures found for *N. vitripennis* suggest that the disorder can take on somewhat different forms. The divergent serine content of these two species may be the cause of the observed differences in conformational propensities. Indeed, computational approaches on repetitive stretches of peptides associated with disorder show that they differ significantly in their tendency to form coils with respect to kinetics and surrounding water molecules (Doruker and Bahar, 1997).

The polyserine region has undergone extensive changes since the divergence of the species *N. vitripennis* and *A. mellifera*. We conclude this study by suggesting that the serine rich domain linker of insect Vgs is a site of high plasticity, both in terms of amino acid sequence and polypeptide conformation. We suggest that residues can be added by transcriptional slippage and kept without much adverse effect at such regions. These plastic regions may subsequently evolve new functionality. Thus, we propose that although AmVg has retained the overall structure as determined for lamprey Vg, it has developed another mechanism for cleavage in its linker region than what is currently known in non-apocritan species. It is likely that multiple phosphorylation by CKII is part of the regulation of this cleavage.

Material and Methods

Sequence alignment of Hymenoptera polyserine tracts and prediction of disorder: Sequences of Hymenoptera Vgs were obtained from UniProtKB (Jain et al., 2009) and NCBI Protein Database in June 2011, aligned using ClustalW2 (Thompson et al., 2002), and cropped at ultra-conserved sites in the N- and the C-terminus surrounding the polyserine tract (honeybee residues 321-407). The alignment was manually adjusted to minimize gaps. A prediction of *A. mellifera* polyserine tract disorder (residues 321-420) and other selected Vg tracts was obtained using Disopred (Ward et al., 2004). An illustration highlighting the region where the polyserine tract is located in insects was produced with Pymol (DeLano, 2002) using lamprey lipovitellin structure (Anderson et al., 1998; Raag et al., 1988).

Limited proteolysis: A limited proteolysis analysis was performed in order to investigate if the natural Vg cleavage, occurring in the vicinity of the serine tract, could be triggered. Vg was purified from honeybee abdomens by GenScript (NY, USA) according to (Havukainen et al., 2011) with size-exclusion and ion-exchange chromatography steps, and the purification was completed with a Superdex 75 column (Sigma-Aldrich, St Louis, MO, USA). Pure Vg was partially digested by incubating 6.4 µg of samples with 0.25, 2.5 and 25 ng of trypsin or chymotrypsin (Sigma-Aldrich, MO, USA) in a total volume of 10.5 µl for 30 min on ice. For digestion of dephosphorylated Vg, samples were incubated for 1 h at 37°C with 1 unit (0.5 µl) calf intestinal phosphatase (New England BioLabs, Inc., Beverly, MA, USA) prior to digestion. Untreated Vg and Vg incubated with the phosphatase served as controls. The samples were run on a 4%-20% SDS-PAGE gel (Bio-Rad, CA, USA) under reducing conditions.

LC-MS/MS of the most protease-resistant part of Vg: The most protease resistant bands of the limited proteolysis assay were excised for LC-MS/MS identification (performed essentially as described in Havukainen et al. 2011 (Havukainen et al., 2011)). The gel pieces were washed, treated with 10 mM DTT (Amersham Biosciences, Piscataway, USA) for cystein reduction and alkylated using iodoacetamide (Sigma-Aldrich), followed by full trypsin digestion. After digestion, the dried samples were dissolved in 11 µl 0,1% formic acid and 6 µl was used for injection. The samples were analyzed on a 4000 QTrap (Applied Biosystems/MDS SCIEX, Concord, ON, Canada).

N-terminal sequencing and MALDI-TOF peptide mass fingerprinting: 150 kDa and 40 kDa fragments were separated on SDS-PAGE gels (Bio Rad) and electroblotted onto a polyvinylidene difluoride membrane followed by Coomassie Brilliant blue staining. N-terminal protein sequencing by Edman degradation was performed using a Procise 494A HT Sequencer (Perkin Elmer, Applied Biosystems Division, Foster City, CA, USA). For peptide mass fingerprint analysis, 40 and 150 kDa Vg fragments were separated on gel, excised and "in gel" digested as described in the preceding paragraph. Peptides generated by enzymatic cleavage were analyzed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry using an Ultraflex TOF/TOF instrument (Bruker-Daltonik GmbH, Bremen, Germany). ExPASy PeptideCutter (Wilkins et al., 1999) was used for *in silico* digestion of AmVg in order to find putative proteases that cut at the cleavage site, between residues 336-427.

Vitellogenin *in vitro* phosphorylation and LC-MSMS: The Vg-derived peptide TDISSSSSSISSSEENDFWQPK (synthesized by the Biochemistry department at Arizona State University) was incubated essentially following the recommendations provided by

the casein kinase II supplier (New England Biolabs, Ipswich, MA, USA). A 20 µl reaction contained 7 µl of a 0.4 mM peptide solution, 4.5 µl ddH_2O, 5 µl of a 0.1 mM ATP solution, 2 µl of 10 times kinase reaction buffer, and 1 µl of a 500 units/µl kinase solution. The reaction was let to proceed for 1h at 30°C and stopped by the addition of 50 µl ice-cold ethanol, extensive mixing, and incubation on ice for 20 min. Subsequently, the kinase was removed by centrifugation (13.400 g at 4°C for 3 min) and the supernatant was concentrated in a speed vacuum. For LC-MS/MS measurements, the peptide sample was dissolved in 0.1% formic acid and injected into a LTQ linear ion trap mass spectrometer (Thermo Scientific, Waltham, MA, USA). Spectra were recorded for the doubly charged peptide with zero to five phosphorylated residues (CID: 35%).

Peptide design and synthesis: Two peptides of the polyserine tract of honeybee and jewel wasp Vg were synthesized by CPC Scientific (CA, USA). The synthesis was performed after isotope-labeling expression attempts in *Escherichia coli* were deemed unsuitable for the project due to very low final yield. Synthesis of serine repeats is hampered by the steric conflicts caused by the chemical group used to protect the side-chains from chemical modification during synthesis. In order to make synthesis of serine repeats feasible and aid assignment, one serine residue was mutated into a residue of similar properties in each peptide. The peptides had the following sequences:

AmVg (residues 358-392):

EKLKQDILNLRTDIS**T**S(Sp)SS(^{15}I)SSSEENDFWQPKPT

AmVg (residues 336-385):

R(^{15}V)SKT(^{15}A)MNSNQI(^{15}V)SDNS(^{15}L)(^{15}S)STEEK(^{15}L)KQDI(^{15}L)N(^{15}L)RTDI(^{15}S)S (^{15}S)(Sp)S(**^{15}A**)IS (^{15}S)(^{15}S)EEND.

The peptides of *N. vitripennis* were:

NvVg (residues 351-385):

EHKHSDES**T**SE(Sp)FES(^{15}I)AD NNDDSYFQRKPKLTEAP

NvVg (residues 335-372):

RPNK(^{15}L)N(^{15}L)QRRHDHKS(^{15}G)EHKHSDE(^{15}S)S(^{15}S)E(Sp)FE(^{15} A)I(**^{15}A)**DNNDD.

Mutated residues are underlined and bolded, phosphorylation is denoted by (Sp), and ^{15}N labeling is denoted by (^{15}X), X being the labeled residue.

Peptides were solubilized in 250 µl of 50 mM phosphate buffer pH 6.7 containing 0.02% NaN_3, resulting in the following concentrations after centrifugation at 20,800 g for 15 minutes: AmVg(358-392) 7.9 mg/ml, AmVg(336-385) 1,8 mg/ml, NvVg(351-385) 20 mg/ml and NvVg(335-372) 2.4 mg/ml. Tetramethylsilane was used as a reference compound.

CD spectroscopy: The samples were diluted to 0.5 mg/ml with 50 mM phosphate buffer pH 6.7. Far-UV spectra (190-260 nm) were recorded on J-810 spectropolarimeter (JASCO, Tokyo, Japan).

NMR structures: An AVII 600 MHz spectrometer (Bruker, Billerica, USA) equipped with a cryoprobe was used for accumulating the following two-dimensional spectra: HSQC, HSQC-NOESY, TOCSY and NOESY. Water suppression was achieved using pulse-sculpting and gradient spin-echo in combination with various modifications of the Watergate sequence. Typical processing parameters were: 4 k number of points in F2 and 2 k in F1; SI1 and SI2 4 k. Squared sine (QSINE) was used as the window function. Decoupled and un-decoupled spectra were recorded in order to identify the labeled residues. The spectra were assigned using Sparky (Goddard and Kneller) and 200 three-

dimensional structures of AmVg(358-392) and NvVg(351-385) were calculated using Aria1.2 and CNS (Linge et al., 2002), of which NvVg(351-385) was calculated only based on the fully assigned peaks.

Accession numbers: Swiss-Prot: *Apis mellifera* **Q868N5**, *Bombus ignitus* **B9VUV6**, *Bombus hypocrita* **C7F9J8**, *Solenopsis invicta*Vg-1 **Q7Z1M0**, *Solenopsis invicta*Vg-2 **Q2VQM6**, *Solenopsis invicta*Vg-3 **Q2VQM5**, *Harpegnathos saltator* **E2BDX3**, *Camponotus floridanus* **E2ANT2**, *Pteromalus puparum* **B2BD67**, *Encarsia Formosa* **AAT48601**, *Pimpla Nipponica* **O17428** and *Athalia rosae* **BAA22791** and *Nasonia vitripennis* with NCBI: **XP_001607388**. Lamprey lipovitellin PDB-ID: 1LSH. The *A. mellifera* polyserine peptide NMR structure: PDB-ID: **2lic**, BMRB-ID: **17888**, and *N. vitripennis* peptide: **2lid** and **17889**.

Acknowledgements: We thank professor Aurora Martinez for outstanding support; this project took place in her group of Biorecognition, at University of Bergen. LC-MS/MS experiments were performed by the PROBE Proteomic Unit at University of Bergen and at Arizona State University. We thank Dr. Hilde Garberg for expert technical support and Dr. Frode Berven for expert advice. The PROBE work was partly supported by the National Program for Research in Functional Genomics (FUGE) funded by the Norwegian Research Council. The N-terminal sequencing and peptide mass fingerprinting were performed by Protein Chemistry Research Group and Core Facility, Institute of Biotechnology, University of Helsinki, and we want to thank Nisse Kalkkinen for expert advice. We thank Atle Aaberg, Nils Åge Føystein and Annette Brenner for practical assistance at the NMR facility. Ø.H. was supported by the Research council (#185306) and Norwegian Cancer Society (#58240001). G.V.A. was supported by the Research Council of Norway (#180504, 185306 and 191699), the National Institute on Aging (NIA P01 AG22500) and the PEW Charitable Trust.

Author Contributions: All experiments except for the in vitro phosphorylation analysis were conceived by Ø.H., G.V.A. and H.H. The *in vitro* phosphorylation and LC-MSMS section was designed and performed by F.W. Ø.H. designed the peptides. H.H., Ø.H. and J.U. recorded and assigned the NMR data. H.H. and J.U. calculated the structures. H.H. performed the rest of the experiments if not otherwise implicated. All authors participated in writing of the manuscript.

References

Alexandru, G., Uhlmann, F., Mechtler, K., Poupart, M. A. and Nasmyth, K. (2001). Phosphorylation of the cohesin subunit Scc1 by Polo/Cdc5 kinase regulates sister chromatid separation in yeast. *Cell* **105**, 459-72.

Amdam, G. V., Nilsen, K. A., Norberg, K., Fondrk, M. K. and Hartfelder, K. (2007). Variation in endocrine signaling underlies variation in social life history. *Am Nat* **170**, 37-46.

Amdam, G. V., Norberg, K., Hagen, A. and Omholt, S. W. (2003). Social exploitation of vitellogenin. *Proc Natl Acad Sci U S A* **100**, 1799-802.

Amdam, G. V. and Omholt, S. W. (2003). The hive bee to forager transition in honeybee colonies: the double repressor hypothesis. *J Theor Biol* **223**, 451-64.

Amdam, G. V., Simoes, Z. L., Hagen, A., Norberg, K., Schroder, K., Mikkelsen, O., Kirkwood, T. B. and Omholt, S. W. (2004). Hormonal control of the yolk precursor vitellogenin regulates immune function and longevity in honeybees. *Exp Gerontol* **39**, 767-73.

Anderson, T. A., Levitt, D. G. and Banaszak, L. J. (1998). The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. *Structure* **6**, 895-909.

Ansari, A. Q., Dolphin, P. J., Lazier, C. B., Munday, K. A. and Akhtar, M. (1971). Chemical composition of an oestrogen-induced calcium-binding glycolipophosphoprotein in Xenopus laevis. *Biochem J* **122**, 107-13.

Ayoub, N. A., Garb, J. E., Tinghitella, R. M., Collin, M. A. and Hayashi, C. Y. (2007). Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. *PloS one* **2**, e514.

Barlos, K., Gatos, D. and Koutsogianni, S. (1998). Fmoc/Trt-amino acids: comparison to Fmoc/tBu-amino acids in peptide synthesis. *The journal of peptide research : official journal of the American Peptide Society* **51**, 194-200.

Barr, P. J. (1991). Mammalian subtilisins: the long-sought dibasic processing endoproteases. *Cell* **66**, 1-3.

Bates, P. A. and DeLuca, N. A. (1998). The polyserine tract of herpes simplex virus ICP4 is required for normal viral gene expression and growth in murine trigeminal ganglia. *J Virol* **72**, 7115-24.

Benore-Parsons, M., Seidah, N. G. and Wennogle, L. P. (1989). Substrate phosphorylation can inhibit proteolysis by trypsin-like enzymes. *Arch Biochem Biophys* **272**, 274-80.

Chen, J. S., Cho, W. L. and Raikhel, A. S. (1994). Analysis of mosquito vitellogenin cDNA. Similarity with vertebrate phosvitins and arthropod serum proteins. *J Mol Biol* **237**, 641-7.

Chen, J. S., Sappington, T. W. and Raikhel, A. S. (1997). Extensive sequence conservation among insect, nematode, and vertebrate vitellogenins reveals ancient common ancestry. *J Mol Evol* **44**, 440-51.

Chu, H. S., Ryum, J., Park, S. Y., Kim, B. G., Kim, D. M. and Won, J. I. (2011). A new cloning strategy for generating multiple repeats of a repetitive polypeptide based on non-template PCR. *Biotechnology letters* **33**, 977-83.

Cohen, P. (2000). The regulation of protein function by multisite phosphorylation--a 25 year update. *Trends Biochem Sci* **25**, 596-601.

Cooper, D. M., Granville, D. J. and Lowenberger, C. (2009). The insect caspases. *Apoptosis* **14**, 247-56.

DeLano, W. L. (2002). The PyMOL Molecular Graphics System. *DeLano Scientific, Palo Alto, CA, USA*.

Denison, R. and Raymond-Delpech, V. (2008). Insights into the molecular basis of social behaviour from studies on the honeybee, Apis mellifera. *Invert Neurosci* **8**, 1-9.

Dolphin, P. J., Ansari, A. Q., Lazier, C. B., Munday, K. A. and Akhtar, M. (1971). Studies on the induction and biosynthesis of vitellogenin, an oestrogen-induced glycolipophosphoprotein. *Biochem J* **124**, 751-8.

Don-Wheeler, G. and Engelmann, F. (1997). The biosynthesis and processing of vitellogenin in the fat bodies of females and males of the cockroach Leucophaea maderae. *Insect Biochemistry and Molecular Biology* **27**, 901-918.

Doruker, P. and Bahar, I. (1997). Role of water on unfolding kinetics of helical peptides studied by molecular dynamics simulations. *Biophysical journal* **72**, 2445-56.

Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M. and Obradovic, Z. (2002). Intrinsic disorder and protein function. *Biochemistry* **41**, 6573-82.

Dyson, H. J. and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature reviews. Molecular cell biology* **6**, 197-208.

Finn, R. N. (2007). Vertebrate yolk complexes and the functional implications of phosvitins and other subdomains in vitellogenins. *Biol Reprod* **76**, 926-35.

Fontana, A., de Laureto, P. P., Spolaore, B., Frare, E., Picotti, P. and Zambonin, M. (2004). Probing protein structure by limited proteolysis. *Acta Biochim Pol* **51**, 299-321.

Goddard, T. D. and Kneller, D. G. SPARKY 3. **University of California, San Francisco, USA**.

Havukainen, H., Halskau, Ø., Skjaerven, L., Smedal, B. and Amdam, G. V. (2011). Deconstructing honeybee vitellogenin: novel 40 kDa fragment assigned to its N-terminus. *Journal of Experimental Biology* **15**, 528-92.

Hawkins, C. J., Yoo, S. J., Peterson, E. P., Wang, S. L., Vernooy, S. Y. and Hay, B. A. (2000). The Drosophila caspase DRONC cleaves following glutamate or aspartate and is regulated by DIAP1, HID, and GRIM. *J Biol Chem* **275**, 27084-93.

