Evolutionary and functional relationships of insect immune proteins

Marco Fabbri

Supervisors:

Prof. Otto Schmidt

Dr. Ulrich Theopold

A thesis submitted for the degree of

Doctor of Philosophy

in the Department of Applied and Molecular Ecology

The University of Adelaide

May 2003

Contents

Summary

Innate immunity has many features, involving a diverse range of pathways of immune activation and a multitude of effectors-functions. This thesis project examined different aspects of innate immunity. In the first part a novel gene family of putative C-type lectins is presented, which may have developmental and immune functions. In the second part, I explored the possible ancestral origin of immune-effector proteins by investigating salivary gland and silk proteins involved in coagulation. In the last part, novel immune genes with similarity to strictosidine synthase are presented and their role as antifeedent is discussed.

Glycodeterminants play an important role in mediating cellular and cell-substrate interactions during development and immune-related reactions enabling an organism to distinguish self determinants from non-self or modified-self determinants. The most studied sugar recognition molecules are lectins. They have a wide range of binding activities and they are organized in multigene families. Here I describe a group of *D. melanogaster* genes that are possible members of the C-type lectin family.

Characterization of a novel *D. melanogaster* hemocyte mucin revealed a gene-locus I71-7, which was identified as a salivary (labial) gland protein. These data suggest that I71-7 is expressed similarly to hemomucin and may take part in hemolymph coagulation and entrapment of microorganisms. To test whether labial gland proteins are expressed in the immune system of other insects, I studied two lepidopteran silk proteins and found them

to be expressed by immune tissues as well. The implications of labial gland secretory protein involvement in coagulation and its role in insect immunity are discussed.

I used conserved protein domains of the *Drosophila* immune receptor hemomucin to identify novel members of a gene family which have similarity to strictosidine synthase (SS), one of the key enzymes in the production of monoterpene indole alkaloids. In addition to the first animal member of the family described previously (hemomucin) a second *D. melanogaster* member could be identified, which appears to differ in subcellular distribution from hemomucin. In *Arabidopsis thaliana*, SS-like genes form a multigene family, compatible with a possible function as antifeedants and antibacterial compounds. In *Caenorhabditis elegans*, two members could be identified and one member each in *Mus musculus* and *Homo sapiens*. Interestingly, the human SS-like gene is strongly expressed in the brain, the very organ many of the indole alkaloids act upon.

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my Thesis, when deposited in the University Library, being available for loan and photocopying.

signed :

Marco Fabbri

Dated this May, 2003.

Introduction

Innate immunity

Genetically 'hard-wired' defense systems termed "innate immunity" are present in all individuals at birth and include all responses that are independent of specialized immunocompetent T and B Lymphocytes. Innate immunity provides a first line of defense against pathogens in vertebrates and is the only defense system in invertebrates. In vertebrates it limits the infection and provides signals required for the development of the adaptive immune response (Biron *et al.*, 1989; Fearon and Locksley, 1996).

Host defense in vertebrates is an intricate interplay between innate and adaptive responses, involving mechanisms that reflect the diversity of pathogens that infect them. It is important that these defense mechanisms be present in readiness at all times and rapidly activated upon infection to allow for the destruction of the infectious agents. It is also obvious that inactivation of pathogens must proceed without harming the host itself. Following this idea, Charles Janeway argued in 1989 that innate immunity must operate by means of receptors that have been selected over evolutionary time to recognize highly conserved and widely distributed features of pathogens, especially features that are not found on the cells of host organisms. He called these features "microbial patterns" and coined the term of "pattern recognition receptors" (PRRs) for the receptors that recognize them (Janeway, 1989).

PRRs are carried by particular types of cells, such as macrophages, natural killer (NK) cells, and probably also epithelial and endothelial cells in the lung, kidney, skin, and gastrointestinal tract (Unanue, 1984; Wright, 1991). They generally have a broad ligand specificity and fail to discriminate between pathogen species. As opposed to clonal molecules such as the immunoglobulins (Igs), their expression is not induced in response

to specific pathogen groups, but rather to broad classes of microbes such as Gram-negative or Gram-positive bacteria, and/or fungi. The innate immune response does not generate immunological memory, nor does it trigger long-lasting protection against disease. However, absence or genetic defects in these defense mechanisms can lead to recurrent infections, demonstrating their importance (Super *et al.*, 1992).

The recognition of pathogens by PRRs triggers their engulfment by immune competent cells leading ultimately to the destruction of the pathogens (Unanue, 1984). PRRs also participate in the elimination of tissue debris that accumulates during infection, inflammation, and wound repair (Akbar *et al.*, 1994; Savill, 1997a). Importantly, this engulfment process, called "phagocytosis", is also important during development, as it participates in the clearance of cells that undergo programmed cell death (PCD, or apoptosis) (Wyllie *et al.*, 1980; Raff, 1992; Savill, 1997b). Rapid phagocytosis of the apoptotic corpse by specific receptors prevents the release of secondary immune signals, which could affect the homeostasis of neighboring tissues (Wyllie *et al.*, 1980; Raff, 1992). Recent studies have also highlighted the importance of phagocytosis receptors with specificity for apoptotic cells in the suppression of inflammation (Akbar *et al.*, 1994; Savill, 1997a). Furthermore, recent data suggest that failure to dispose of apoptotic corpses could ultimately result in the activation of immune responses against self-antigens, which may contribute to autoimmune diseases such as lupus (Laderach *et al.*, 1998; Botto *et al.*, 1998).

Insects possess a complex and efficient system of biological defense against pathogens and parasites. This system comprises three different means: the integument and gut as first line of defense to infection, the responses within the hemocoel when these barriers are breached and the induced synthesis of antimicrobial peptides and proteins from various tissues, such as the fat body.

Recognition molecules

In order to identify a potentially damaging object or organism an immune system must be capable of recognising diagnostic features at a molecular level. Furthermore, this recognition process must be sufficiently precise in order to allow a defence reaction to take place in an accurate, effective and controlled fashion. Therefore, the process of recognising potentially damaging structures must be considered as the key event for a specific immune response.

Since the insect immune system is not adaptive, an immune response relies on a fixed number of recognition molecules that are specific for common microbial epitopes (Hultmark, 1993). They are thought to be representative for a broad spectrum of microbes, as indicated by the term "pattern recognition" (Janeway, 1994).

However, pattern recognition is probably not a complete description of innate immune recognition processes since abiotic and chemically inert particles (e.g. those made from Nylon), are able to stimulate immune responses *in vivo* (Salt, 1965). Therefore, it is conceivable that physical properties, like the net charge and the hydrophobicity of a surface, play a crucial role in immune reactions, whether or not a foreign object is recognised as "non-self" (Lackie, 1988). In fact, objects with neutral surfaces, when injected into the hemocoel of cockroaches, provoked a less intensive response than charged surfaces (Lavine and Strand, 2001). On the other hand, negatively charged surfaces were not encapsulated by hemocytes of certain caterpillars or locusts (Lackie, 1986). A few potential recognition molecules have been identified to date and some of them are described in the following sections.

Hemolin, or P4, a 48kDa protein found in the hemolymph of *Hyalophora cecropia* and *Manduca sexta* is considered as one of the candidate recognition molecules (Sun *et al.*, 1990; Ladendorff and Kanost, 1991). Interestingly, this molecule contains four internal repeats showing homology to C2-type immunoglobulin-like domains, indicating that it is a member of the immunoglobulin superfamily. Within this superfamily, hemolin shows the closest similarity to cell adhesion molecules, particularly to the insect proteins neuroglian and amalgam and the vertebrate proteins NCAM and L1 (Sun *et al.*, 1990). In unchallenged larvae, it is present at low levels - a prerequisite for putative recognition molecules. After bacterial infection, however, the concentration of hemolin increases up to 18-fold (Andersson and Steiner, 1987). Sun *et al.* (1990) could show that hemolin binds to the surface of bacteria, forming a complex with two other proteins. However, hemolin itself seems to display no direct antibacterial activity supporting the idea that hemolin functions as a recognition molecule. In fact, several lines of evidence indicate that hemolin is involved in the regulation of hemocyte behaviour. Upon immune challenge *in vitro*, hemolin binds to hemocytes and has the ability to inhibit the aggregation of hemocytes (Zhao and Kanost, 1996; Ladendorff and Kanost, 1991; Kanost *et al.*, 1994; Lanz-Mendoza *et al.*, 1996). In addition to its role in regulating hemocyte adhesiveness, hemolin is capable in *in vitro* experiments to stimulate the phagocytic activity of insect blood cells (Lanz-Mendoza *et al.*, 1996). However, the *in vivo* function of hemolin remains to be determined. Interestingly, (Bettencourt *et al.*, 1997) found evidence of a 52kDa membrane form of hemolin associated with hemocytes, indicating a multifunctional role of this putative recognition molecule in humoral and cellular immunity.

For another class of recognition factors, the lectins, it has been clearly shown that they play a role in insect immunity. However, their precise mode of action has still to be resolved. In general, lectins are widely occurring proteins or glycoproteins, sometimes with multiple binding sites, which specifically recognize certain carbohydrate moieties. They are inducible and show no enzymatic activity. Due to their ability to agglutinate vertebrate erythrocytes, bacteria, and other microorganisms, they are also referred to as agglutinins (Götz and Boman, 1985). In insects, they are found to be associated with the plasma membrane of hemocytes and as soluble hemolymph proteins (Ratcliffe, 1993b). For various lectins it has been shown that lectin-mediated recognition or clumping of foreign objects makes these objects susceptible to both cellular and humoral defense reactions. For example, a lectin from the grasshopper *Melanoplus differentialis*, described by (Wheeler *et al.*, 1993) was shown to display an opsonic activity against fungal blastospores. In *Periplaneta americana*, two lectins, named *Periplaneta* lectin (Kawasaki *et al.*, 1993) and LPS-binding protein (Jomori and Natori, 1992), respectively, have been isolated that participate in clearing of bacteria from the hemolymph. Both of them recognise lipopolysaccharide (LPS) from the outer membrane of gram-negative bacteria and are related in sequence to the C-type lectins of vertebrates. Another, galactose-specific lectin from the hemolymph of the flesh fly *Sarcophaga peregrina*, was found to play an active role in the lysis of sheep red blood cells injected into the hemocoel (Komano and Natori, 1985). These few examples from invertebrates show that lectins display opsonic activity and can be considered as recognition molecules capable of eliciting immune responses. Likewise C-type lectins in mammals are also involved in pattern recognition of microorganisms (Hoffmann *et al.*, 1999).

Another important group of recognition molecules are proteins involved in the activation of the phenoloxidase pathway. The end product of this pathway, melanin, plays an important role in the sclerotisation and tanning of the insect cuticle (Anderson *et al.*, 1985) and is responsible for sealing off cellular capsules that have been formed around an foreign object in the hemocoel (Salt, 1970). As exemplified in the case of already mentioned lectins that bind LPS, sugar-determinants exposed on the surface of an foreign object constitute important signals for the identification of bacterial microbes. In addition to LPS from gram-negative bacteria, the peptidoglycans from gram-positive bacteria and the mannans or beta-1,3-glucans from fungi have been described to evoke immune reactions, including the activation of the prophenoloxidase activating system (Sugumaran and Kanost, 1993; Söderhäll *et al.*, 1994). Ashida and colleagues were able to purify two soluble proteins from the hemolymph of the silkworm *Bombyx mori*, one of which is specific for peptidoglycans from bacterial cell walls, and the other for beta-1,3-glucans from fungal cell walls (Ashida and Yamazaki, 1990). When bound to their respective ligands, these proteins activate a proteolytic cascade, named prophenoloxidase activating system, that leads to the activation of the enzyme phenoloxidase which is generally considered to be the key enzyme for the synthesis of melanin (Sugumaran and Kanost, 1993). Since many cellular capsules are melanised, the question can be asked whether PO or prophenoloxidase (proPO), the non-activated form of PO, serves as a recognition molecule triggering the encapsulation reaction. In this context (Rizki and Rizki, 1990) could show that in larvae of proPO-deficient mutants of *Drosophila melanogaster* parasitoid eggs are encapsulated but not melanised. From this result they concluded that phenoloxidase is involved in the cross-linking and melanisation of capsules, but not in the recognition of foreign objects in *Drosophila* larvae. It seems that the recognition events

leading to the production of melanin and the ones leading to encapsulation are separate events.

Cellular response

The cellular defence reactions are mediated by the insect blood cells in the hemolymph, the hemocytes. Hemocytes from different insects show large variations regarding their morphological and functional characteristics. Therefore, hemocyte classification has proven to be difficult. However, two types of hemocytes, the plasmatocytes and the granulocytes, are generally thought to be the most important ones involved in cellular defence reactions (Ratcliffe, 1993a). Both plasmatocytes and granulocytes are polymorphic and variable in size. A distinguishing feature of granulocytes is that they possess numerous cytoplasmic granules. Another characteristic of this hemocyte type is the presence of a microtubule band near the periphery of the cell. In contrast, plasmatocytes are agranular or contain granules in the cytoplasm, which are considerably finer and less electron-dense; peripheral microtubule bands are absent (Gupta, 1991).