Hoon Kim, D., Jeon Choi, S., Kook, S., Kim, W. and Keun Song, W. (2003). Phosphorylation-dependent cleavage of p130cas in apoptotic rat-1 cells. *Biochem Biophys Res Commun* **300**, 141-8.

Howard, M. B., Ekborg, N. A., Taylor, L. E., Hutcheson, S. W. and Weiner, R. M. (2004). Identification and analysis of polyserine linker domains in prokaryotic proteins with emphasis on the marine bacterium Microbulbifer degradans. *Protein Sci* **13**, 1422-5.

Hu, X., Lawrence, B., Kohler, K., Falick, A. M., Moore, A. M., McMullen, E., Jones, P. R. and Vierra, C. (2005). Araneoid egg case silk: a fibroin with novel ensemble repeat units from the black widow spider, Latrodectus hesperus. *Biochemistry* **44**, 10020-7.

Huntley, M. A. and Golding, G. B. (2004). Neurological proteins are not enriched for repetitive sequences. *Genetics* **166**, 1141-1154.

Huntley, M. A. and Golding, G. B. (2006). Selection and slippage creating serine homopolymers. *Mol Biol Evol* **23**, 2017-25.

Jain, E., Bairoch, A., Duvaud, S., Phan, I., Redaschi, N., Suzek, B. E., Martin, M. J., McGarvey, P. and Gasteiger, E. (2009). Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics* **10**, 136.

Kristoffersen, B. A., Nerland, A., Nilsen, F., Kolarevic, J. and Finn, R. N. (2009). Genomic and proteomic analyses reveal non-neofunctionalized vitellogenins in a basal clupeocephalan, the Atlantic herring, and point to the origin of maturational yolk proteolysis in marine teleosts. *Mol Biol Evol* **26**, 1029-44.

Li, J., Huang, J., Cai, W., Zhao, Z., Peng, W. and Wu, J. The vitellogenin of the bumblebee, Bombus hypocrita: studies on structural analysis of the cDNA and expression of the mRNA. *J Comp Physiol B* **180**, 161-70.

Linge, J. P., Habeck, M., Rieping, W. and Nilges, M. (2002). ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* **19**, 315-316.

Mann, C. J., Anderson, T. A., Read, J., Chester, S. A., Harrison, G. B., Kochl, S., Ritchie, P. J., Bradbury, P., Hussain, F. S., Amey, J. et al. (1999). The structure of vitellogenin provides a molecular model for the assembly and secretion of atherogenic lipoproteins. *J Mol Biol* **285**, 391-408.

Nelson, C. M., Ihle, K. E., Fondrk, M. K., Page, R. E. and Amdam, G. V. (2007). The gene vitellogenin has multiple coordinating effects on social organization. *PLoS Biol* **5**, e62.

Piulachs, M. D., Guidugli, K. R., Barchuk, A. R., Cruz, J., Simoes, Z. L. P. and Belles, X. (2003). The vitellogenin of the honey bee, Apis mellifera: structural analysis of the cDNA and expression studies. *Insect Biochemistry and Molecular Biology* **33**, 459-465.

Romans, P., Tu, Z., Ke, Z. and Hagedorn, H. H. (1995). Analysis of a vitellogenin gene of the mosquito, Aedes aegypti and comparisons to vitellogenins from other organisms. *Insect Biochem Mol Biol* **25**, 939-58.

Rouille, Y., Duguay, S. J., Lund, K., Furuta, M., Gong, Q., Lipkind, G., Oliva, A. A., Jr., Chan, S. J. and Steiner, D. F. (1995). Proteolytic processing mechanisms in the biosynthesis of neuroendocrine peptides: the subtilisin-like proprotein convertases. *Frontiers in neuroendocrinology* **16**, 322-61.

Raag, R., Appelt, K., Xuong, N. H. and Banaszak, L. (1988). Structure of the lamprey yolk lipid-protein complex lipovitellin-phosvitin at 2.8 A resolution. *J Mol Biol* **200**, 553-69.

Seehuus, S. C., Norberg, K., Gimsa, U., Krekling, T. and Amdam, G. V. (2006). Reproductive protein protects functionally sterile honey bee workers from oxidative stress. *Proc Natl Acad Sci U S A* **103**, 962-7.

Sharp, T. R., Morris, R., Horan, G. J., Pezzullo, L. H. and Stroh, J. G. (2006). Method for determining the average degree of substitution of o-vanillin derivatized porcine somatotropin. *J Pharm Biomed Anal* **40**, 185-9.

Silva-Neto, M. A., Fialho, E., Paes, M. C., Oliveira, P. L. and Masuda, H. (2002). Cyclic nucleotide-independent phosphorylation of vitellin by casein kinase II purified from Rhodnius prolixus oocytes. *Insect Biochem Mol Biol* **32**, 847-57.

Smolenaars, M. M., Madsen, O., Rodenburg, K. W. and Van der Horst, D. J. (2007). Molecular diversity and evolution of the large lipid transfer protein superfamily. *J Lipid Res* **48**, 489-502.

Stroh, J. G., Loulakis, P., Lanzetti, A. J. and Xie, J. (2005). LC-mass spectrometry analysis of N- and C-terminal boundary sequences of polypeptide fragments by limited proteolysis. *J Am Soc Mass Spectrom* **16**, 38-45.

Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* **Chapter 2**, Unit 2 3.

Trenczek, T. and Engels, W. (1986). Occurrence of Vitellogenin in Drone Honeybees (Apis-Mellifica). *International Journal of Invertebrate Reproduction and Development* **10**, 307-311.

Tufail, M. and Takeda, M. (2008). Molecular characteristics of insect vitellogenins. *J Insect Physiol* **54**, 1447-58.

Uversky, V. N., Oldfield, C. J. and Dunker, A. K. (2008). Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annual review of biophysics* **37**, 215-46.

Walter, J., Schindzielorz, A., Grunberg, J. and Haass, C. (1999). Phosphorylation of presenilin-2 regulates its cleavage by caspases and retards progression of apoptosis. *Proc Natl Acad Sci U S A* **96**, 1391-6.

Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. and Jones, D. T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* **337**, 635-45.

Wheeler, D. E. and Kawooya, J. K. (1990). Purification and characterization of honey bee vitellogenin. *Arch Insect Biochem Physiol* **14**, 253-67.

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D. and Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* **112**, 531-52.

Wishart, D. S., Sykes, B. D. and Richards, F. M. (1992). The Chemical-Shift Index - a Fast and Simple Method for the Assignment of Protein Secondary Structure through Nmr-Spectroscopy. *Biochemistry* **31**, 1647-1651.

Figure legends

Figure 1. Alignment of Hymenoptera vitellogenin polyserine tracts and their predicted disorder. (A) Sequence alignment of a selection of Vgs at the polyserine tract. The very long serine stretches of *A. rosae* were condensed (marked with slashes). The serine proportion of the sequence shown is indicated for each Vg. (B) Disopred (Ward et al., 2004) disorder prediction for selected sequences. Disordered parts are marked with asterisks, (*). (C) The loop colored with red in lamprey lipovitellin crystal structure visualizes the linker region between the conserved N-terminal b-barrel-like domain and the major lipid-binding cavity; the polyserine linker is located here in insects. Vertebrate Vgs, like lamprey lipovitellin, lack the polyserine tract. The location of the conserved sequence LMY (lamprey) / LVY (honeybee) is indicated (also the beginning of alignment 1A).

Figure 2. Limited proteolysis of honeybee Vg by trypsin and chymotrypsin. (A-B) The circles, fainter next to a faint band, emphasize the 150 and 40 kDa cleavage products of the full-length Vg (180 kDa). S = standard; Vg = pure, untreated vitellogenin. (A) The proteases do not prefer the polyserine tract. The most resistant bands (t/ct; arrow) were cut for LC-MS/MS analysis. A small C-terminal piece is easily cleaved, observed as doubling of both 180 and 150 kDa bands (two circles at 150 kDa). (B) When dephosphorylated, the unspecific cleavage pattern faints (see the arrow-marked band), but enhanced persistence of the 150 + 40 kDa pattern is observed (the red circles). Since full-length Vg (180) is digested as in the natural sample (A), the altered pattern can be explained by proteases switching to prefer the natural cleavage site. CIP = calf intestinal phosphatase. (C) LC-MS/MS-based identifications of the most proteolysis-resistant Vg region. The trypsin sample identifications are highlighted red, and the chymotrypsin identifications are underlined with red color. The trypsin/chymotrypsin cleavage sites at the polyserine region (grey) are indicated with open and filled triangles, respectively. The framed residues show the final MALDI-TOF hit of the natural 40 kDa and the first hit of the natural 150 kDa fragment; within this frame, the natural cleavage takes place.

Figure 3. *In vitro* phosphorylation of a *A. mellifera* polyserine tract peptide by casein kinase II. The peptide TDISSSSSSISSSEENDFWQPK was *in vitro* phosphorylated using casein kinase II (CKII) and up to a quintuple phosphorylation was monitored using LC-MS/MS analysis. Loss of phosphate from the parent mass (pm) and some diagnostic y ions are indicated in the spectra. It was observed that serine residues 11 and 12 are among the phosphorylated residues. Additional phosphorylations could not be located to a specific site.

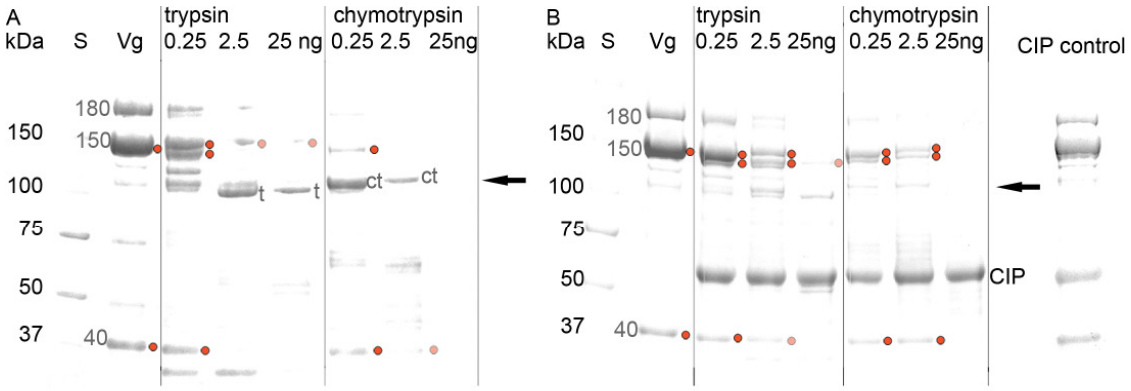
Figure 4. TOCSY and far-CD spectra of AmVg(358-392) polyserine peptide and NvVg(351-385) peptide. (A) A close-up of the TOCSY H-α-region region of AmVg, where also the collapsed H-β-peaks of the serine residues are visible at around 3.85 ppm. (B) The CD-spectrum of AmVg shows a typical curve of a random coil protein. (C) The TOCSY- and (D) CD-spectrum of NvVg indicate a more structured peptide compared to AmVg.

Figure 5. Graphical presentation of sequential (d_{NN}, $d_{\alpha N}$) and medium-range ($d_{\alpha N}$(i,i+1-4)) proton-proton NOE connectivities, chemical shift indices (CSI) and secondary structure prediction of the polyserine peptides. Thicker bars indicate stronger NOE intensities. The phosphorylated serine residues are indicated with PO_4 above the residue in question. The N15-labeled residues are shown in grey. Gaps in the CSI bar indicate residues whose α-proton was not applicable for CSI analysis, such as the phosphoserine residues, or was not identified with certainty. (A) Sequence alignment of the polyserine tract of honeybee and *N. vitripennis*. The synthesized fragments are marked. (B,C) *Apis mellifera* peptides: AmVg(358-392) (B) and AmVg(226-385) (C). (D,E) *Nasonia vitripennis* peptides: NvVg(351-385) (D) and NvVg(335-372) (E). The sequence location of the *N. vitripennis* helix between residues 371 and 377 is indicated (D, Observed).

Figure 6. Honeybee and jewel wasp polyserine tracts show structural differences. (A) The twenty calculated structures with lowest energy showing the outstretched random coil structure of the polyserine tract of AmVg(358-392). (B) An energetically minimized, averaged structure with surface electrostatics and with serine residues highlighted as thicker sticks. (C) The twenty calculated structures with lowest energy of NvVg(351-385), aligned at residues 364-377, show a compact architecture with a helix. (D) The minimized, averaged structure with surface electrostatics and serine residues highlighted.

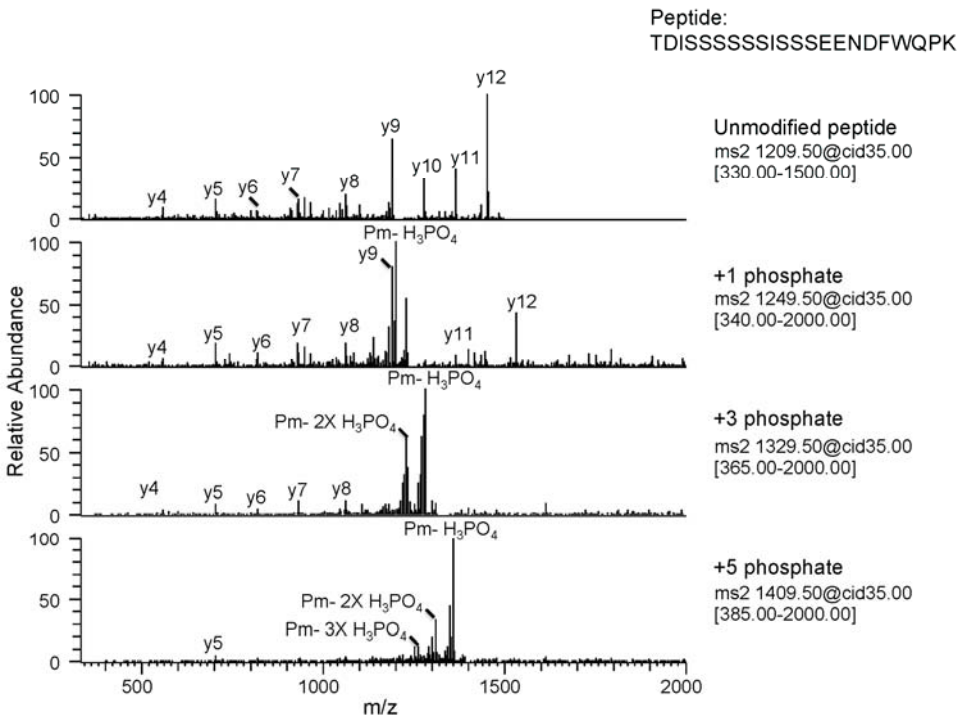
Table 1. NMR statistics of the solved peptides structures.