The invasion of foreign particles into the insect hemocoel evokes dramatic changes to the hemocyte population, involving surface morphology and adhesive properties (Wago, 1980b; Wago, 1980a). Granular hemocytes, for example, form numerous filipodia on their cell surface, important for entrapping foreign particles, and become increasingly adhesive to each other and to other surfaces (Wago, 1980a). When granulocytes get in contact with a foreign surface, they undergo degranulation, a process involving the discharge of granules and other cytoplasmic contents into the surrounding environment. The released components, which are generally thought to include phenoloxidase and lectins, act as opsonins that attract additional hemocytes and, in general, invoke the insect's defence reactions (Ratcliffe, 1993a). Nevertheless, the mechanisms by which

hemocytes recognise a foreign object are still the least understood facet of the cellular defence reactions. However, recognition molecules associated with hemocytes and present in cell-free hemolymph seem to be involved (Ratcliffe, 1993a). Depending on the number and size of the foreign particles in the insect hemocoel, three major defense reactions can be distinguished: phagocytosis, encapsulation, and nodule formation.

The principle means of removing small foreign objects such as bacteria from the hemolymph is phagocytosis, a process comprising recognition, attachment, and internalisation of the foreign object by hemocytes (Gupta, 1991). Phagocytic activity has been reported for both granulocytes and plasmatocytes. After a particle is recognized as foreign and attached to a hemocyte, it becomes ingested by endocytosis. In this process, the foreign particle is taken up by the hemocyte into either a coated vesicle, a membrane bound phagosome or a pinosome, each of which eventually fuses with a lysosome. Lysosomes contain enzymes and antimicrobial agents that finally digest the foreign object (Gupta, 1991). In the case of microbes, this process is believed to lead to the release of cell wall components, like lipopolysaccharides or peptidoglycans, that in turn are strong elicitors of antibacterial and antifungal defense reactions (Iketani and Morishima, 1993).

Foreign objects like metazoan parasites too large to be phagocytised are surrounded by hemocytes and become engulfed in a multi-layered cellular capsule. Similar to phagocytosis, both granulocytes and plasmatocytes have been shown to participate in encapsulation (Pech and Strand, 1996). However, in some species, only granulocytes are capable of encapsulation, a property that appears to be dependent on the granulocyte microtubule band (Gupta, 1991). The initial stage of encapsulation involves the

recognition of foreign antigens by granulocytes. After recognition, the granulocytes degranulate, release stored components, and eventually lyse. These components adhere to foreign surfaces and, in turn, attract additional hemocytes (Gupta, 1991). The next step leads to the formation of multiple cellular layers around the foreign object. The hemocytes in contact with the foreign particle form the innermost layer (Salt, 1970). They show elongation, flattening, and lysis. Melanization occurs, a process which physically separates and eliminates the foreign object from the hemolymph. However, gradually more hemocytes become attached, building up new layers. Finally, the capsule develops an outer layer which does not invoke any further attachment of hemocytes (Salt, 1970; Gupta, 1991). Interestingly, in *Pseudoplusia includens*, this outer layer is formed by a monolayer of granulocytes (Pech and Strand, 1996).

When large numbers of foreign objects, such as microbes, have entered the insect's hemocoel, nodule formation occurs (Salt, 1970; Ratcliffe, 1993a). This defence reaction is similar to encapsulation, and it also includes phagocytic activity by hemocytes. Like encapsulation, the initial step upon recognition of a foreign surface is the degranulation of granulocytes. But in this process, the granulocyte-derived material forms a coagulum containing degranulated and lysed granulocytes together with entrapped and phagocytosed microorganisms. Around this mass hemocytes proceed to attach and flatten, resulting in large cellular aggregates or nodules. These aggregates leave circulation by adhering to tissues and finally become encapsulated. Melanisation may also occur at the core of the nodule (Gupta, 1991).

Hemolymph coagulation

Hemolymph coagulation in insects is not well understood, but hemolymph coagulation in the horseshoe crab is the most completely characterized invertebrate defense response (Iwanaga *et al.*, 1998). Horseshoe crab hemocytes are the basis of the *Limulus Amebocyte Lysate* (LAL) assay for endotoxin, and this clinical application is an example for the high sensitivity of invertebrate immune recognition.

Horseshoe crab hemolymph coagulation is initiated by the presence of two types of pathogen surface molecules, the bacterial lipopolysaccharides (LPS) and the fungal beta-1,3-glucans. Factor C, a protein with both a serine protease catalytic domain and an LPS binding region, binds to LPS in the hemolymph, undergoes a change in conformation, and autoactivates. Activated Factor C cleave another serine protease, Factor B, which then cleaves a third serine protease, proclotting enzyme (PCE). Factor B and PCE are both members of the clip domain family of serine proteases. These serine proteases have an N-terminal domain characterized by six cysteine residues, that fold the domain into a disulfide knot resembling a paper clip (Muta *et al.*, 1990). They are often associated with invertebrate immune processes (Jiang and Kanost, 2000). Unfortunately, the function of clip domains is unknown, but they are thought to regulate catalytic domain activity by interacting with regulatory molecules or substrates. PCE can also be activated by Factor G, whose serine protease subunit becomes activated after a second subunit binds to beta-1,3-glucan. Finally, activated PCE catalyzes the last proteolytic step in the pathway, the cleavage of coagulogen to form coagulin. Coagulin self-aggregates to create a gel-like clot that traps pathogens. Whether insects have a similar defense response is unknown.

Virtually all multicellular organisms have developed unique modalities to detect and respond to microbial surface antigens, such as lipopolysaccharide (LPS), peptidoglycan and beta-glucan. Because both invertebrates and vertebrate animals respond to these substances, it is likely that a system recognizing these epitopes emerged at a very early stage in the evolution of animal species. It is well known that components of the bacterial cell wall elicit various responses, depending on the animal species or cell type. In invertebrates, hemolymph coagulation (Iwanaga, 1993) and melanin formation are the most visible responses (Söderhäll and Cerenius, 1998). In addition to these enzymatic cascades, a variety of antimicrobial substances and lectins also cooperate with the reactions for killing the invaders (Boman, 1998).