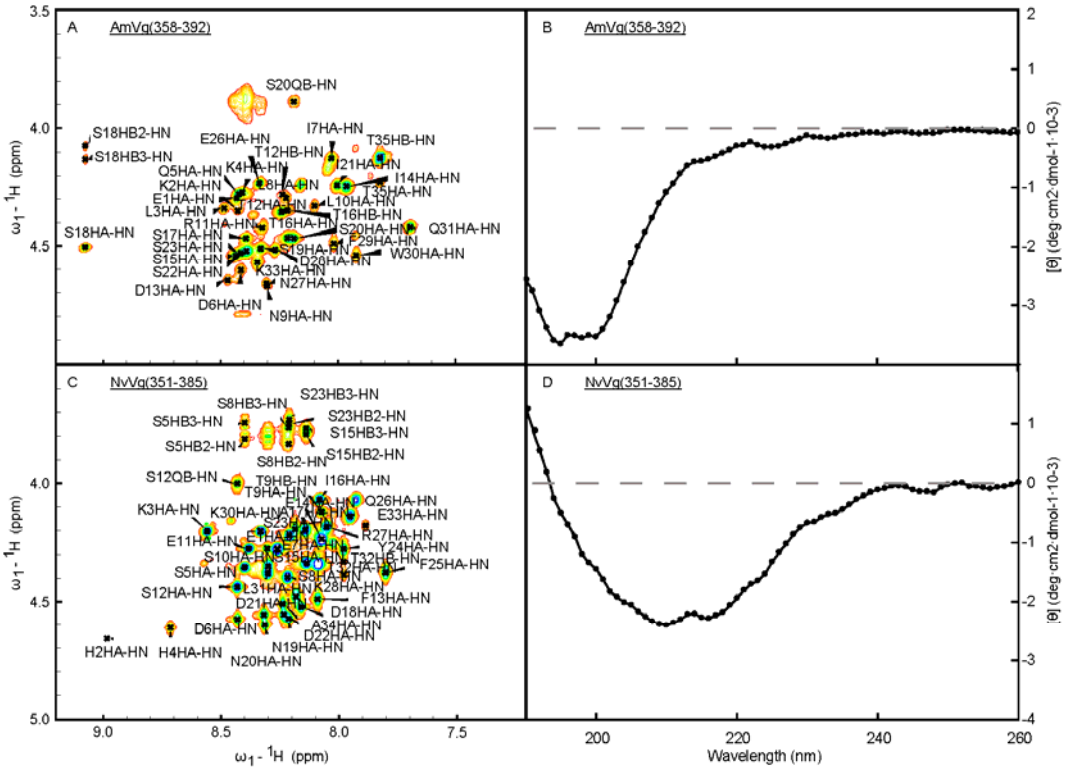

A
Suborder: Apocrita
Superfamily: Apoidea
Serine %
A.mellifera
B.ignitus
B.hypocrita
Superfamily: Chalcidoidea
N.vitripennis
P.puparum
E.formosa
Superfamily: Ichneumonoidea
P.nipponica
Superfamily: Formicidea
S.invictaVg-2
S.invictaVg-3
S.invictaVg-1
H.saltator
C.floridanus
Suborder: Symphyta
Superfamily: Tenthredinoidea
A. rosae
B
A. mellifera
B. ignitus
N. vitripennis
P. nipponica
S. invictaVg-2
H. saltator
A. rosae
C
LMY / LVY

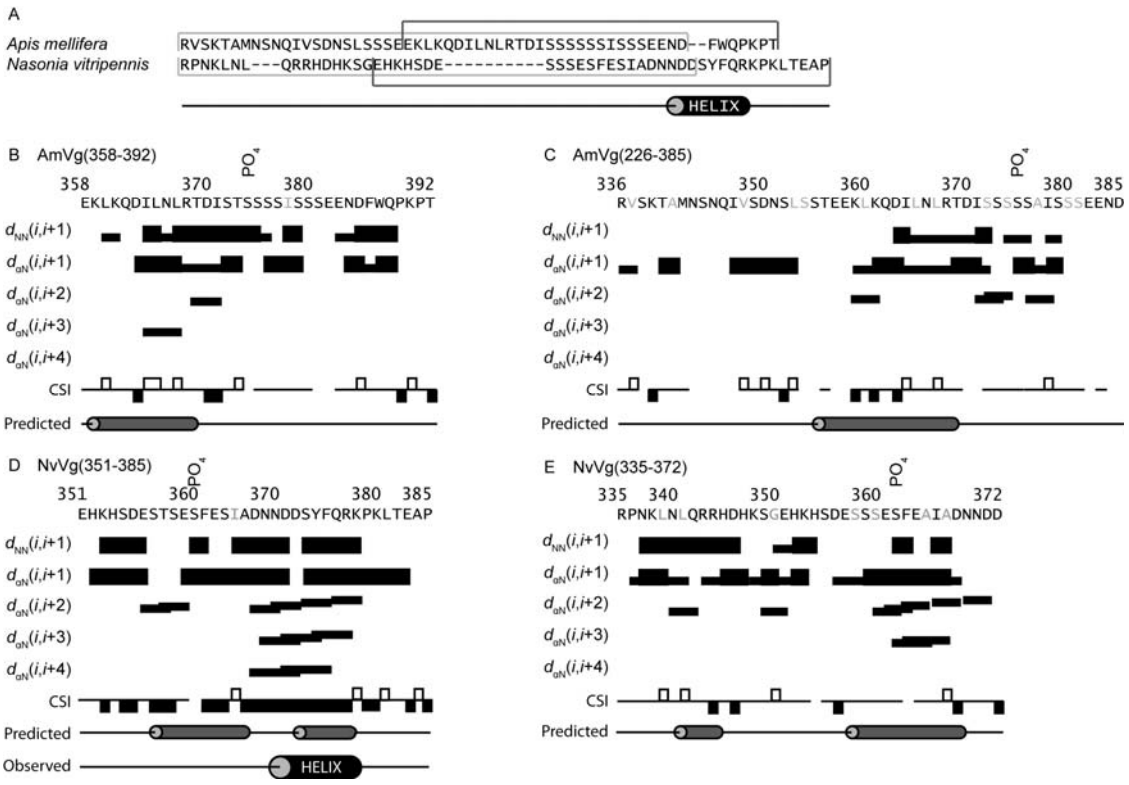


C

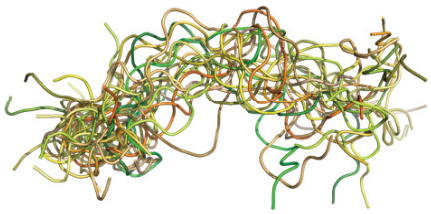
```
   1 MLLLLTLLLF AGTVAADFQH NWQVGNEYTY LVRSRTLTSL GDLSDVHTGI LIKALLTVQA KDSNVLAAKV
  71 WNGQYARVQQ SMPDGWETEI SDQMLELRDL PISGKPFQIR MKHGLIRDLI VDRDVPTWEV NILKSIVGQL
 141 QVDTQGENAV KVNSVQVPTD DEPYASFKAM EDSVGGKCEV LYDIAPLSDF VIHRSPELVP MPTLKGDGRH
 211 MEVIKIKNFD NCDQRINYHF GMTDNSRLEP GTNKNGKFFS RSSTSRIVIS ESLKHFTIQS SVTTSKMMVS
 281 PRLYDRQNGL VLSRMNLTLA KMEKTSKPLP MVDNPESTGN LVYIYNNPFS DVEERRVSKT AMNSNQIVSD
 351 NSLSSSEEKL KQDILNLRTD ISSSSSSISS SEENDFWQPK PTLEDAPQNS LLPNFVGYKG KHIGKSGKVD
 421 VINAAKELIF QIANELEDAS NIPVHATLEK FMILCNLMRT MNRKQISELE SNMQISPNEL KPNDKSQVIK
 491 QNTWTVFRDA ITQTGTGPAF LTIKEWIERG TTKSMEAANI MSKLPKTVRT PTDSYIRSFF ELLQNPKVSN
 561 EQFLNTAATL SFCEMIHNAQ VNKRSIHNNY PVHTFGRLTS KHDNSLYDEY IPFLERELRK AHQEKDSPRI
 631 QTYIMALGMI GEPKILSVFE PYLEGKQQMT VFQRTLMVGS LGKLTETNPK LARSVLYKIY LNTMESHEVR
 701 CTAVFLLMKT NPPLSMLQRM AEFTKLDTNR QVNSAVKSTI QSLMKLKSPE WKDLAKKARS VNHLLTHHEY
 771 DYELSRGYID EKILENQNII THMILNYVGS EDSVIPRILY LTWYSSNGDI KVPSTKVLAM ISSVKSFMEL
 841 SLRSVKDRET IISAAEKIAE ELKIVPEELV PLEGNLMINN KYALKFFPFD KHILDKLPTL ISNYIEAVKE
 911 GKFMNVNMLD TYESVHSFPT ETGLPFVYTF NVIKLTKTSG TVQAQINPDF AFIVNSNLRL TFSKNVQGRV
 981 GFVTPFEHRH FISGIDSNLH VYAPLKISLD VNTPKGNMQW KIWPMKGEEK SRLFHYSVVP FVSNHDILNL
1051 RPLSMEKGTR PMIPDDNTSL ALPKNEGPFR LNVETAKTNE EMWELIDTEK LTDRLPYPWT MDNERYVKVD
1121 MYMNLEGEQK DPVIFSTSFD SKVMTRPDTD SENWTPKMMA VEPTDKQANS KTRRQEMMRE AGRGIESAKS
1191 YVVDVRVHVP GESESETVLT LAWSESNVES KGRLLGFWRV EMPRSNADYE VCIGSQIMVS PETLLSYDEK
1261 MDQKPKMDFN VDIRYGKNCG KGERIDMNGK LRQSPRLKEL VGATSIIKDC VEDMKRGNKI LRTCQKAVVL
1331 SMLLDEVDIS MEVPSDALIA LYSQGLFSLS EIDNLDVSLD VSNPKNAGKK KIDVRAKLNE YLDKADVIVN
1401 TPIMDAHFKD VKLSDFGFST EDILDTADED LLINNVFYED ETSCMLDKTR AQTFDGKDYP LRLGPCWHAV
1471 MTTYPRINPD NHNEKLHIPK DKSVSVLSRE NEAGQKEVKV LLGSDKIKFV PGTTSQPEVF VNGEKIVVSR
1541 NKAYQKVEEN EIIFEIYKMG DRFIGLTSDK FDVSLALDGE RVMLKASEDY RYSVRGLCGN FDHDSTNDFV
1611 GPKNCLFRKP EHFVASYALI SNQCEGDSLN VAKSLQDHDC IRQERTQQRN VISDSESGRL DTEMSTWGYH
1681 HNVNKHCTIH RTQVKETDDK ICFTMRPVVS CASGCTAVET KSKPYKFHCM EKNEAAMKLK KRIEKGANPD
1751 LSQKPVSTTE ELTVPFVCKA
```


Peptide:
TDISSSSSSISSSEENDFWQPK
Unmodified peptide
ms2 1209.50@cid35.00
[330.00-1500.00]
+1 phosphate
ms2 1249.50@cid35.00
[340.00-2000.00]
+3 phosphate
ms2 1329.50@cid35.00
[365.00-2000.00]
+5 phosphate
ms2 1409.50@cid35.00
[385.00-2000.00]
Relative Abundance
m/z
Pm- H3PO4
Pm- 2X H3PO4
Pm- 3X H3PO4

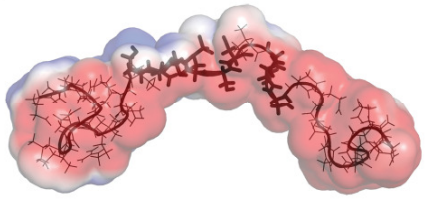

A AmVq(358-392)
B AmVq(358-392)
C NvVq(351-385)
D NvVq(351-385)
S20QB-HN
S18HB2-HN
S18HB3-HN
E26HA-HN
I7HA-HN
T35HB-HN
T12HB-HN
I21HA-HN
Q5HA-HN
K4HA-HN
I14HA-HN
K2HA-HN
T35HA-HN
E1HA-HN
L10HA-HN
L3HA-HN
T16HB-HN
R11HA-HN
T16HA-HN
S20HA-HN
Q31HA-HN
S18HA-HN
S17HA-HN
F29HA-HN
S23HA-HN
S19HA-HN
W30HA-HN
S15HA-HN
D28HA-HN
S22HA-HN
K33HA-HN
N27HA-HN
D13HA-HN
D6HA-HN
N9HA-HN
S8HB3-HN
S23HB3-HN
S5HB3-HN
S23HB2-HN
S15HB3-HN
S5HB2-HN
S15HB2-HN
S8HB2-HN
S12QB-HN
T9HB-HN
I16HA-HN
T9HA-HN
Q26HA-HN
K3HA-HN
K30HA-HN
E33HA-HN
R27HA-HN
E11HA-HN
Y24HA-HN
S10HA-HN
T32HB-HN
F25HA-HN
S5HA-HN
S12HA-HN
L31HA-HN
K28HA-HN
F13HA-HN
D21HA-HN
D18HA-HN
D6HA-HN
A34HA-HN
D22HA-HN
H2HA-HN
H4HA-HN
N19HA-HN
N20HA-HN
ω1 - 1H (ppm)
[θ] (deg·cm2·dmol-1·10-3)
Wavelength (nm)


A
Apis mellifera
Nasonia vitripennis
RVSKTAMNSNQIVSDNSLSSSEEKLKQDILNLRTDISSSSSSISSSEEND--FWQPKPT
RPNKLNL---QRRHDHKSGEHKHSDE----------SSSESFESIADNNDDSYFQRKPKLTEAP
HELIX
B AmVg(358-392)
PO4
358 370 380 392
EKLKQDILNLRTDISTSSSSSSSEENDFWQPKPT
dNN(i,i+1)
dαN(i,i+1)
dαN(i,i+2)
dαN(i,i+3)
dαN(i,i+4)
CSI
Predicted
C AmVg(226-385)
PO4
336 350 360 370 380 385
RVSKTAMNSNQIVSDNSLSSTEEKLKQDILNLRTDISSSSSAISSSEEND
dNN(i,i+1)
dαN(i,i+1)
dαN(i,i+2)
dαN(i,i+3)
dαN(i,i+4)
CSI
Predicted
D NvVg(351-385)
PO4
351 360 370 380 385
EHKHSDESTSESFESIADNNDDSYFQRKPKLTEAP
dNN(i,i+1)
dαN(i,i+1)
dαN(i,i+2)
dαN(i,i+3)
dαN(i,i+4)
CSI
Predicted
Observed
HELIX
E NvVg(335-372)
PO4
335 340 350 360 372
RPNKLNLQRRHDHKSGEHKHSDESSSESFEAIADNNDD
dNN(i,i+1)
dαN(i,i+1)
dαN(i,i+2)
dαN(i,i+3)
dαN(i,i+4)
CSI
Predicted

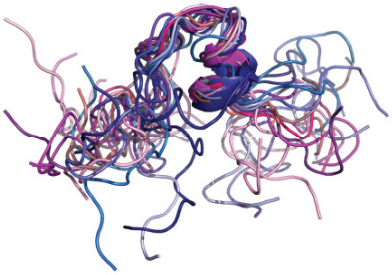
A *Apis mellifera*



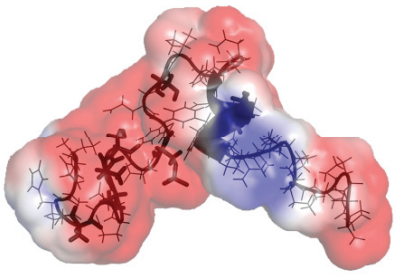
B



C *Nasonia vitripennis*



D

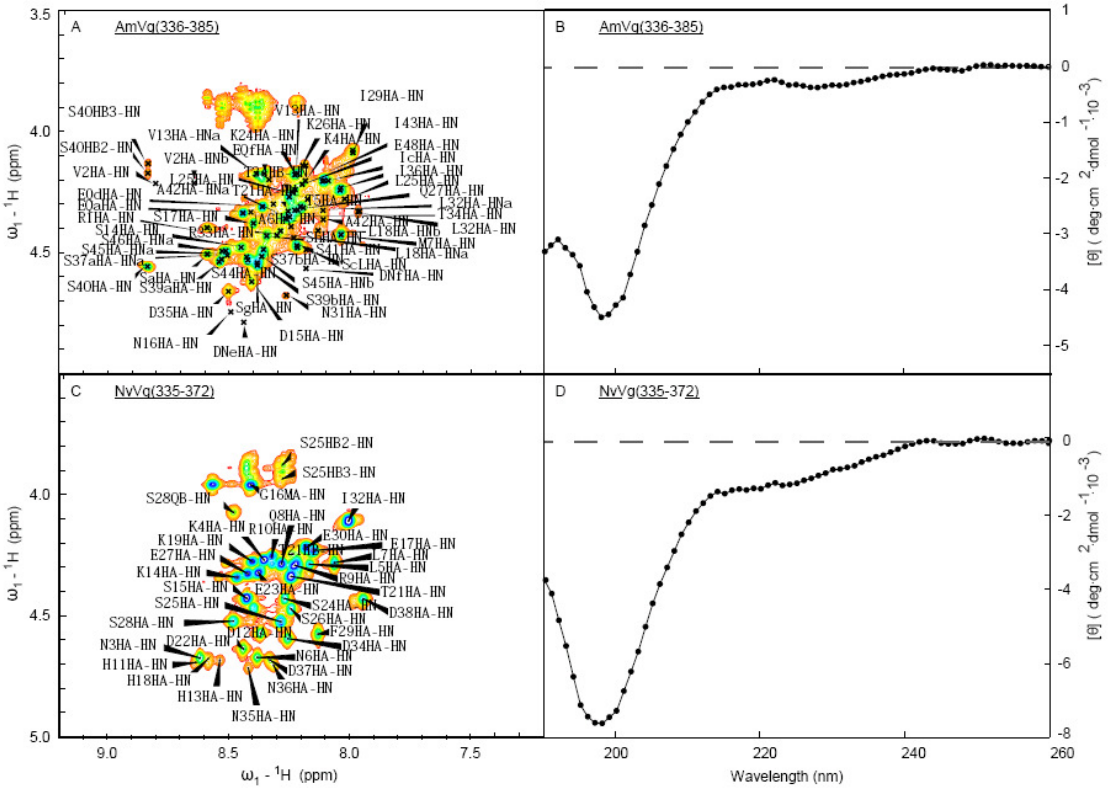


	AmVg(358-392)	NvVg(351-385)
Medium-range NOEs	8	39.0
Long-range NOEs	2	19.0
RMSD, backbone	3.4 ± 0.53	1,98 ± 0.41
RMSD, all residues	8.34 ± 1.33	5.8 ± 1.33
NOE violations > 0.3 Å	0	0
NOE violations > 0.1 Å	2.00 ± 1.45	3.7 ± 1.71