To date, two types of clotting mechanisms have been reported in invertebrate animals. One of these is found in crustaceans (lobster and crayfish) and insects (cockroach and grasshopper) (Kopacek *et al.*, 1993; Brehelin, 1979) where a gel is formed through the polymerization of clottable protein(s), catalyzed by Ca^{2+}-dependent transglutaminase. The translutaminase is released from the hemocytes or muscle cells through an unknown mechanism. The clottable proteins isolated from lobster and crayfish hemolymph plasma have similar characteristic to each other (Fuller and Doolittle, 1971; Kopacek *et al.*, 1993). They are large glycoproteins consisting of dimers with about 200kDa subunits. Their amino-terminal sequences suggest that these are vitellogenin-like proteins (Fuller and Doolittle, 1971; Doolittle and Riley, 1990; Hall *et al.*, 1999). In Insects, another abundant plasma protein, lipophorin, seems to be cross-linked upon gelation (Barwig, 1985; Brehelin, 1979; Duvic and Brehelin, 1998).

The other type of coagulation is activated by a cascade-type reaction composed of serine protease zymogens as in the mammalian blood coagulation system. In horseshoe crab (or limulus), two pathways involving either a LPS-mediated or a beta-glucan mediated

coagulation reaction have recently been established and all essential clotting factors are now disclosed (Muta *et al.*, 1995; Muta and Iwanaga, 1996).

Exocytosis of hemocytes mediated by lipopolysaccharide

Hemocytes in the hemolymph of the horseshoe crab play a major role in the defense system of this animal. In contrast to hemocytes of insects or other invertebrates, a single type of hemocyte circulates in horseshoe crab hemolymph (Iwanaga, 1993). The hemocyte, which is also called granulocyte or amebocyte, contains two types of secretory granules: large granules and small but dense, granules (Toh, 1991). This cell is extremely sensitive to bacterial endotoxins (e.g. LPS). When Gram-negative bacteria invade the hemolymph, the hemocyte detects LPS molecules on the surface of the pathogen causing the release of granules by rapid exocytosis. The released granular components include two regulators of the coagulation reaction, factor C and factor G (Muta *et al.*, 1995). These serine protease zymogens are autocatalytically activated by LPS and beta-glucans, which are major cell wall components of Gram-negative bacteria and fungi, respectively. The activation of these two zymogens triggers the coagulation cascades, resulting in the conversion of coagulogen to an insoluble coagulin gel. The invaders in the hemolymph are thus engulfed or immobilized by the clot. They are subsequently cell-agglutinated and killed by various lectins and antimicrobial substances that are also released from the granules (Muta and Iwanaga, 1996).

Coagulation mediated by lipopolysaccharide

The LPS-mediated coagulation cascade involves three serine protease zymogens (factor C, factor B, and proclotting enzyme) and a clottable protein, coagulogen. Factor C

(123kDa) is a biosensor that responds to LPS (Muta *et al.*, 1991). In the presence of LPS or synthetic lipid A analogues, it is autocatalytically activated to an active form, factor C which is composed of three chains held together by disulfide bonds: H chain (8kDa), N chain (7.9kDa) and B chain (34kDa). Factor EN (64kDa) is then activated by factor C and in turn, its active form (factor B) activates proclotting enzyme to clotting enzyme (54kDa) (Muta *et al.*, 1993). The active clotting enzyme converts coagulogen to an insoluble coagulin gel (Iwanaga, 1993).

All the three zymogens contain a serine protease domain at their carboxyl terminus. Their amino-terminal end, however, shows interesting structures, indicating a mosaic protein probably derived from exon shuffling. The H chain of factor C, which binds LPS, contains five "sushi" domains, an epidermal growth factor (EGF)-like domain, and a C-type lectin like domain (Muta *et al.*, 1991). The finding of "sushi" (also called SCR or CCP) domains in factor C made it the first protein in invertebrates that has been discovered to have this type of domain. The fact that this initiator of the horseshoe crab clotting cascade contains "sushi" domains that are found mainly in mammalian complement factors, led us to speculate that both coagulation and mammalian complement systems may have evolved from a common origin.

The amino-terminal L chains of factor B and proclotting enzyme contain a small compact domain with three disulfide bonds, called "clip" domain (formerly called "disulfide-knotted" domain) (Muta and Iwanaga, 1996). The "clip" domain has been found in the *Drosophila snake* and *easter* protease precursors (Smith and DeLotto, 1992). Both *easter* and *snake* proteins are indispensable for the normal embryonic developments in flies (Chasan and Anderson, 1989; DeLotto and Spierer, 1986). The presence of this type of domain in *Drosophila* strongly suggests the existence of a protease cascade system similar to that of the horseshoe crab.

The folding pattern of the three disulfide bridges in the "clip" domain is identical to that of "big defensin", which was recently identified as the third antimicrobial protein in horseshoe crab hemocytes (Saito *et al.*, 1995). As the carboxy-terminal end of the "clip" domain in proclotting enzyme constitutes a hinge region susceptible to protease attack (Muta *et al.*, 1990), the "clip" domain(s) might be released during the activation of the zymogens to work as antimicrobial substances. If this is the case, the coagulation cascade itself can produce antimicrobial substances during the activation. The system has dual action: coagulation to seal wound and the killing of invading microorganism.

Coagulation mediated by beta-glucan

The purified factor G zymogen is autocatalitically activated in the presence of beta-glucan, without any other proteins (Muta *et al.*, 1995). The resulting active factor G activates proclotting enzyme directly, which links with the coagulin gel formation. As factor G co-localizes in the large granules together with components participating in the LPS-mediated coagulation cascade, it can be released into hemolymph upon cell activation. This beta-glucan mediated coagulation pathway is activated on the surface of fungi.

Lectins

Lectins are protein complexes with multiple binding sites, which specifically recognize certain carbohydrate moieties, agglutinate cells or precipitate polysaccharides and glycoproteins. Because they are mostly polyvalent, each lectin molecule has at least two carbohydrate binding sites allowing cross linking between cells (by combining with

sugars on their surfaces) or between sugar-containing macromolecules (Kaltner and Stierstorfer, 1998).

Historically, the specificity of sugars has dominated the assignments of lectin categories. Since this feature does not reflect an evolutionary relationship, it has been replaced where possible by categorization based on data sets consisting of protein domain sequences and topologies. Knowledge about carbohydrate recognition domain (*CRD*), the basic unit for legitimate classification of a protein as lectin, is enabling computer assisted homology searches to find further members of lectins subgroups among the steadily increasing number of sequenced genomes (Gabius, 1997).

The lectins are classified into five different groups, based on structural alignments. The main group is the C-type lectin, the two prerequisites for defining a C-type lectin are the dependence of sugar binding on the presence of Ca^{2+} ions and the preservation of a common sequence motif of 14 invariable and 18 highly conserved amino acid residues (Gabius, 1997). This family has already been divided into several subgroups, because sequence alignments have enabled the detection of regions with similarity to the initially defined carbohydrate recognition domain in structurally otherwise unrelated proteins (Gabius, 1997). This obvious dissemination of a homologous CRD can reflect an adaptation to the physiological need to confer the capacity for protein-carbohydrate recognition in a different context (Gabius, 1997).

Little is known about their functions, but it is generally believed that lectins serve primarily as recognition determinants. In mammalia they augment innate immunity before the establishment of a mature network of clonally selected antibody gene systems with recombinatorial diversity. Lacking the adaptive immunity based on B and T cells, invertebrates have developed a defense system which shows prodigious similarities to innate immune mechanisms in vertebrates (Hultmark, 1993) attracting invading

organisms (bacteria) and associated substances, such as lipopolysaccharides, and (in the course of normal development) of altered tissue constituents that are no longer required. Thus recognition in defence and development is counted among the major functional roles of invertebrate C-type lectins (Kaltner and Stierstorfer, 1998).

Because insect lectins have been implied in immune recognition, it is tempting to speculate that the lectin family might form part of the pattern recognition repertoire, which have been postulated to play a role in innate immune recognition. However, carbohydrate determinants, which are expressed and modified as part of a normal developmental program, might also be recognized as immune elicitors in the same organism.

Glycosylation

Glycosylation is a post-translational modification that occurs in the rough endoplasmic reticulum and Golgi apparatus. Secretory and membrane-bound proteins are glycosylated to give specific biological function, structure and stability.

There are two types of glycosylation, N- and O-linked, distinguished by the enzymes involved and the amino acid to which the sugars are added. The nitrogen of the asparagine is glycosylated by the addition N-acetyl-D-Glucosamine (GlcNAc) in N-glycosylation (Rademacher *et al.*, 1988). O-glycosylation involves the addition of N-Acetyl-D-Galactosamine (GalNAc) to a Serine or a Threonine residue (Rademacher *et al.*, 1988).

Glycosylation reactions are of great importance to both mammalia and insects, and require the coordinated action of a large number of enzymes the glycosyltransferases. These enzymes transfer the sugar from an activated nucleotide sugar to an acceptor, which may be a lipid or a protein (Breton and Imberty, 1999).

Oligosaccharides exist mainly in covalent association with proteins affecting both intramolecular and intermolecular functions. Changes in protein stability and solubility that occur when glycosylation variants are generated suggest that intramolecular interactions involving both carbohydrate and polypeptide define the physical properties of glycoproteins. The structural and functional roles of oligosaccharide moieties of proteins are interdependent and variation of sugar modifications at one position affects the functional role of another.

At any developmental stage, cells may have solved the biosynthetic problem of controlled variation by making not just one glycoprotein, but coding for large repertoires of proteins, each variant with slightly different covalently attached oligosaccharides. This pool of protein glycoforms contains distinct members, each of which may have a unique spectrum of biological activities while maintaining the intramolecular characteristics necessary for structural integrity (Adams and Watt, 1993). Hence a glycoprotein has a composite activity, which a cell can control via a post-translational process by varying the relative incidence of each glycoform (Adams and Watt, 1993). Displaying the entire spectrum of oligosaccharides simultaneously may be necessary to avoid immunogenicity at later developmental stages (Adams and Watt, 1993).

Glycodeterminants are frequently found on cell surfaces; this location together with the fact that glycodeterminants are often large in comparison to the proteins to which they are attached, makes them well suited for the mediation of interaction and recognition events (Geisow, 1991). Different glycodeterminants may be assembled to form combinatorial structures. The three dimensional structures allow them to be recognized by receptors (lectins) and to mediate events such as cell adhesion, cell trafficking and immune recognition.

Glycosylation in Insects

Important roles for mammalian glycoproteins in various biological functions are now well documented, while insect glycoproteins remain poorly investigated. The data available at present are mainly concerned with glycoproteins having N-linked chains showing some similarity between the mechanism of N-glycosylation in insects and vertebrates (März *et al.*, 1995). Insect O-glycoproteins are much less characterized (März *et al.*, 1995).

The first evidence for the developmental importance of insect glycoprotein was obtained in neural cells (Bastiani *et al.*, 1987). During studies on *Drosophila* and grasshopper nervous systems two glycoproteins, fasciclin I and II, were discovered on specific subsets of axon pathways, which differ in their glycoforms; these two glycoproteins are dynamically expressed during specific embryonic stages. These findings led to the speculation that, as in vertebrate organisms, structural alterations in insect glycodeterminants may act as a modulator of cell recognition and adhesion. More recently several authors showed the pivotal role of glycodeterminants as a modulator during different developmental stages in a variety of tissues (Adams and Watt, 1993; März *et al.*, 1995; Mann and Waterman, 1998).

Glycosyltransferases

Glycosyltransferases catalyze the synthesis of glycoproteins by transferring an activated sugar to a protein for the initiation or elongation of the carbohydrate chain (Kapitonov and Yu, 1999).

Because the glycosylation is highly specific with respect to both the sugar residue and the site and configuration of the linkage, it is expected that unique domain structures for

substrate recognition and nucleotide-sugar binding are located within the enzymatic cleft of the molecule. Indeed common amino acid sequences have been deducted for homologous binding sites.

During the past few years the amino acid sequence of a number of glycosyltransferases has been identified using protein sequence data provided by the complete genomic sequences obtained for a number of organisms, including *C. elegans, S. cerevisiae, D. melanogaster*, human and mouse genome projects.

Structural studies dominating the field of glycobiology have revealed an extensive array of naturally occurring oligosaccharides configurations resulting from the coordinated activity of many glycosyltransferase enzymes. Most of these enzymes are expressed in a temporal- and/or tissue-specific manner and numerous hypotheses have emerged regarding the biological significance of the glycoconjugates formed, however no functional role has been deduced (Varki, 1993).

The modulation of glycosyltransferases *in vitro* has yielded considerable information regarding the functions of specific oligosaccharides structures in protein folding, cell adhesion and signal transduction.