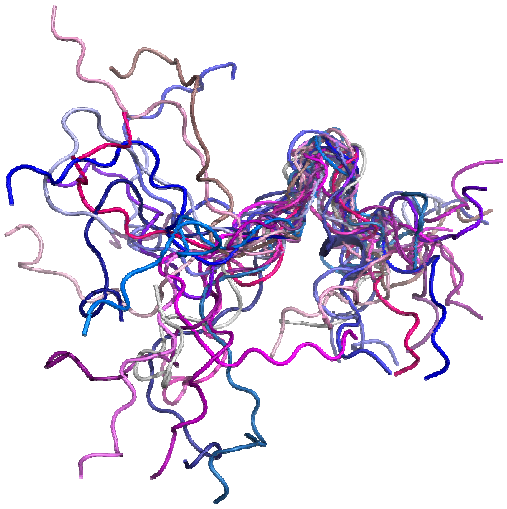
Supporting information

```
   1 mlllltlllf agtvaadfqh nwqvgneyty lvrsrtltsl gdlsdvhtgi likalltvqa
  61 kdsnvlaakv wngqyarvqq smpdgwetei sdqmlelrdl pisgkpfqir mkhglirdli
 121 vdrdvptwev nilksivgql qvdtqgenav kvnsvqvptd depyasfkam edsvggkcev
 181 lydiaplsdf vihrspelvp mptlkgdgrh mevikiknfd ncdqrinyhf gmtdnsrlep
 241 gtnkngkffs rsstsrivis eslkhftiqs svttskmmvs prlydrqngl vlsrmnltla
 301 kmektskplp mvdnpestgn lvyiynnpfs dveerrvskt amnsnqivsd nslssseekl
 361 kqdilnlrtd isssssssiss seendfwqpk ptledapqns llpnfvgykg khigksgkvd
 421 vinaakelif qianeledas nipvhatlek fmilcnlmrt mnrkqisele snmqispnel
 481 kpndksqvik qntwtvfrda itqtgtgpaf ltikewierg ttksmeaani msklpktvrt
 541 ptdsyirsff ellqnpkvsn eqflntaatl sfcemihnaq vnkrsihnny pvhtfgrlts
 601 khdnslydey ipflerelrk ahqekdspri qtyimalgmi gepkilsvfe pylegkqqmt
 661 vfqrtlmvgs lgkltetnpk larsvlykiy lntmeshevr ctavfllmkt npplsmlqrm
 721 aeftkldtnr qvnsavksti qslmklkspe wkdlakkars vnhllthhey dyelsrgyid
 781 ekilenqnii thmilnyvgs edsviprily ltwyssngdi kvpstkvlam issvksfmel
 841 slrsvkdret iisaaekiae elkivpeelv plegnlminn kyalkffpfd khildklptl
 901 isnyieavke gkfmnvnmld tyesvhsfpt etglpfvytf nvikltktsg tvqaqinpdf
 961 afivnsnlrl tfsknvqgrv gfvtpfehrh fisgidsnlh vyaplkisld vntpkgnmqw
1021 kiwpmkgeek srlfhysvvp fvsnhdilnl rplsmekgtr pmipddntsl alpknegpfr
1081 lnvetaktne emwelidtek ltdrlpypwt mdneryvkvd mymnlegeqk dpvifstsfd
1141 skvmtrpdtd senwtpkmma veptdkqans ktrrqemmre agrgiesaks yvvdvrvhvp
1201 gesesetvlt lawsesnves kgrllgfwrv emprsnadye vcigsqimvs petllsydek
1261 mdqkpkmdfn vdirygkncg kgeridmngk lrqsprlkel vgatsiikdc vedmkrgnki
1321 lrtcqkavvl smlldevdis mevpsdalia lysqglfsls eidnldvsld vsnpknagkk
1381 kidvraklne yldkadvivn tpimdahfkd vklsdfgfst edildtaded llinnvfyed
1441 etscmldktr aqtfdgkdyp lrlgpcwhav mttyprinpd nhneklhipk dksvsvlsre
1501 neagqkevkv llgsdkikfv pgttsqpevf vngekivvsr nkayqkveen eiifeiykmg
1561 drfigltsdk fdvslaldge rvmlkasedy rysvrglcgn fdhdstndfv gpknclfrkp
1621 ehfvasyali snqcegdsln vakslqdhdc irqertqqrn visdsesgrl dtemstwgyh
1681 hnvnkhctih rtqvketddk icftmrpvvs casgctavet kskpykfhcm ekneaamklk
1741 kriekganpd lsqkpvstte eltvpfvcka
```

Supplement 1. MALDI-TOF peptide hits of vitellogenin 40 and 150 kDa fragments. The 40 kDa fragment peptide hits are red and bolded; the 150 kDa fragment peptide hits are red and underlined. The polyserine area at the intersection of these two fragments is highlighted grey.



Supplement 2, Figure 1. TOCSY and far-CD spectra of AmVg(336-385) polyserine peptide and the corresponding NvVg(335-372) peptide. (A) A close-up of the TOCSY H-α-region region of AmVg, where also the collapsed H-β-peaks of the serine residues are visible at around 3.85 ppm. (B) The CD-spectrum of AmVg shows a typical curve of a random coil protein. (C) The TOCSY- and (D) CD-spectrum of NvVg indicate a random coil structure.



Supplement 2, Figure 2. The twenty calculated NMR structures of NvVg(335-372) peptide aligned at residues 349-364 that have the lowest RMSD. The structure is a collapsed random coil, although some weak helical tendency is present at residues 361-364.

Supplement 2, Table 1. NMR statistics of the NvVg(335-372) peptide.

	NvVg(335-372)
Medium-range NOEs	28.5
Long-range NOEs	4
RMSD, backbone	3.17 ± 0.42
RMSD, all residues	9.59 ± 1.57
NOE violations > 0.3 Å	0
NOE violations > 0.1 Å	0.55 ± 0.80

Paper V

Membrane binding of vitellogenin

Havukainen, H., Münch, D., Baumann, A., Halskau, Ø., Amdam, G. V.
Manuscript.

Abstract

Vitellogenin (Vg) is a near universal egg-yolk protein. It is also a central life-history regulator in honey bees where Vg supports immune cell viability, protects against oxidative stress, and suppresses risky foraging behavior. The molecular mechanisms of these effects are not well understood. In fish, Vg is an antibacterial protein, and homologous lipoproteins act in human inflammation responses. In general, lipoproteins have varied roles affiliated with animal immune systems, several of which depend on interactions with self- and non-self hydrophobic portions (hyppos). We therefore tested whether such interactions could also include direct membrane association in Vg. Using immunodetection approaches, we first demonstrated that honey bee Vg is present at cell membranes and intracellular vesicles in bee tissues, and we documented binding of Vg to insect cells. Direct binding to bilayers was verified with a membrane mimicking system (liposomes). These experiments located the binding ability largely to an evolutionarily conserved α-helical domain, of which we prepared a homology model. With a fluorescence leakage assay we further tested whether binding of Vg on negatively charged membrane could provoke membrane disruption, as demonstrated for several antibacterial polypeptides. This test and a follow-on antibacterial assay toward *Escherichia coli* were both negative. Thus, we suggest that instead of a destructive binding to non-self membranes, Vg may bind to honey bee membranes to confer protective effects. This prediction is consistent with the positive influence of Vg on cell function and survival in honey bees.

Heli Havukainen[1,2], Daniel Münch[1], Anne Baumann[2], Øyvind Halskau[3], Gro V. Amdam[1,4]

[1]Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, P.O. Box 5003 1432 Aas, Norway
[2]Department of Biomedicine, University of Bergen, Jonas Lies vei 91, 5009 Bergen, Norway
[3]Department of Molecular Biology, University of Bergen, Thormøhlensgate 55, 5020 Bergen, Norway
[4]School of Life Sciences, Arizona State University, P.O. Box 874501, Tempe, AZ 85287-4501, USA

Abbreviations:

LLTP	large lipid transfer proteins
Vg	vitellogenin
apoB	apolipoprotein B
MTP	microsomal triglyceride transfer protein
hlVg	vitellogenin purified from hemolymph
fbVg	vitellogenin purified from fatbody
IR	immunoreactivity
vWFD	von Willebrand Factor D
SEC	size-exclusion chromatography
P	porcine brain phosphatidylserine
E	egg yolk phosphatidylcholine
ANTS	8-aminonaphtalene-1,3,6-trisulphonic acid
DLS	dynamic light scattering
SPR	surface plasmon resonance
DPX	p-xyelene-bis-N-pyrimidium bromide
PAMP	pathogen-associated molecular pattern
LDL	low-density lipoprotein

Introduction

Vitellogenin (Vg) is a lipoprotein that functions as a yolk precursor in oviparous animals, including insects and fish (1,2). The main production site of Vg is the fatbody in insects and the liver in vertebrates (3). In insects, Vg can be stored in the trophocyte cells of the fatbody and is excreted to the hemolymph (blood) for transport to target tissues (4,5). In vertebrates and insects, Vg is taken up from the circulation by oocytes through receptor-mediated endocytosis (reviewed by (3,6)).

In the honey bee (*Apis mellifera*), an advanced social insect, the Vg receptor is expressed strongly in the ovaries of reproductive queens, but also at lower levels in the fatbody, the hypopharyngeal glands and the midgut of essentially sterile female helpers called workers (7). Non-reproductive functions of Vg are documented for queen and worker bees, and may also influence males (8); these are behavioral and physiological effects that improve survival. For example, Vg can suppress foraging behavior, a risky activity that accelerates senescence in workers (9,10). High Vg levels further correlate with improved immune cell survival *in vitro* and improved immune function *in vivo* (11,12), and honey bee Vg protects both queens and workers against damage from oxidative stress-inducing agents (13). Excessive buildup of Vg in workers (up to 100 mg/ml in hemolymph (14)) is characteristic of an extremely long-lived and stress resistant worker phenotype called winter (*diutinus*) bees (15). Yet, although honey bee Vg is a molecularly, physiologically, ecologically, and evolutionary interesting protein (16,17), little is known about how it confers its effects.

Vgs belong to a family of large lipid transfer proteins (LLTP) which includes human microsomal triglyceride transfer protein (MTP) and apolipoprotein B (apoB) (18,19). All LLTPs are involved in lipid transport, but these proteins have additional functions in immunity. Human MTP regulates the maturation of CD1 molecules that present lipid antigens to T-cells (20,21). Human apoB is an inflammation responsive protein (22), which recognizes pathogen-associated molecular patterns (PAMPs) such as lipopolysaccharides (23). In several fish, including salmon (*Salmo salar*) (24,25) and zebrafish (*Danio rerio*) (26), Vg recognizes PAMPs and has antibacterial, antiviral and antifungal effects (reviewed by (27)).

The antibacterial effect of Vg in fish is thought to rely on at least two mechanisms: Vg damages the cell wall of bacteria (28) and opsonizes bacteria for phagocytosis (27). Another generally well-known antibacterial mechanism is the disruption of the negatively charged bacterial cell membrane by a positively charged peptide (reviewed by (29)). For example, several human lipoproteins, including apoB, have amphipathic α-helical stretches with positively charged residues that can disrupt bacterial membranes (30).

Many proteins have non-disruptive affinity toward self-membranes in the host organism. These interactions can be linked to diverse functions, from signaling to transport, and depend on membrane-bound receptors or act via direct lipid binding (31). Direct membrane binding of Vg to somatic cells has not been studied in detail, but for apoB there is evidence of direct binding to epithelial cell membrane (32) and to the membrane of the endoplasmic reticulum (33,34) via its Vg domain, which shares homology with the N-sheet and the α-helical domain of Vgs (19,35). In general, human lipoproteins show affinity for hydrophobic portions, hyppos, which PAMPs and immunostimulatory self-derived hydrophobic molecules belong to (reviewed by (36)). Because Vgs belong to the LLTP family, they might share this general affinity toward hydrophobic molecules, as exemplified by the PAMP binding of Vg in fish (27).

In this study, we report membrane association of honey bee Vg. Membrane binding was detected *in vitro* in bee tissues by immunohistochemistry and immunoblotting, and we observed Vg adherence toward the insect cells of *Spodoptera frugiperda* (Sf9) cell line. We questioned the interpretation that these observations could be explained purely by Vg-receptor interaction (7) and used a phospholipid membrane-mimicking system – liposomes – to test direct binding to bilayers. Binding to both neutral and negatively charged bilayers was confirmed, and we identified the amphipathic α-helical domain of Vg as largely responsible for the binding.

Since antibacterial effects of Vg are insufficiently studied in the bee this far, we next tested for binding of Vg to *Escherichia coli*, which was confirmed. Thereafter, we investigated whether such interaction could cause disruption of negatively charged (as are bacterial membranes) liposomes, a mechanism which could be similar to that of the antibacterial α-helical stretches in human lipoproteins and add to those mechanisms reported for Vg in fish. However, no such disruptive action was found for negative

liposomes in the presence of Vg. Using *E. coli*, we followed up with an antibacterial test, which corroborated that honey bee Vg might lack general bactericidal potential. We conclude that this Vg is unlikely to combat bacteria directly. Since Vg is known for life-prolonging effects in honey bees, we propose an alternative explanation: The membrane adherence of Vg serves protective functions in honey bee tissues.

This study is the first to describe direct somatic membrane binding of any Vg, and to identify a domain involved in the binding. It remains to be determined whether our findings reveal properties that are specific to the Vg of honey bees, or are shared with related species, or that are of broad phylogenetic importance. It was, however, recently confirmed that regions of Vg that include its α-helical domain are under strong positive selection in the genus *Apis* (37). The specific pattern of positive selection indicates that Vg protein evolution contributes a survival benefit to several species in this phylogenetic group. Generally, our findings corroborate that LLTPs are committed to a broad range of interactions and are constitutively affiliated with somatic maintenance systems.

Results

1. Immunodetection of Vg at biological membranes

We found Vg in the membrane fraction of proteins extracted from abdomen carcasses of honey bee workers (Fig. 1A; the abdominal carcass consists of the cuticular wall with adhering fatbody, plus minor amounts of muscles, trachea, nerves and glands). Then, we tested Vg-cell binding using the insect Sf9 cell line as model target, and the protein was again found to bind to these cells (Fig. 1B-C). Both the membranes and the Sf9 cells were washed thoroughly before blotting to avoid contamination by unbound Vg. In addition to the full-length protein (180 kDa), the immunoblots (Fig. 1A-C) detected a 150 kDa Vg fragment. This fragment has been previously identified as the protein's major lipid-binding cavity, including the α-helical domain of Vg (38).

We next asked if Vg immunoreactivity (VgIR) in anatomical tissue sections supports Vg accumulation at membranes. The confocal micrographs shown in Figure 2 represent sections of N = 5 individuals for both abdominal (Fig. 2A-C, G-L) and head tissue

samples (Fig. 2D-F). As previously described (e.g., (38)), we found intense VgIR in Vg rich vesicles of abdominal fatbody trophocytes (Fig. 2A-C). Adding to this, our high-resolution images reveal intense VgIR close to vesicular membranes in a subset of these vesicles (Fig. 2B). As expected for an abundant hemolymph protein, VgIR was scattered across different sites (Fig. 2C, E, F) apart from the known localization within fatbody (Fig. 2A-C) and hypopharyngeal glands (Fig. 2D). Intense VgIR was detected at boundary layers, e.g. between single cells (Fig. 2D) or between syncitial cells of muscle fibers (Fig. 2D, F). We conclude that high VgIR is found at both vesicular and cellular membranes, indicative of Vg accumulation at different membranous compartments. Controls for auto-fluorescence, detector noise and unspecific labeling by the secondary antibody were negative (Fig. 2H, J, L for N = 5), as compared to test samples recorded with similar confocal scan settings (Fig. 2G, I, K for N = 5).

2. Vg binding on phospholipid bilayers *in vitro*

The VgIR we localized to honey bee Vg membranes *in vivo* might be explained by Vg interacting with its membrane-bound receptor, but Vg was also observed on muscle cells that are not a suggested target for Vg or a site of Vg receptor expression. We therefore tested whether receptor-free binding could occur. Surface plasmon resonance (SPR) allows us to quantify a membrane binding response using L1 sensor chips designed for immobilizing lipid bilayers in the form of liposomes. A characteristic for proteins displaying intrinsic affinities for receptor-free lipid membranes is a heightened affinity for negatively charged bilayers relative to neutral, zwitterionic counterparts (31). Therefore, we prepared two liposome systems from 100% egg-yolk phosphatidylcholine (E), and E in 1:1 mixture with porcine brain phosphatidylserine (P; P:E). These two lipid extracts are zwitterionic (neutral) and negatively charged, respectively, at the physiological conditions employed here.

Vg affinities for E and P:E were tested using the following native protein samples: full-length Vg purified from hemolymph (hlVg) and predominantly fragmented (150 + 40 kDa fragments) Vg purified from fatbody (fbVg; the 150 kDa band is visible in Fig. 1A). Also, two domains produced in *E. coli* were used: α-helical domain (α-helical) and

vWFD-domain (vWFD), which both localize to the 150 kDa fragment of Vg. While the vWFD-domain sample showed no binding (Fig. 3), the sensograms of all the other samples indicate that some of the bound protein was permanently associated with the membrane (Fig. 3A). The bound Vg samples showed affinity toward both the neutral E and the negatively charged P:E bilayers, with higher binding response for the latter (Fig. 3B). The cleaved fatbody derived Vg had lower affinity toward both bilayers than the full-length Vg. The α-helical domain showed a similar response to the cleaved native Vg (Fig. 3B), in particular when binding to the negatively charged bilayer.

3. Identification of a candidate bilayer-binding region

Liposome-combined limited proteolysis can reveal which part of the protein enters the lipid bilayer, since the lipid-embedded region can be shielded against proteolytic digestion (39). In our liposome-combined proteolysis experiment, one segment of the protein was less degradable in the presence of both neutral and negatively charged liposomes (Fig. 4A). Titration digestions with smaller amount of chymotrypsin show that liposomes as such do not affect the digestion rate (Fig. 4B). We found that the amount of the less degradable fragment of Figure 4A, measured as its band intensities in the liposome samples, was significantly higher than control intensities from the corresponding gel region of digestions where no liposomes were provided (Kruskal-Wallis ANOVA: $H = 10.75$, $p = 0.005$; Fig. 4C). The net charge of the liposomes (E or P:E) had no effect on the fragment intensity (Mann-Whitney post-hoc test: $Z = 0.16$, $p = 0.873$; Fig. 4C).

As an additional reference test for these analyses, we chose two more Vg fragments in the range of 30 kDa (Fig. 4A). We measured the intensity of these reference bands, and ran them through the same statistical analysis (Fig. 4C). The reference samples showed no significant difference between the control and the liposome digestions (Kruskal-Wallis ANOVA $H = 0.047$, $p = 0.977$; Fig. 4C).

The protease resistant ~120 kDa Vg segment was identified with LC-MSMS. The LC-MSMS hits span amino acid residues 135-1613 (Supplement 1). This region contains the α-helical domain (residues 416 - 765).

4. The α-helical Vg domain is amphipathic with an overall positive charge

Our homology model (Fig. 5) contains a total of 49 positively charged residues, 34 of which point directly toward the solvent (Fig. 5A). The side facing the lipid-binding cavity is less charged (Fig. 5B). The secondary structure prediction program PSIPRED (40) assigns the domain exclusively to helices with connective loops that mach with our model (data not shown). Residues 579-604 could not be modeled, since honey bee Vg has additional sequence compared to the template (lamprey, *Ichthyomyzon unicuspis,* lipovitellin); PSIPRED prediction suggests that this missing sequence, which we assign insect-specific (Fig. 5C), is a random coil.

5. Vg interacts with *E. coli*, but does not disrupt negatively charged liposomes or kill *E. coli*

Since binding on negatively charged membranes can be linked to antibacterial activity, we tested Vg-bacteria interaction. Hemolymph and fatbody protein extract were incubated with gram-negative *E. coli*, and unbound protein was removed by extensive washing. The bacteria were lysed and all protein was solubilized. Vg antibody recognized Vg in the *E. coli* samples that had been incubated with hemolymph or fatbody extract (Fig. 6A). Some antibacterial proteins act through direct interaction with the lipids of the membrane causing membrane disruption (41,42). Therefore, we tested the effect of hemolymph and fatbody purified Vg on the integrity of a negatively charged bilayer by an extrinsic fluorescence leakage assay with ANTS-DPX-filled P:E liposomes. The fluorescence signal was not notably changed until the final 2.5 μM protein concentration (the same as used in SPR), thus, Vg mediated membrane disruption was not observed (Fig. 6B). The assay was controlled with Triton X-100, a detergent that causes 100% leakage. Finally, we tested with a plating assay adapted from a fish study on Vg (24),

whether honey bee Vg can have general antibacterial properties. We did not observe an antibacterial effect of purified Vg on *E. coli* (Fig. 6C, Supplement 2). Six replicates were run before the significance of the total 22 samples was calculated (t-value = 1.745, p = 0.0875, see Supplement 2 for the details on the data).

Discussion

The major contribution of this work is the new insights on the membrane binding ability of a Vg. In current literature, only one study documents a Vg-membrane association, and this interaction is with reproductive, not somatic cells: in the amphibian *Discoglossus pictus*, lipovitellin (i.e. cleaved Vg from eggs) appears at the surface of each egg at a particular sperm binding site, where lipovitellin is involved in sperm-egg interaction (43). In contrast, honey bee Vg has roles beyond the egg, and we believe that separate regions of the protein are largely responsible for its reproductive versus somatic roles. The N-terminal domain is the predicted receptor binding site of Vgs and presumably essential for the uptake of the protein by oocytes (44). This domain, however, is absent from the 150 kDa fragment of Vg (38) that we found in the membrane fraction of bee abdomens, that associates with Sf9 cells (Fig. 1), and that binds to liposomes (Fig. 3; Fig. 4). These observations suggest that Vg is able to bind on membranes without receptor interaction. In honey bees, the N-terminal domain of Vg is under strong negative selection (i.e. mutational change is not permitted) while the remaining protein has been freer to evolve (37). In fact, the 150 kDa Vg fragment – the lipid-binding cavity of the lipoprotein – has been (and still is) under considerable positive selection, not only in honey bees but in several *Apis* species (37). We speculate that this evolution of Vg has facilitated an affinity to somatic cell membranes, as we here document accumulation of Vg at such membranes, including those of muscle cells.

In our SPR study, the α-helical domain alone could bind to receptor-free lipid bilayers with notable affinity (for comparative protein-liposome SPR values, see e.g. (45)), and vWFD was excluded as a non-membrane binding domain. The α-helical domain, as noted above, is under positive selection in bees (37) but still well-conserved in the LLTP

family compared to the generally low sequence identity of the lipid-binding cavity (35). We localized an insect-specific loop in the middle of the α-helical domain, which might provide flexibility as a hinge for lipid-interaction in insects (Fig. 5). The otherwise general sequence conservation at this domain could point to essential functions in extracellular membrane interactions and/or sub-cellular membrane binding. The latter option is supported by human LLTP data; the α-helical domain is included in the minimal part of apoB needed for its direct endoplasmic membrane interaction during the assembly of apoB-lipid complex (LDL) (34,35,46). Our model of the honey bee α-helical domain shows many solvent-facing positively charged amino acid residues, which can initiate electrostatic protein interaction with negatively charged membranes (47). We believe these positively charged residues are involved in the binding to negatively charged membranes, and possibly also in charge-matching with phosphate groups found on all cellular membranes. These propositions are in accordance with our findings of notable yet lower affinity of the α-helical domain, full-length Vg and its 150 kDa fragment to neutral bilayers.

After documenting Vg-membrane binding, we investigated whether this could be a disruptive antibacterial mechanism, inspired by the antibacterial effects of LLTPs in fish (24,27,48) and human (22,23). But neither membrane disruptive- nor bactericidal effects were observed for honey bee Vg. A recent fish study assigns the bactericidal effect of the protein to a vertebrate Vg domain called phosvitin (26), and it is therefore possible this property is a gain-of-function restricted to vertebrates. On the other hand, since honey bee Vg was found to interact with *E. coli*, we cannot exclude that this protein participates innate immunity without direct bactericidal actions by stimulating the phagocytic cells, as described in fish (28), or has effects on other bacteria than *E. coli*. These possibilities may be worthwhile to address in future experiments.

Without support for an antibacterial role, can the Vg binding we document here at membranes of worker bees still reflect an adaptive function? We believe that a candidate mechanism has been indicated before: Vg might shield membranes against oxidative damage from free radicals that can cause cellular and organismal senescence (49). We base this hypothesis on the work of Seehuus and colleagues (13), who showed that honey bee Vg reduces cellular and organismal damage induced by a free radical producing

agent, and who confirmed that Vg is a preferential target of oxidative carbonylation — a protein modification caused by free radicals. Alternatively or additionally, Vg might be attracted by modified, for instance oxidized, lipid surfaces and act as an anti-inflammatory quencher. Human apoB is already such a proposed immuno-suppressing hyppo-quencher (22).

Overall, our discovery of Vg-membrane binding emphasizes the versatility of this protein that has significant effects on reproduction, social behavior, immunity and survival in honey bees, and perhaps also more widely in *Apis*. Vg is expressed copiously in honey bee workers, which are globally abundant and commercially available model insects. Honey bee Vg is, accordingly, easy to obtain in reasonable quantities and can offer an attractive future model system for LLTPs.

Material and methods

Material sampling and protein purification

Handling of colony sources, honey bee tissue dissection, Vg extraction and Vg purification was performed as described elsewhere (38). Single ion-exchange purification was used for hlVg purification. FbVg was purified from fatbody protein extract (38) by GenScript (NY, USA) with size-exclusion (SEC) and ion-exchange steps like described before (38), and the purification was completed with a Superdex 75 column (Sigma-Aldrich, MO, USA). The Vg fragments of the α-helical region (416-778) and vWFD (1444-1615) were produced by GenScript: The vWFD DNA was subcloned into pUC57 vector, and α-helical region was subcloned into pCold trigger factor (TF) vector. Both fragments contain an N-terminal hexa histidine-tag, which was used for one-step affinity purification. The α-helical fragment is a fusion construct with TF that is a highly soluble protein with chaperon activity (50), and separation of the α-helical fragment from its fusion partner by HRV 3C cleavage was not feasible due to fragmentation of the α-helical domain. The quality of the domains was tested using dynamic light scattering (DLS) and SEC (not shown).

Membrane protein immunoblot

Membrane protein extraction protocol was modified from (51) and (52). The sample was kept on ice or at 4°C during all steps, and all buffers contained a protease inhibitor cocktail (Roche, Mannheim, Germany). Five frozen *diutinus* bee abdomens, emptied of gut and ovary, were homogenized (like (38)) and centrifuged at 800 g for 10 min in order to remove exoskeleton and nuclei. The supernatant was centrifuged at 30,000 g for 20 min in order to collect the membranes. The resulting supernatant was filtrated through a syringe filter of 0.2 µm pore size (Pall, Washington, USA) and stored as the cytosolic fraction. The membrane-containing pellet was washed three times by suspending it in 1 ml HBS buffer, in order to remove soluble material, and finally centrifuged at 30,000 g for 20 min and suspended in HBS containing 0.1% Triton X-100. 24 µg of the cytosolic and the membrane proteins (measured with Nanodrop, Thermo Fisher Scientific Inc, MA, USA) were applied on a 7.5% SDS-PAGE gel and blotted on a nitrocellulose membrane that was incubated with 0.5% Tween containing TBS with 2.5% fat-free milk powder overnight. The membrane was then incubated for 1 h with Vg antibody tested before (38,53) (1:25,000; Pacific Immunology, Ramona, CA, USA) followed by incubation for 1 h with a horse-radish peroxidase conjugated secondary antibody like before (38,54). The bands were visualized using Immun-Star kit and Bio-Rad (CA, USA) ChemiDoc XRS+ imager. All blotting reagents were purchased from Bio-Rad.

Cell binding assay

The adherence of Vg toward cells was tested using Sf9 insect cells (Invitrogen, Paisley, UK). The assay is modified from fish Vg bacterial binding assay (24). Centrifugations were 5 min 2,000 g in room temperature and the wash volume was 0.5 ml of PBS (pH 7.4), if not specified otherwise. The cells were thawed at a 37°C water bath, washed and centrifuged. The pellet was resuspended in PBS for final concentration of 1.5 x 10^7 cells/ml at 28°C. 25 µl of cells were mixed with 100 µl of hemolymph diluted 1:10 in PBS, total protein concentration of 1.5 mg/ml, or with 100 µl of fatbody protein extract

(38), concentration 5.7 mg/ml. Prior to the experiment, both hemolymph and fatbody extract were filtrated (pore size 0.2 µm) and centrifuged for 20 min, 20,800 g, at 4°C. Two negative controls were used: 25 µl cells with 100 µl PBS but no protein, and 100 µl hemolymph/fatbody protein extract with 25 µl PBS but no cells. The latter secures that the experiment does not measure general aggregation of Vg. 0.1 µl hemolymph and 0.5 µl fatbody extract were kept as untreated controls. The samples were incubated at 28°C for 50 min with gentle agitation and then washed a total of six times to remove unbound protein. The final pellet was resuspended in 20 µl of 4 M urea in 10 mM PBS (pH 8), agitated for 15 min and centrifuged (20 min; 20,800 g). The supernatants were run on a gel and blotted like above (see Membrane protein immunoblot).

Histology

Dissection, tissue preparation and immunohistochemistry were largely performed as described previously (15). In brief, tissue samples were fixed in paraformaldehyde (4% in PBS), embedded in London Resin White (Electron Microscopy Science, Hatfield, USA), and cut with a Reichert Jung ultra-microtome (Leica, Wetzlar, Germany; section thickness: 1-2 µm). Mounted sections were rinsed with PBS-NTx (0.25% Triton X-100 in PBS), pre-incubated with 2% BSA (Sigma-Aldrich) in PBS-NTx for 60 min, and then incubated overnight with a polyclonal anti-Vg antibody at 1:100 (see section on Membrane protein immunoblot). After washing in PBS-NTx, the Cy5 conjugated secondary antibody (Jackson ImmunoResearch, West Grove, PA, USA, 1:200) and the nuclear stain 4′,6-diamidino-2-phenylindole (DAPI; 1:1,000 from 0.5 mg/ml stock, Sigma-Aldrich) were co-applied. Finally, sections were rinsed again in PBS-NTx and cleared in glycerol (30% in PBS). Separation of immunopositive and false positive signals is complicated by the fact that production sites release Vg into the hemolymph. This can entail a widespread dispersal of Vg and highly variable immunosignal intensities. Therefore, to further rule out false positives due to unspecific secondary antibody binding, autofluorescence as well as detector noise, we included controls that were not incubated with the primary antibody. In order to directly compare signal intensities in these controls and regularly treated test samples, controls and tests were

scanned in an alternating fashion with all scanning settings kept constant. Two controls and two test samples were prepared for each of 5 individuals. Confocal images were acquired on a Leica TCS SP5 Laser Scanning Confocal Microscope (LSCM; Leica, Wetzlar, Germany) using a 63x oil immersion objective (numerical aperture = 1.4). Image stacks (z = 2 µm, Δz = 0.5 µm) or single optical sections (control vs. test comparisons) were viewed and processed in ImageJA v. 1.44b (public domain, http://pacific.mpi-cbg.de/wiki/index.php/ImageJA).

Liposome preparation

The phospholipids were purchased from Avanti Polar Lipids (Alabaster, USA) and other reagents from Sigma-Aldrich. Egg yolk phosphatidylcholine (E) and porcine brain phosphatidylserine (P) mixed 1:1 with E were dissolved in chloroform, which was evaporated resulting in formation of a lipid film. For lyophilisation, samples were left on Christ Alpha 1-4 lyophilisation system (Martin Christ Gefriertrocknungsanlagen GmbH, Osterode am Harz, Germany) for two hours to remove traces of chloroform. The lipids were resuspended in HBS (for SPR) or in 10 mM Hepes, 100 mM NaCl including 12.5 mM 8-aminonaphtalene-1,3,6-trisulphonic acid (ANTS) and 45 mM p-xyelene-bis-N-pyrimidium bromide (DPX), pH 7.0 (for leakage) and left overnight in the dark at room temperature. The hydrated phospholipid solutions were then frozen and thawed seven times, followed by extrusion (11 times) with the Avanti Mini-Extruder using a membrane of 0.1 µm pore size. The leakage-intended liposomes were cleared of unencapsulated fluorochrome (ANTS) and quencher (DPX) with a Sephadex G-75 column. Determination of phospholipid concentration was based on the determination of total phosphorus using a protocol from Avanti Polar Lipids (55). The liposome size was determined by DLS using a Zetasizer Nano-Instrument resulting in: E 102 nm and P:E 108 nm for SPR, and E 122 nm and P:E 191 nm for leakage.