Mucins

Mucins are in general large, highly glycosylated, O-linked glycoproteins. Up to 90% of the total weight of mucins can be carbohydrates or sugars (Gum, 1995). Mucins can be membrane bound or secreted and can have different properties. Membrane bound glycoproteins, such as episialin (Muc I), act as cell surface receptors and protect against pathogens as they protect the cell membrane from invasion. Muc II, a secreted form of Muc I, is larger than the membrane bound form and has gel forming properties and high viscosity (Gum, 1995).

Mucins are most commonly known in vertebrate immunity (Shimizu and Shaw, 1993). Mucus is a layer of O-linked glycoproteins that protects the epithelial surface of tissues, such as those in the gastrointestinal and respiratory tracts. Mucus consists of many proteins secreted by epithelial specialised cells. Surface mucins, glycoproteins proteins attached to the cell membrane, are involved in extravasation, the movement of leukocytes through the endothelial membrane to the site of infection (Shimizu and Shaw, 1993). Mucins are present on the surface of cells and bind to a specialised type of lectin, called selectins. This cell to cell interaction between cell surface molecules aids in the rolling, activation, adhesion and trans-endothelial migration of circulating blood cells involved in extravasation (Shimizu and Shaw, 1993).

Butters *et al.* (1978) initially discovered O-glycosylated glycoproteins in insects in a mosquito cell line. Since Butters pioneering work, O-glycosylated glycoproteins have been found in many insects, with a diverse range of functions.

Sialic acid, a neuraminic acid derivative, adds charge to a protein. Sialic acid has been only recently reported in insects by a small number of research groups (Davidson *et al.*, 1991, Roth *et al.*, 1992). Sialic acid appeared to be developmentally regulated from the blastoderm to the third larval instar. Polysialic acid (PSA) was detected on the cephalic brow and thought to be a neural cell adhesion molecule similar to those in chick embryos as reported by Kerscher *et al.* (1995). Highly sialylated and negatively charged molecules present early in development in immature chick embryos restrict homophilic binding until the chain length is reduced and the charge is neutral. Once shortened, the neural cells bind and develop into the central nervous system. Differential glycosylation of N-CAM aids in the development of neurons (Kersher *et al.*, 1995). In accordance with Roth's (1992) results, PSA may be involved with neural development in *Drosophila*. These

results are inconclusive though, as Roth *et al.* (1992) performed no structural analysis of the sugars detected with *Limax flavis* lectin (which may only detect sialic acid).

Pearson (1996) discovered a putative O-glycosylated cell surface receptor, a scavenger receptor, dSR-C on the surface of hemocytes. Scavenger receptors bind foreign antigens such as lipopolysaccharide, glucans and bacteria. This binding leads to an activation of the innate immune system.

Antibacterial peptides, such as cecropins, are small cationic proteins that bind to infectious particles after release during an immune response. Bulet *et al.* (1993) described the discovery of drosocin an O-glycosylated antibacterial peptide in *Drosophila*.

Hemomucin

Another o-glycosylated protein, hemomucin has been described on the membrane of hemocytes in *D. melanogaster* (Theopold *et al.*, 1996). The isolation of hemomucin, exhibiting similar characteristics to immune-related mucins in vertebrates was facilitated by the use of a lectin, from *Helix pomatia* (*Hp*) which binds to mucins such as episialin and leukosialin on vertebrate leukocytes, as well as membrane-bound mucin on hemocytes and hemocyte derived-cells (mbn-2 cell line). This mucin, now named hemomucin, was found to have a role in insect immunity, inducing an immune response and the release of cecropins when the lectin was applied to the cells (Theopold *et al.*, 1996).

In addition to hemocytes, the membrane-bound form of hemomucin was discovered on the peritrophic membrane of the gut and in the ovaries. In ovaries, hemomucin was deposited on the outer membrane of the egg-shell - the chorion. Follicle cells, that secrete protein components for the chorion assembly, showed a strong response in *in situ*

hybridisation experiments and are postulated to be the source of chorion hemomucin (Theopold *et al.*, 1996).

Two forms of hemomucin, distinguished only by their size have been reported: the 100 and 220kDa forms of hemomucin are products from the same gene as they have the same protein sequence, but it is unknown how the 220kDa iso-form is produced (Theopold *et al.*, 1996). More recently, a 105kDa form was reported (Theopold *et al.* 2001). The difference between the 100 and 105kDa glycoform is the presence of one type of sugar. The 105kDa form contains a beta-attached galactose to the N-Acetyl-Galactosamine (GalNAc), which is recognized by peanut agglutinin (PNA). Interestingly, this form of the protein is only found in the ovary, whereas the 100kDa form is found in the larval organs. Hemomucin has a role as a cell surface protein in immunity, but it may also be involved in development as hemomucin is activated by ecdysone treatment of hemocyte-like cells. Different glycoforms are expressed in different developmental stages of *D. melanogaster* in the presence of ecdysone. The differential expression of hemomucin begs the question as to whether it has a role in development and whether it is hormone regulated. One possible function is for hemomucin to label cells and tissues for remodelling during metamorphosis. This labelling could aid in recognition of "larval" for destruction or programmed cell death in ecdysone-mediated processes. This observed change in glycoform from 100kDa in the larvae to 105kDa might be due to activation or modification of glycosyltransferases by ecdysone.

Also of interest is the possible function of hemomucin in the embryo and on the eggshell. It is still unknown which function hemomucin has on the eggshell and why the PNA-glycoform is produced.

Strictosidine synthase

The toxicity of plants towards some plant-eating animals, which contributes to their ability to protect themselves against predation, is partially related to the diversity of small metabolites that they synthesize. Alkaloids, which display a large variety of effects, some with pharmaceutical activities, comprise one of the major classes of plant metabolites and accumulate in various plant tissues in about 20% of all plant species (De Luca and Laflamme, 2001).

Amongst plant alkaloids, some monoterpenoid indole alkaloids have attracted particular interest due to their ability to act as toxins (strichnin), antimalarial (quinine), antineoplastic (vincristine and vinblastine) (Kutchan, 1995) or antipsychotic drugs (reserpine) (Bennett, 1998). This led to the identification of a number of enzymes that are part of their biosynthetic pathway including tryptophane decarboxylase and strictosidine synthase (SS), two key enzymes in the production of monoterpenoid indole alkaloids (Kutchan *et al.*, 1988). SS has been isolated from two plant species, which are known for their pharmaceutical benefits, namely *Catharanthus roseus* (madagascar periwinkle) (McKnight *et al.*, 1990) and *Rauvolfia serpentina* (sarpagandha plant) (Kutchan *et al.*, 1988).