Surface plasmon resonance

Biacore T200 (GE Healthcare, Waukesha, USA) with an L1 sensor chip designed for liposome capture was used in the SPR experiments (56). The liposome capturing wizard method of the Biacore control software was used as a method template. The L1 chip was conditioned with 40 mM octyl glucoside (30 s, flow rate 30 µl/min). Concentrations of 0.5 mM E and 1.2 mM P:E liposomes were captured on separate channels of the chip (30 s, 10 µl/min), and unbound structures were washed with 10 mM NaOH, followed by a stabilization time of 300 s. The binding temperature was 25°C, while the sample compartment was kept at 4°C. The liposome surface coverage was verified according to Biacore user manual's instructions using BSA that binds on the naked chip, but has low affinity toward a liposome-covered chip. After optimization, following protein concentrations of hlVg, fbVg, α-helical fragment and vWFD were used: 0, 100, 500, 1,500 and 2,500 nM. For fbVg, the molecular weight of 150 kDa fragment was used for calculations, since this fragment is dominating the sample. After each sample injection (60 s, flow rate 30 µl/min, dissociation time 600 s) the chip was regenerated by injecting 2:3 isopropanol:NaOH (30 s, 30 µl/min), since the protein did not dissociate completely from the lipid-covered chip. Regeneration removed all protein and lipids from the chip, recovering the RU-level baseline. The final protein-membrane binding experiments were performed in row using a fresh chip and the same batch of liposomes.

Liposome-combined proteolysis and LC-MSMS

Neutral (size 120 nm) and negative (99 nm) liposomes were incubated with fbVg (see protein purification) for 1 h at room temperature in the dark and digested with 25 µg chymotrypsin for 30 min on ice. The volumes and concentrations in the reaction were: 1 µl of 0.1 µM liposomes and 8 µl of 0.7 mg/ml Vg – the total volume was adjusted to 12 µl with HBS. As controls, we had Vg without liposomes or protease, Vg with P:E or E liposomes, and Vg with chymotrypsin. Also, we tested digestion with a smaller amount (0.25 µg) of chymotrypsin in the presence and absence of both liposomes to ensure that liposomes as such do not affect the digestion rate. The samples were run on 4-20% SDS-PAGE (Bio-Rad). Band intensities were measured with the Bio-Rad ChemiDoc XRS system. The protease resistant band in the liposome samples was excised and analyzed

using LC-MSMS. The gel pieces were washed, treated with 10 mM DTT (Amersham Biosciences, Piscataway, USA) for cysteine reduction and alkylated using iodoacetamide (Sigma-Aldrich), followed by trypsin digestion. After digestion, the dried samples were dissolved in 11 µl 0.1% formic acid and 6 µl was used for injection. The samples were analyzed on 4000 QTrap (Applied Biosystems /MDS SCIEX, Concord, ON, Canada).

Modeling

A homology model based on lamprey lipovitellin X-ray structure (57,58) was built using software Bodil (59). The following Vgs were used for the alignment: *Bombus ignitus* (Swiss-prot accession number B9VUV6), *Pimpla nipponica* (O17428), *Pteromalus puparum* (B2BD67), *Leucophaea maderae* (Q5TLA5), *Acipenser transmontanus* (Q90243) and *Xenopus laevis* (P18709). Multiple LLTP alignment (by (60)) was used as a guide. The model was energetically minimized with program Discovery Studio (Accelrys, San Diego, CA, USA). A segment absent in lamprey was verified insect-specific with a multiple alignment against the honey bee amino acid residues L564-L604. In addition to the species above, this alignment included Vgs of *Bombus hypocrite* (C7F9J8), *Solenopsis invicta* Vg-2 (Q2VQM6), *Aedes aegypti* (Q16927), *Riptortus clavatus* (O02024), *Anthonomus grandis* (Q05808), *Graptopsaltria nigrofuscata* (Q9U5F1), *Antheraea pernyi* (Q9GUX5), *Periplaneta americana* (Q9U8M0), *Rhyparobia maderae* (Q9GR96), *Blattella germanica* (O76823), *Athalia rosae* (BAA22791), *Encarsia formosa* (AAT48601), *Oreochromis aureus* (Q9YGK0), *Oncorhynchys mykiss* (Q92093) and *Fundulus heteroclitus* (Q90508) and human apoB (P04114) and human MTP (P55157).

Bacterial binding assay

Hemolymph and fatbody protein extract were used for testing Vg-binding on living bacteria, adapted from the fish Vg bacteria binding experiment by Tong *et al*. (2009). The experiment was performed at room temperature, centrifugation steps were 3,000 g for 5 min, and wash volume was 0.5 ml of 0.9% NaCl, if not mentioned otherwise. Shortly, *E.*

coli were washed and resuspended in 100 µl of 20 mM Tris-HCl containing 150 mM NaCl (pH 7.5). The *E. coli* suspension (2 x 10^9 cells/ml) was mixed with equal volume of hemolymph dilution or with fatbody protein extract. Two negative controls were used: 100 µl bacteria with equal volume of saline but no protein, and 100 µl fatbody protein extract with equal volume of saline but no bacteria. 0.1 µl hemolymph and 0.5 µl fatbody extract were kept as untreated controls. The samples were incubated at 26°C for 50 min with agitation. Subsequently, the bacteria were washed with saline a total of six times. The final pellet was resuspended in 40 µl of 4 M urea in 10 mM Tris-HCl (pH 8), agitated for 15 min and centrifuged. The supernatants were blotted like above.

Extrinsic fluorescence measurement

The fluorescence spectra of the leakage assay using the hlVg and fbVg samples were recorded at 25°C on a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies). Excitation of ANTS was performed at 355 nm and its subsequent emission measured at 510 nm, as this is close to the λ_{max} of ANTS, at a scan rate of 600 nm/min. 10- and 20-nm slit widths in the excitation and emission pathways, respectively, were used. A 10 mm quartz cuvette was filled with 1 mM P:E (see Liposome preparation) and 2.5 µM sample was added. The fluorescence was measured every 0.25 min over a time range of 20 min. To estimate the maximum attainable fluorescence, Triton X-100 was added (a final concentration of 2 mM) for the disruption of the liposome bilayer and for the release of the encapsulated contents at the end of the time scan. The leakage was evaluated as a percentage of complete release with Triton X-100. The initial intensity (at 0 min) of the liposome and sample solution was set to 0% leakage and the final intensity after adding Triton X-100 corresponds to 100% leakage.

Antibacterial activity assay

The antibacterial activity assay was modified from Tong *et al*. (2009), where the zebra fish (*Danio rerio*) Vg was found antibacterial. In our test, purified Vg from bee fatbody was used (see Material sampling and protein purification). 5 µl of 0.9% saline was added

to the control (20 µl of 50 mM Tris, pH 8) or to 20 µl 0.8 mg/ml Vg in the Tris buffer. The control and the Vg sample were incubated with 1 µl of ampicillin resistant Epicurian Gold *E. coli* (10^6 cells/ml) at 26°C for 60 min, diluted 1,000-fold with 0.9% saline and plated on three LB-ampicillin agar plates, 20 µl per plate. After incubation at 37°C for 16 h, the bacterial colonies were calculated. The experiment was repeated six times: four experiments with N = 3 and twice with N = 5.

Statistics

The lipid bilayer binding responses measured by SPR were plotted against the sample concentration using SigmaPlot (Systat Software, Washington, USA). The saturation half value, i.e. the equilibrium dissociation constant ($S_{0.5}$) and its standard error were calculated by fitting a curve to the values using equation $y = B_{max} X / (S_{0.5} + X)$ with SigmaPlot (Systat Software, Washington, USA). In the liposome-combined proteolysis test, the significance of the band intensities versus control were determined with Kruskal-Wallis test, and Mann Whitey U test was used on the two liposome samples. The bacterial colony counts were analyzed by the t-test.

Acknowledgements

The work was mostly performed in the laboratory of Biorecognition at University of Bergen, headed by Prof. Aurora Martinez, whom we thank for extensive support. H.H., Ø.H. and G.V.A. conceived and designed the experiments. H.H., Ø.H. and D.M. collected the samples. D.M. performed the immunohistochemistry section. A.B. prepared the liposomes and performed the leakage experiment, and assisted with SPR. The mass-spectrometric analysis was performed by PROBE proteomic unit of University of Bergen. Other experiments were performed by H.H. All authors participated in the writing. We thank Ole Horvli for assistance with the SPR instrument, Marte Flydal for assistance with the bacteria assay, Helene Bustad Johannessen and Åge Skjevik for help with the statistics and Assoc. Prof. Knut Teigen for the model minimization. We thank GenScript for satisfying collaboration. The PROBE work was partly supported by the National Program for Research in Functional Genomics (FUGE) funded by the Norwegian Research Council. Ø.H. was supported by the Research council (#185306) and Norwegian Cancer Society (#58240001). G.V.A. was supported by the Research Council of Norway (#180504, 185306 and 191699), the National Institute on Aging (NIA P01 AG22500) and the PEW Charitable Trust.

References

1. Tufail, M., and Takeda, M. (2008) *J Insect Physiol* **54**, 1447-1458
2. Finn, R. N. (2007) *Biol Reprod* **76**, 926-935
3. Sappington, T. W., and Raikhel, A. S. (1998) *Insect Biochem Mol Biol* **28**, 277-300
4. Roma, G. C., Bueno, O. C., and Camargo-Mathias, M. I. (2010) *Micron* **41**, 395-401
5. Yin, L., Nordin, J. H., Lucches, P., and Giorgi, F. (2001) *Cell Tissue Res* **304**, 391-399
6. Tufail, M., and Takeda, M. (2009) *J Insect Physiol* **55**, 87-103
7. Guidugli-Lazzarini, K. R., do Nascimento, A. M., Tanaka, E. D., Piulachs, M. D., Hartfelder, K., Bitondi, M. G., and Simoes, Z. L. (2008) *J Insect Physiol* **54**, 1138-1147
8. Trenczek, T., Zillikens, A., and Engels, W. (1989) *Journal of Insect Physiology* **35**, 475-481
9. Guidugli, K. R., Nascimento, A. M., Amdam, G. V., Barchuk, A. R., Omholt, S., Simoes, Z. L., and Hartfelder, K. (2005) *FEBS Lett* **579**, 4961-4965
10. Nelson, C. M., Ihle, K. E., Fondrk, M. K., Page, R. E., and Amdam, G. V. (2007) *PLoS Biol* **5**, e62
11. Amdam, G. V., Simoes, Z. L., Hagen, A., Norberg, K., Schroder, K., Mikkelsen, O., Kirkwood, T. B., and Omholt, S. W. (2004) *Exp Gerontol* **39**, 767-773
12. Amdam, G. V., Aase, A. L., Seehuus, S. C., Kim Fondrk, M., Norberg, K., and Hartfelder, K. (2005) *Exp Gerontol* **40**, 939-947
13. Seehuus, S. C., Norberg, K., Gimsa, U., Krekling, T., and Amdam, G. V. (2006) *Proc Natl Acad Sci U S A* **103**, 962-967
14. Amdam, G. V., Hartfelder, K., Norberg, K., Hagen, A., and Omholt, S. W. (2004) *J Econ Entomol* **97**, 741-747
15. Smedal, B., Brynem, M., Kreibich, C. D., and Amdam, G. V. (2009) *J Exp Biol* **212**, 3795-3801
16. Denison, R., and Raymond-Delpech, V. (2008) *Invert Neurosci* **8**, 1-9
17. Omholt, S. W., and Amdam, G. V. (2004) *Sci Aging Knowledge Environ* **2004**, pe28
18. Babin, P. J., Bogerd, J., Kooiman, F. P., Van Marrewijk, W. J., and Van der Horst, D. J. (1999) *J Mol Evol* **49**, 150-160
19. Baker, M. E. (1988) *Biochem J* **255**, 1057-1060
20. Zeissig, S., Dougan, S. K., Barral, D. C., Junker, Y., Chen, Z., Kaser, A., Ho, M., Mandel, H., McIntyre, A., Kennedy, S. M., Painter, G. F., Veerapen, N., Besra, G. S., Cerundolo, V., Yue, S., Beladi, S., Behar, S. M., Chen, X., Gumperz, J. E., Breckpot, K., Raper, A., Baer, A., Exley, M. A., Hegele, R. A., Cuchel, M., Rader, D. J., Davidson, N. O., and Blumberg, R. S. (2010) *J Clin Invest* **120**, 2889-2899
21. Sagiv, Y., Bai, L., Wei, D. G., Agami, R., Savage, P. B., Teyton, L., and Bendelac, A. (2007) *J Exp Med* **204**, 921-928
22. Cho, N. H., and Seong, S. Y. (2009) *Immunology* **128**, e479-486

23. Bartolome, N., Aspichueta, P., Martinez, M. J., Vazquez-Chantada, M., Martinez-Chantar, M. L., Ochoa, B., and Chico, Y. (2010) *Innate Immun*
24. Tong, Z., Li, L., Pawar, R., and Zhang, S. (2009) *Immunobiology* **215**, 898-902
25. Garcia, J., Munro, E. S., Monte, M. M., Fourrier, M. C., Whitelaw, J., Smail, D. A., and Ellis, A. E. (2010) *Fish Shellfish Immunol*, 293-297
26. Wang, S., Wang, Y., Ma, J., Ding, Y., and Zhang, S. (2011) *J Biol Chem* **286**, 22653-22664
27. Zhang, S., Wang, S., Li, H., and Li, L. (2011) *Int J Biochem Cell Biol* **43**, 303-305
28. Li, Z., Zhang, S., and Liu, Q. (2008) *PLoS One* **3**, e1940
29. Epand, R. F., Mor, A., and Epand, R. M. *Cell Mol Life Sci* **68**, 2177-2188
30. Kelly, B. A., Harrison, I., McKnight, A., and Dobson, C. B. *BMC Immunol* **11**, 13
31. Halskau, O., Muga, A., and Martinez, A. (2009) *Current Protein & Peptide Science* **10**, 339-359
32. Sivaram, P., Choi, S. Y., Curtiss, L. K., and Goldberg, I. J. (1994) *J Biol Chem* **269**, 9409-9412
33. Hussain, M. M., Rava, P., Pan, X., Dai, K., Dougan, S. K., Iqbal, J., Lazare, F., and Khatun, I. (2008) *Curr Opin Lipidol* **19**, 277-284
34. Ledford, A. S., Weinberg, R. B., Cook, V. R., Hantgan, R. R., and Shelness, G. S. (2006) *J Biol Chem* **281**, 8871-8876
35. Mann, C. J., Anderson, T. A., Read, J., Chester, S. A., Harrison, G. B., Kochl, S., Ritchie, P. J., Bradbury, P., Hussain, F. S., Amey, J., Vanloo, B., Rosseneu, M., Infante, R., Hancock, J. M., Levitt, D. G., Banaszak, L. J., Scott, J., and Shoulders, C. C. (1999) *J Mol Biol* **285**, 391-408
36. Seong, S. Y., and Matzinger, P. (2004) *Nat Rev Immunol* **4**, 469-478
37. Kent, C., Issa, A., Bunting, A., and Zayed, A. (2011) *Molecular Ecology* **In press.**
38. Havukainen, H., Halskau, O., Skjaerven, L., Smedal, B., and Amdam, G. V. (2011) *J Exp Biol* **214**, 582-592
39. Bigay, J., Casella, J. F., Drin, G., Mesmin, B., and Antonny, B. (2005) *EMBO J* **24**, 2244-2253
40. Jones, D. T. (1999) *J Mol Biol* **292**, 195-202
41. Yu, L. L., Guo, L., Ding, J. L., Ho, B., Feng, S. S., Popplewell, J., Swann, M., and Wohland, T. (2009) *Biochimica Et Biophysica Acta-Biomembranes* **1788**, 333-344
42. Abraham, T., Marwaha, S., Kobewka, D. M., Lewis, R. N. A. H., Prenner, E. J., Hodges, R. S., and McElhaney, R. N. (2007) *Biochimica Et Biophysica Acta-Biomembranes* **1768**, 2089-2098
43. Campanella, C., Caputo, M., Vaccaro, M. C., De Marco, N., Tretola, L., Romano, M., Prisco, M., Camardella, L., Flagiello, A., Carotenuto, R., Limatola, E., Polzonetti-Magni, A., and Infante, V. (2011) *Mol Reprod Dev* **78**, 161-171
44. Li, A., Sadasivam, M., and Ding, J. L. (2003) *J Biol Chem* **278**, 2799-2806
45. Halskau, O., Ying, M., Baumann, A., Kleppe, R., Rodriguez-Larrea, D., Almas, B., Haavik, J., and Martinez, A. (2009) *Journal of Biological Chemistry* **284**, 32758-32769
46. Richardson, P. E., Manchekar, M., Dashti, N., Jones, M. K., Beigneux, A., Young, S. G., Harvey, S. C., and Segrest, J. P. (2005) *Biophys J* **88**, 2789-2800