Drosophila melanogaster as a model

D. melanogaster is a little insect about 3 mm long that accumulates around fermenting fruit. It is also one of the most valuable model organisms in biological research, particularly in genetics and developmental biology given that there are more mutants and techniques available than for most other model organisms. *Drosophila* has been used as a model organism for research for almost a century, and today the interest is even greater

because of the finalization of the genomic project. Therefore, it is possible to identify *Drosophila* homologues of known genes using bioinformatics tools. Furthermore as an invertebrate, *Drosophila* does not have an adaptive immune response. Therefore, it allows us to study the innate immune system without the added complication of adaptive responses.

A Lectin multigene family in *Drosophila melanogaster*

Introduction

Carbohydrate determinants constitute an informational system, the "glycocode", with even greater combinatorial diversity than both nucleic acids and polypeptides, due to the variability in the configuration of glycosidic bonds (Kaltner and Stierstorfer, 1998; Mann and Waterman, 1998). The regulatory potential of the information contained in the glycocode has been recognised during recent years (Acta Anatomica, issue on glycobiology, 161, 1998). Glycodeterminants play a role in mediating cellular and cell-substrate interactions during development and during immune reactions (Gabius, 1997). In addition, they confer signals to the immune system, which allow an organism to distinguish self determinants from non-self or modified-self determinants (Lanier, 1998). In order to be able to recognise a wide range of activities, sugar recognition molecules (lectins) are often organised in multigene families. One such system is encoded by a gene locus described in natural killer cells (NK cells) (Lanier, 1998; Unkeless and Jin, 1997). This locus codes for a variety of lectins, some of which recognise self-determinants and deliver inhibitory signals to the cells via a specialised intracellular inhibitory domain, whereas others activate NK cells upon binding of non-self carbohydrate determinants as found on pathogens and aberrant cells including tumour cells (Lanier, 1998). The combined activities of a cell's surface lectins determine on its activation status.

In invertebrates, a number of lectins are known to bind to self-determinants, whereas others recognise foreign elicitors of the immune response, like lipopolysaccharides (LPS) (Kawasaky *et al.*, 1996). A family of lectins has been

described in the American cockroach (Kawasaky *et al.*, 1996). In *Drosophila melanogaster*, a number of lectins or lectin-like molecules have been characterised, including a humoral member of the C-type lectin family with specificity for galactose (Haq *et al.*, 1996) and a member of the selectin family (Leshko-Lindsay and Gorces, 1997). C-type lectins share a number of conserved amino acid residues that can be found in both vertebrate and invertebrate members of the family. Here a group of 19 *Drosophila* genes is described, these genes are possible members of the C-type lectin family, using the information available through the BDGP/HHMI EST Project.

Materials and Methods

Sequence similarity searches

Sequence similarity searches were performed on the NCBI server or the BDGP server using the blast algorithms (Altschul *et al.*, 1990). The search results were analyzed for the presence of amino acids which are known to be conserved in C-type lectins, in particular conserved cysteines. Sequences with a significant similarity to the original sequence were downloaded into the Lasergene program package (DNASTAR Inc., Madison, WI, USA) and further aligned using Megalign. Using sequences from the original search with lower similarity scores, the procedure was repeated until no additional sequence similarities were found and the search considered exhaustive. In order to exclude intron sequences, only regions, which showed significant sequence homology extending well beyond the region used for the comparison in Fig. 1 were included in further analysis. In addition, the genomic regions were analysed using Genfinder. Only regions, which were well

located within a predicted exon were analysed further. Using Genfinder the results were in agreement with the genomic structure of the only characterised *Drosophila* lectin (Haq *et al.*, 1996), for which the genomic structure could be determined using the genomic sequences in comparison to the cDNA sequence. It was not possible to determine the genomic structure for all lectin-like sequences using Genfinder. In some cases, inspection by eye allowed to identify regions in the predicted open reading frame, which showed significant homology to other lectins but failed to be identified by Genfinder. Chromosomal localisations for P1 clones and BACs were available from the BDGP/HHMI EST Project (Hartl et. al., 1994; Kimmerly *et al.*, 1996).

Results

Novel lectin-like sequences in *Drosophila*

In order to identify novel C-type lectins from *Drosophila*, the protein sequence of known *Drosophila* lectins were used in sequence comparisons with the sequences obtained so far by the BDGP/HHMI EST Project (unpublished). The best results were obtained using a galactose-specific lectin (Haq *et al.*, 1996). Further analysis were run for those sequences that showed conservation of amino acid residues that are known to be essential in C-type lectins (Fig. 1). A total of 19 additional open reading frames where obtained which could be aligned with the original lectin used for the search. The majority of the ORFs is derived from genomic clones, some are from ESTs. The EST sequences were only analysed further if they covered the region, which is relatively well conserved among C-type lectins. For a number of sequences, both genomic and EST sequences could be identified. Two of the EST

sequences had previously been identified as lectin-like, one of them showed similarity to incilarin, a snail lectin (Yuasa *et al.*, 1998), a second showed similarity to a lectin-like protein from *Periplaneta americana* (Kawasaki *et al.*, 1996). One cDNA with similarities to lectin-like proteins has previously been isolated and shown to code for a *Drosophila* male accessory gland protein (Wolfner *et al.*, 1997). Some of the ORFs are encoded by the same genomic clone and are thus physically linked in the genome. As an example, two additional lectin-like ORFs were identified on the genomic fragment that encodes the galactose-specific lectin. In addition, the majority of the ORFs map to the same chromosome arm (2L) between positions 21-37 (Hartl *et al.*, 1994; Kimmerly *et al.*, 1996). For a more detailed sequence comparison, a particular protein region was chosen, since it is well conserved in C-type lectins including four conserved cysteines and a number of additional amino acid residues, which were present in all ORFs obtained. All sequences show significant similarity to each other and to a consensus sequence, which has been established for C-type lectins. While lectin sequences showed a number of amino acids in certain positions to be highly conserved, many positions in between were highly variable, including several gaps. This is reminiscent of the situation in the immunoglobulin superfamily (Sun *et al.*, 1990), where a number of highly conserved amino acids are known to stabilise the protein folding with areas of highly variable regions in between being responsible for the binding properties.