47. Takahashi, D., Shukla, S. K., Prakash, O., and Zhang, G. (2010) *Biochimie* **92**, 1236-1241
48. Shi, X., Zhang, S., and Pang, Q. (2006) *Fish Shellfish Immunol* **20**, 769-772
49. Finkel, T., and Holbrook, N. J. (2000) *Nature* **408**, 239-247
50. Thapa, A., Shahnawaz, M., Karki, P., Raj Dahal, G., Golam Sharoar, M., Yub Shin, S., Sup Lee, J., Cho, B., and Park, I. S. (2008) *Biotechniques* **44**, 787-796
51. Fischer, H. D., Gonzalez-Noriega, A., and Sly, W. S. (1980) *J Biol Chem* **255**, 5069-5074
52. Kirankumar, N., Ismail, S. M., and DuttaGupta, A. (1997) *Insect Biochemistry and Molecular Biology* **27**, 671-679
53. Seehuus, S. C., Norberg, K., Krekling, T., Fondrk, K., and Amdam, G. V. (2007) *J Insect Sci* **7**, 1-14
54. Amdam, G. V., Norberg, K., Hagen, A., and Omholt, S. W. (2003) *Proc Natl Acad Sci U S A* **100**, 1799-1802
55. Fiske, C. H., and Subbarow, Y. (1925) *J Biol Chem* **66**, 374-389
56. Besenicar, M., Macek, P., Lakey, J. H., and Anderluh, G. (2006) *Chem Phys Lipids* **141**, 169-178
57. Anderson, T. A., Levitt, D. G., and Banaszak, L. J. (1998) *Structure with Folding & Design* **6**, 895-909
58. Raag, R., Appelt, K., Xuong, N. H., and Banaszak, L. (1988) *J Mol Biol* **200**, 553-569
59. Lehtonen, J. V., Still, D. J., Rantanen, V. V., Ekholm, J., Bjorklund, D., Iftikhar, Z., Huhtala, M., Repo, S., Jussila, A., Jaakkola, J., Pentikainen, O., Nyronen, T., Salminen, T., Gyllenberg, M., and Johnson, M. S. (2004) *J Comput Aided Mol Des* **18**, 401-419
60. Avarre, J. C., Lubzens, E., and Babin, P. J. (2007) *BMC Evol Biol* **7**, 3

Figure legends

Figure 1. Vg is found in the membrane fraction of honey bee abdominal tissue, and the protein binds to insect cells *in vitro*. *A*, An anti-Vg-immunoblot shows Vg in the cytosol and in the membrane fraction of abdominal protein extract. An amount of 24 μg total protein was loaded on each lane. *B-C*, Cell binding assay using hemolymph (*B*) or fatbody protein extract (*C*) as sample. The last lane shows the Vg associated to Sf9 cells after incubation with hemolymph/fatbody extract and washes. The negative controls show hemolymph/fatbody extract incubated and washed without cells and cells incubated without these samples. S = molecular weight standard.

Figure 2. Immunolabeling indicates the presence of Vg at vesicular (asterisks) and cellular membranes (arrowheads). Representative optical sections of abdominal (*A-C*) and head tissue (*D-F*) are labeled with anti-Vg (green) and DAPI nuclear stain (magenta). *A-C,* Trophocytes are identified by irregularly shaped nuclei (TcN), and contain a large number of Vg rich vesicles. Intense VgIR also was observed in other abdominal cell types, for example in muscle tissue (Mu). *D,* Hypopharyngeal glands with intense VgIR in Vg rich granules, and less intensely labeled cellular membranes (arrowheads). *E-F,* Syncitial cells of muscle fibers with marked VgIR at their outer membranes (arrowheads). *G-L,* Test sections treated with anti-Vg (*G, I, K*) as compared to alternate control sections treated only with the secondary antibody (*H, J, L*). The nuclear stain (DAPI) confirms that control and test images were taken from similar sites and cell types (fatbody trophocytes). Scale bars: A = 50 μm, B-F = 20 μm, G-L = 20 μm.

Figure 3. SPR sensograms for the 2,500 nM samples (*A*) and binding curves based on the maximum responses of the sensograms of all concentrations (*B*). Red color is used for negatively charged bilayer (P:E) and black for neutral bilayer (E). Negative charge enhances Vg-membrane association in all cases of samples that bind (red vs. black curves). Vg purified from hemolymph is indicated as thick lines, Vg purified from fatbody as dashed lines and α-helical fusion domain Vg is indicated as dot lines. The vWFD domain is illustrated in grey. *A*, All sensograms of the samples except from the vWFD domain show irreversible binding toward the membranes: none of the samples that bound returned back to the baseline during the dissociation time (600 s). *B*, Response Units as a function of concentration. These curves were fitted based on R_{max} values of the measurements done at sample concentrations of 0, 100, 500, 1,500 and 2,500 nM, marked with dots, triangles or crosses. The fastest association is seen with full-length Vg on P:E (the steepest red curve). The α-helical recombinant domain binding pattern on P:E is nearly identical to the binding pattern of fatbody purified Vg (the red curves in dotted and dashed lines, respectively). hlVg = Vg purified from hemolymph. fbVg = Vg purified from fatbody. α-helical and vWFD = domains of Vg.

Figure 4. A Vg segment is resistant toward proteolysis by binding to liposomes. *A,* Two gels show how a part of Vg (the band marked with black arrows) is not digested by chymotrypsin (ct) in the presence of neutral (E) or negative (P:E) liposomes. The grey arrows show two bands, whose intensity on all lanes was used as a reference. *B*, Titration digestion using smaller chymotrypsin amount showed that the proteolysis rate was not, in general, affected by the liposomes since the controls and the liposome-incubated samples (E, P:E) share the same digestion products (faint bands under the 150 kDa and 40 kDa Vg fragments). *C*, The intensity of the bands marked with black arrows in 4*A* is significantly higher than the intensity of the corresponding gel region in the controls. The reference bands (grey columns; grey arrows in 4*A*) showed no significant difference. The bars are the standard deviations of the six band intensities. S = molecular weight standard. Vg = untreated, purified Vg. P:E = Vg with P:E liposomes. E = Vg with E liposomes.

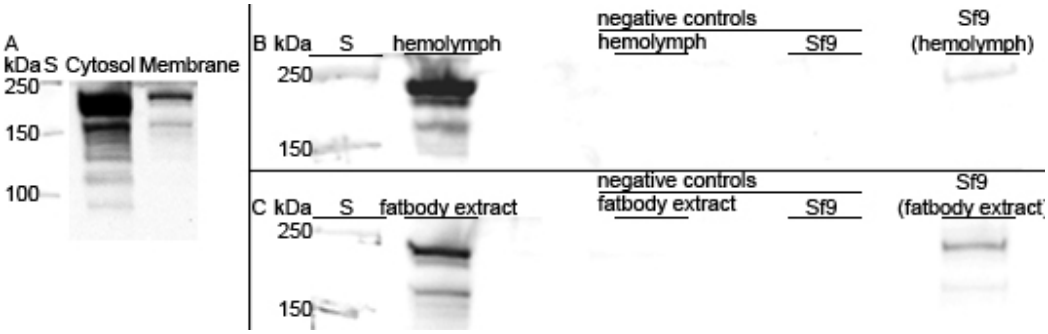
Figure 5. A homology model of the 18-helix Vg domain. The positively charged residues are shown as blue sticks. The 25 residue region that could not be modeled is marked with arrows. The model begins at serine 416 and ends at leucine 765 of honey bee Vg. *A,* The helices facing the solvent (green). *B,* The helices facing the lipid-binding cavity (grey). *C*, A multiple alignment of the segment that is absent in lamprey (arrows in *A* and *B*). This predicted loop is conserved in the 16 insect

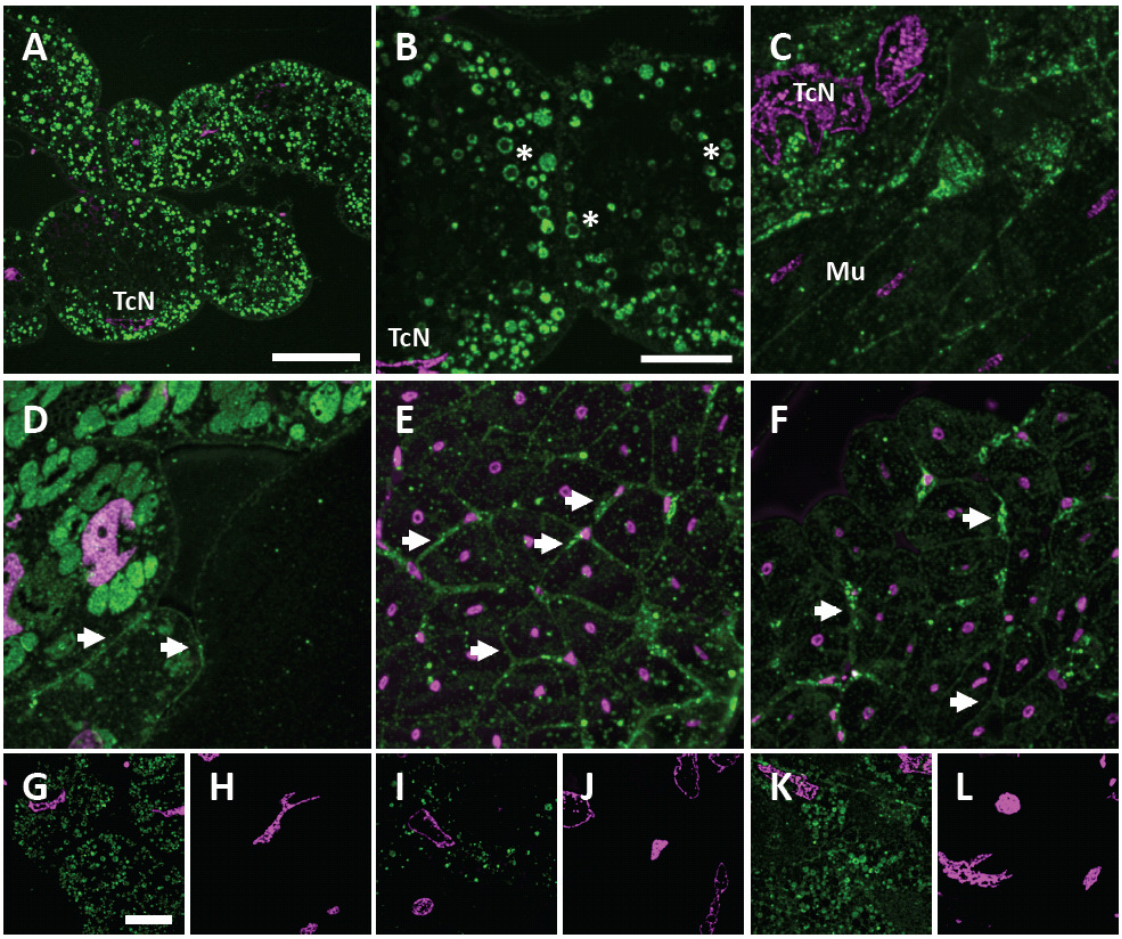
species included in the alignment, but the five fish, an amphibian (*Xenopus laevis*) and the human homologues lack this prolonged loop.

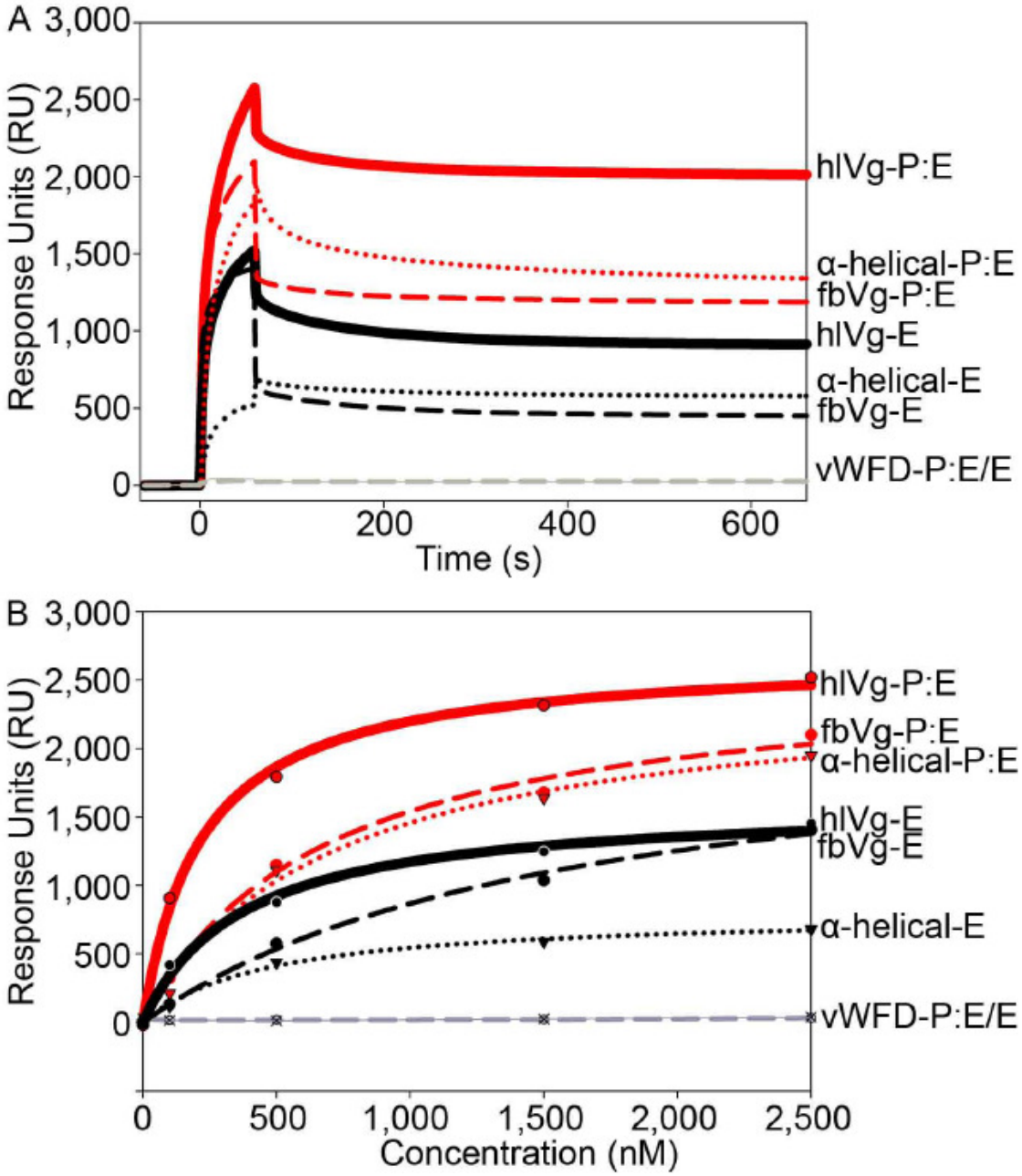
Figure 6. Vg associates with *E. coli*, but does not disrupt negative liposome membranes or kill *E. coli*. *A*, *E. coli* were incubated with hemolymph, fatbody protein extract or saline (treatment in parenthesis). After thorough washing of the bacteria, Vg was found associated with *E. coli* by immunoblotting. Hemolymph and fatbody extract were used as positive controls. Negative controls were fatbody extract incubated without *E. coli* and the saline treated bacteria. S = molecular weight standard. *A* is a combination of two blots. *B*, Content release from P:E liposomes in response to Vg purified from hemolymph (black) or from fatbody (grey) at the concentration of 2.5 μM. The fluorescence signal at 0 min was set to 0% and 100% leakage corresponds to the signal of liposomes fully disrupted by Triton X-100. Neither of the Vg samples causes notable leakage, which means that the integrity of the negatively charged liposomes is not effected by Vg. *C*, The bacteria colony count graph is a combination of six distinct antibacterial assays (total N = 22) with control *E. coli* and *E. coli* incubated with Vg prior to plating. The control and the Vg incubated sample do not significantly differ. The bar shows the standard deviation.

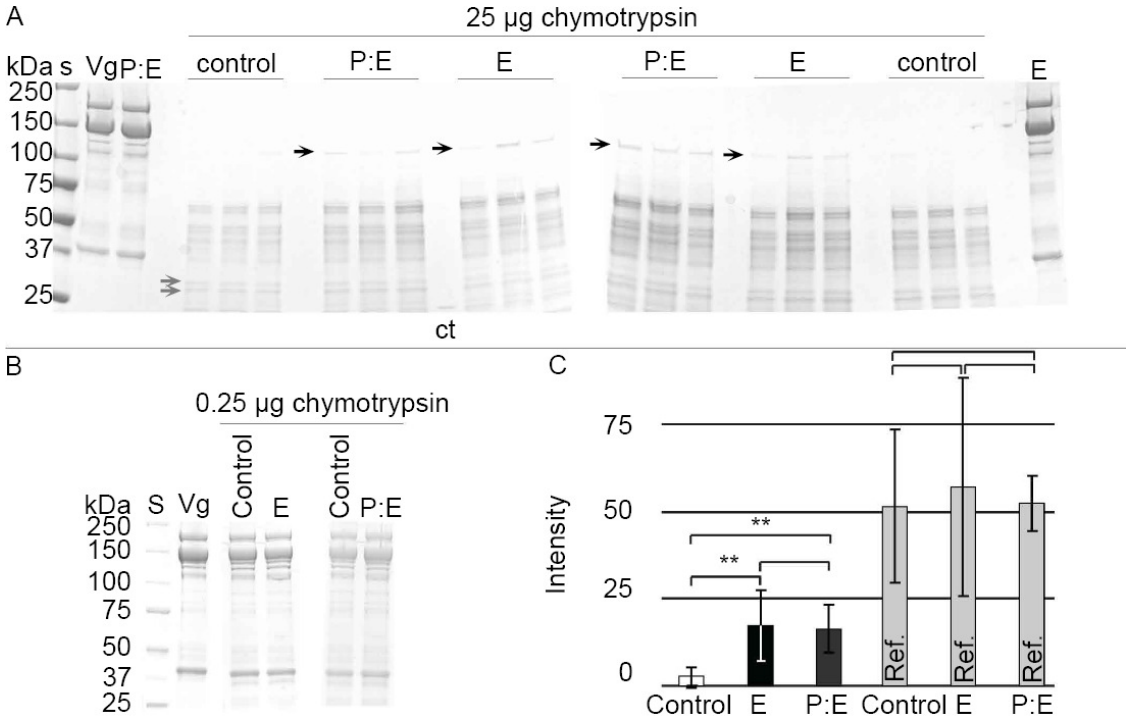
Table 1. Numerical presentation of Vg membrane affinity. $S_{0.5}$ is the protein concentration that yields half of the maximal response. $S_{0.5}$ was calculated based on the affinity curve of the five measurements (Fig. 3B). The given standard error is based on the curve fitting.

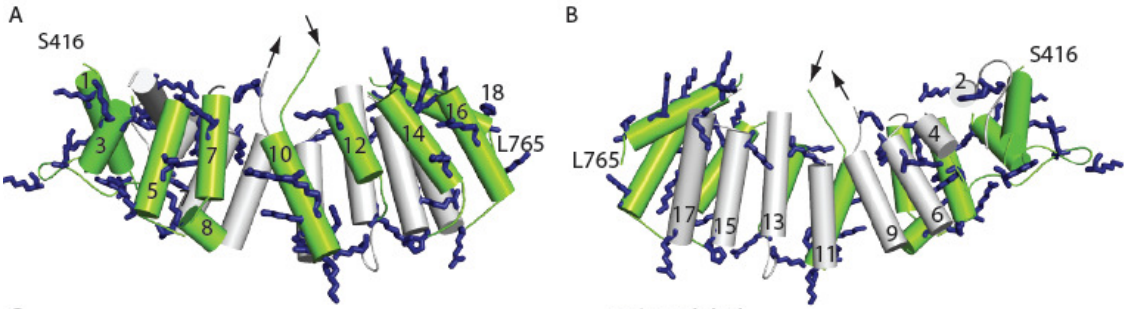
Figures


A
kDa S Cytosol Membrane
250
150
100
B kDa S hemolymph
250
150
negative controls
hemolymph
Sf9
Sf9
(hemolymph)
C kDa S fatbody extract
250
150
negative controls
fatbody extract
Sf9
Sf9
(fatbody extract)


A
TcN
B
TcN
C
TcN
Mu
D
E
F
G
H
I
J
K
L


A
Response Units (RU)
3,000
2,500
2,000
1,500
1,000
500
0
0
200
400
600
Time (s)
hlVg-P:E
α-helical-P:E
fbVg-P:E
hlVg-E
α-helical-E
fbVg-E
vWFD-P:E/E
B
Response Units (RU)
3,000
2,500
2,000
1,500
1,000
500
0
0
500
1,000
1,500
2,000
2,500
Concentration (nM)
hlVg-P:E
fbVg-P:E
α-helical-P:E
hlVg-E
fbVg-E
α-helical-E
vWFD-P:E/E


A
25 µg chymotrypsin
kDa s Vg P:E
control
P:E
E
P:E
E
control
E
250
150
100
75
50
37
25
ct
B
0.25 µg chymotrypsin
kDa S Vg Control E Control P:E
250
150
100
75
50
37
25
C
Intensity
75
50
25
0
**
**
Ref.
Ref.
Ref.
Control E P:E Control E P:E



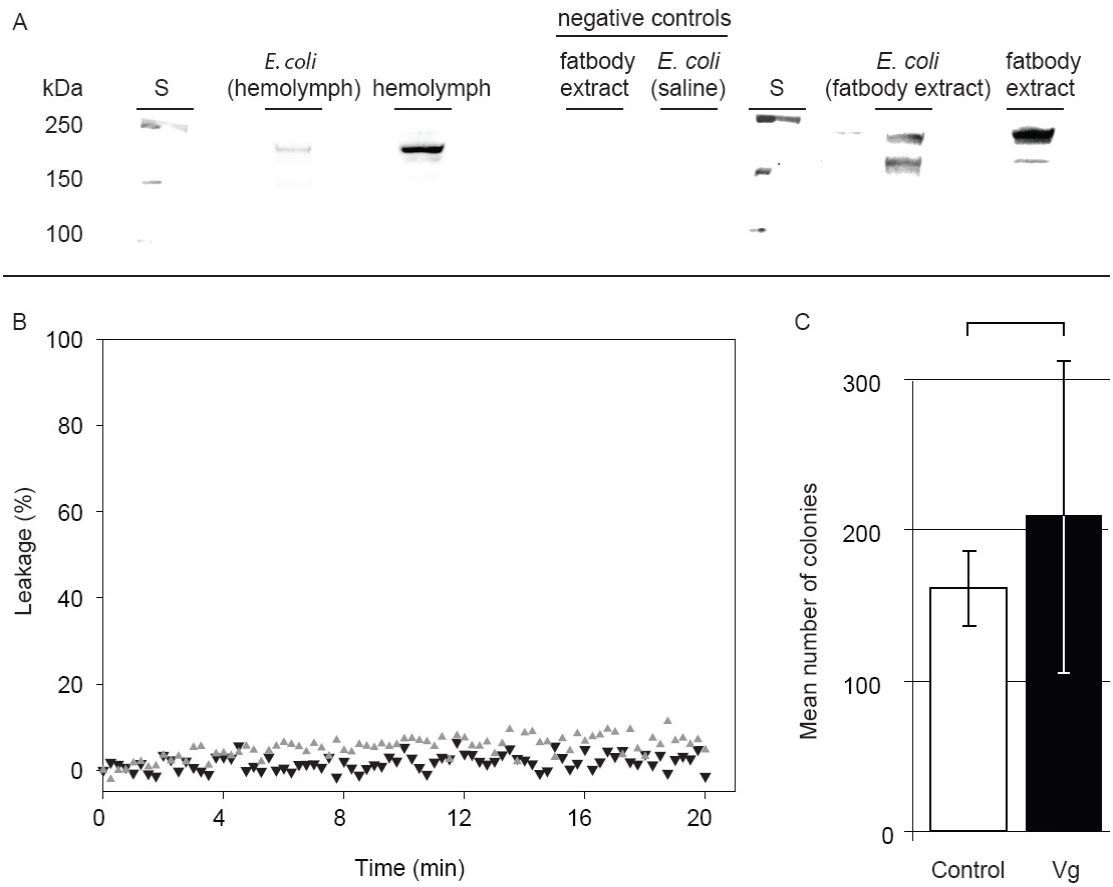
C

not modeled

```
A. mellifera          LNTAATLSFCEMIHNAQVNKRSIH---NNYPVHTFGRL---TSKHDNSLYDEYIPFL 51
B. ignitus            LNSSAIIAFSELVYNAQVSRKGLH---NHYPVHTYGRL---TPKHNKAVTEYYIPYL 51
B. hypocrite          LNSSAIIAFSELVYNAQVNRKGLH---NHYPVHTYGRL---TPKHDKTVTEYYIPYL 51
S. invicta (Vg2)      VNVSAPIALAELLR----NSYIGQ---NYYPIYSFGFI---TLKKNDEVVGKYINYL 47
P. nipponica          LNTTALFTFSELVRYSQMDNPSSH---TRYPVHTFGRF---SDKINPEVFRMYVPYL 51
P. puparum            LRDASILAFADLIRHAMVNQRSAH---NRYPVHAFGRL---ISKDSKNLLDKYLSYM 51
A. aegypti            LNATALFAYSNFVNQAHVSNRSAY---NYYPVFSFGRL---ADADYKIIEHKIVPWF 51
R. clavatus           LNTTALFAFTELSRKSQVDYDVAH---NRYPVHLHAKN---SQEHRQAVPQKYIPWL 51
A. grandis            LNQTCILSYTNLVHKVYINRNESH---NQFPVHAFGSFY--TKKGREFVKTTVIPHL 52
G. nigrofuscata       MNTSAILALGNMLRRTQVNKHIFH---SRYPTYTFGPM---AAKNN-HVLHEYTDKI 50
A. pernyi             LNTTALIAATRFINMGQVNNDTAH---NFYPTHMYGRL---ARKHDNFVLEQVLPPL 51
P. americana          VNSTGLLALATLHRQVH-DAEFSH---NNYPVHAFGRM----VPRNYNATNDFIDYL 49
L. maderae            VNETVVIAFSNAYRFIHARLKRPYI--S--PYFIKYLFQEFENAYRRQNTTQMQVYV 56
B. germanica          VNTSLVLAFSEVIHQVEMHQVRDLKIKSVYIPYLVQEFDDAVKENNSIKIQLYTHAL 57
A. rosae              LNNTAILAFADLVRESQVNNASAH---GRYPVHSFGRL---SPKNNPAVQGQWIPYI 51
E. formosa            LRDSSVLSFANLIRRAIVNKRSAH---NRYPVHVFGRM---FNRENSNLHEKYIPYI 51

I. unicuspis          LRKTAVLGYGSLVFRYCANTV-------------------------SCPDELLQP-L 31
A. transmontanus      LRKHAVLAYGSMVNRYCAETL-------------------------NCREEALKP-L 31
O. aureus             LNEIVMLGWGTVISRFCKAQP-------------------------SCSSDLVTP-V 31
O. mykiss             LRELTMLGYGTMVSKYCVEHP-------------------------NCPAELVKP-I 31
F. heteroclitus       LREVVFLGYGTMVNKYCNKTV-------------------------DCPVELIKP-I 31
X. laevis             VHKAAILAYGTMVRRYCDQLS-------------------------SCPEHALEP-L 31

human ApoB            RSRATLYALSHAVNNYHKTNP-------------------------TGTQELLDI-A 31
human MTP             IRETVMIITGTLVRKLCQNE-------------------------GCKLKAVVE-A 30
```



Peptide	$S_{0.5}$ (nM) E (neutral)	P:E (negative)
hlVg	367 ± 80	219 ± 26
fbVg	1603 ± 330	674 ± 130
α-helical	469 ± 81	694 ± 143
vWFD	No binding	No binding

Supporting information

```
   1 MLLLLTLLLF AGTVAADFQH NWQVGNEYTY LVRSRTLTSL GDLSDVHTGI
  51 LIKALLTVQA KDSNVLAAKV WNGQYARVQQ SMPDGWETEI SDQMLELRDL
 101 PISGKPFQIR MKHGLIRDLI VDRDVPTWEV NILKSIVGQL QVDTQGENAV
 151 KVNSVQVPTD DEPYASFKAM EDSVGGKCEV LYDIAPLSDF VIHRSPELVP
 201 MPTLKGDGRH MEVIKIKNFD NCDQRINYHF GMTDNSRLEP GTNKNGKFFS
 251 RSSTSRIVIS ESLKHFTIQS SVTTSKMMVS PRLYDRQNGL VLSRMNLTLA
 301 KMEKTSKPLP MVDNPESTGN LVYIYNNPFS DVEERRVSKT AMNSNQIVSD
 351 NSLSSSEEKL KQDILNLRTD ISSSSSSISS SEENDFWQPK PTLEDAPQNS
 401 LLPNFVGYKG KHIGKSGKVD VINAAKELIF QIANELEDAS NIPVHATLEK
 451 FMILCNLMRT MNRKQISELE SNMQISPNEL KPNDKSQVIK QNTWTVFRDA
 501 ITQTGTGPAF LTIKEWIERG TTKSMEAANI MSKLPKTVRT PTDSYIRSFF
 551 ELLQNPKVSN EQFLNTAATL SFCEMIHNAQ VNKRSIHNNY PVHTFGRLTS
 601 KHDNSLYDEY IPFLERELRK AHQEKDSPRI QTYIMALGMI GEPKILSVFE
 651 PYLEGKQQMT VFQRTLMVGS LGKLTETNPK LARSVLYKIY LNTMESHEVR
 701 CTAVFLLMKT NPPLSMLQRM AEFTKLDTNR QVNSAVKSTI QSLMKLKSPE
 751 WKDLAKKARS VNHLLTHHEY DYELSRGYID EKILENQNII THMILNYVGS
 801 EDSVIPRILY LTWYSSNGDI KVPSTKVLAM ISSVKSFMEL SLRSVKDRET
 851 IISAAEKIAE ELKIVPEELV PLEGNLMINN KYALKFFPFD KHILDKLPTL
 901 ISNYIEAVKE GKFMNVNMLD TYESVHSFPT ETGLPFVYTF NVIKLTKTSG
 951 TVQAQINPDF AFIVNSNLRL TFSKNVQGRV GFVTPFEHRH FISGIDSNLH
1001 VYAPLKISLD VNTPKGNMQW KIWPMKGEEK SRLFHYSVVP FVSNHDILNL
1051 RPLSMEKGTR PMIPDDNTSL ALPKNEGPFR LNVETAKTNE EMWELIDTEK
1101 LTDRLPYPWT MDNERYVKVD MYMNLEGEQK DPVIFSTSFD SKVMTRPDTD
1151 SENWTPKMMA VEPTDKQANS KTRRQEMMRE AGRGIESAKS YVVDVRVHVP
1201 GESESETVLT LAWSESNVES KGRLLGFWRV EMPRSNADYE VCIGSQIMVS
1251 PETLLSYDEK MDQKPKMDFN VDIRYGKNCG KGERIDMNGK LRQSPRLKEL
1301 VGATSIIKDC VEDMKRGNKI LRTCQKAVVL SMLLDEVDIS MEVPSDALIA
1351 LYSQGLFSLS EIDNLDVSLD VSNPKNAGKK KIDVRAKLNE YLDKADVIVN
1401 TPIMDAHFKD VKLSDFGFST EDILDTADED LLINNVFYED ETSCMLDKTR
1451 AQTFDGKDYP LRLGPCWHAV MTTYPRINPD NHNEKLHIPK DKSVSVLSRE
1501 NEAGQKEVKV LLGSDKIKFV PGTTSQPEVF VNGEKIVVSR NKAYQKVEEN
1551 EIIFEIYKMG DRFIGLTSDK FDVSLALDGE RVMLKASEDY RYSVRGLCGN
1601 FDHDSTNDFV GPKNCLFRKP EHFVASYALI SNQCEGDSLN VAKSLQDHDC
1651 IRQERTQQRN VISDSESGRL DTEMSTWGYH HNVNKHCTIH RTQVKETDDK
1701 ICFTMRPVVS CASGCTAVET KSKPYKFHCM EKNEAAMKLK KRIEKGANPD
1751 LSQKPVSTTE ELTVPFVCKA
```

Supplement 1. The LC-MSMS hits (red) of the honey bee Vg segment that resists chymotrypin digestion in the presence of liposomes. The α-helical domain is highlighted grey. The most reliable peptide hits with high Mascot Ion score (> 75) are underlined.

Supplement 2. Table 1. Statistics of the antibacterial activity test on honey bee Vg. The control and the Vg sample column show the mean colony number and the standard deviation.

Replicate	Control	Vg	N
1	219.0 ± 8.7	165.7 ± 11.0	3
2	235.3 ± 102.6	348.7 ± 68.2	3
3	50.7 ± 27.7	70.0 ± 25.2	3
4	117.7 ± 56.0	198.0 ± 82.5	3
5	201.6 ± 24.9	171.4 ± 19.7	5
6	132.8 ± 43.8	277.6 ± 107.2	5