Discussion

It can be proposed the existence of a multigene familiy in *Drosophila*, which potentially codes for proteins with lectin-like activity. The gene structure seems to be relatively simple, a fact that allowed to analyse ORFs directly without the need

to determine the exact organisation of the gene using cDNAs. In order to exclude pseudogenes, the analysis was restricted to parts of the ORFs, which could be unambiguously aligned with sequences coding for fully characterised lectins. In one case, an ORF was identified, showing significant sequence homology with C-type lectins but contained stop codons (AC004296, localised in the region 55 B1-55 B4). The fact that lectin genes have been duplicated a number of times points to the possible existence of different ligands for each lectin. Most members of this family are located on the left arm of chromosome 2 (Tab. 1). This is in contrast to the distribution of other gene families, including a number of UDP-glucuronosyltransferases (UGTs)(Tab. 2), which are scattered throughout the genome even though their sequences are more conserved than the lectin-like sequences described here. A sequence comparison of *Drosophila* UGTs (Fig. 2) along with the genomic localisation of the sequences (Tab. 2) are shown. The observation of a lectin gene cluster on a chromosome section might point towards a more recent origin of the lectin-like sequences as compared to the UGTs (Tab. 1 and 2).

Because lectins have been implied in immune recognition, it is tempting to speculate that the lectin family might form part of the pattern recognition repertoire, which has been postulated for innate immune recognition (Medzhitov and Janeway, 1997). However, carbohydrate determinants, which are expressed as part of a normal developmental program in the same organism, might also be recognized by members of the family. In fact the two possible functions are not mutually exclusive. Whatever the ligands for the putative lectins might turn out to be, the fact that this multigene family was identified in an insect adds to the similarities between invertebrates and vertebrates, which use multiple lectins for self and non-

self recognition. Using prediction programs to identify the complete ORF on the genomic sequences, the efforts to identify signal transduction domains similar to the inhibitory domain found in several vertebrate NK cell lectins have been without success so far (Unkeless and Jin, 1997). Whether inhibitory lectins are present in invertebrates has to await for the completion of the sequencing of the complete *Drosophila* genome to be conclusive. Nevertheless the identification of an interesting multigene family in an invertebrate organism suggests that the preconditions for the recognition of a variety of sugar determinants exist in *Drosophila*. This observation can therefore be used to study the regulatory function using the genetic and molecular techniques available in this organism.

Figures

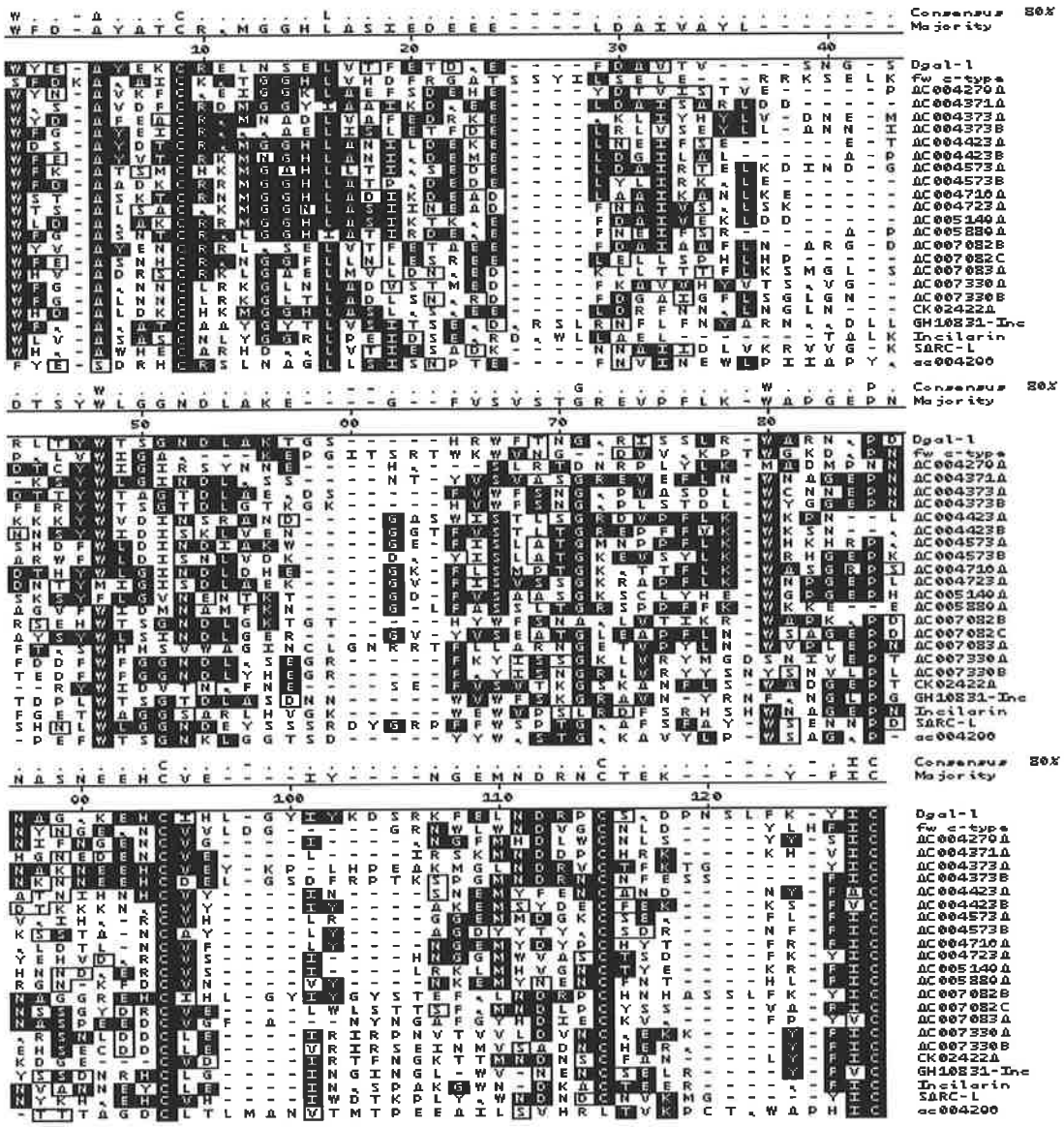

Fig. 1: Sequence comparison between the galactose specific *Drosophila* lectin (Dgal-1), the C-type lectin domain of the furrowed gene product (fw C-type), *Sarcophaga* lectin (SARC-L), Incilarin and newly identified *Drosophila* ORFs with sequence homology to C-type lectins (see table 1). Two consensus sequences are given, one which is shared by the majority of sequences and a second one, which is common to more than 80% of the sequences (consensus 80%).

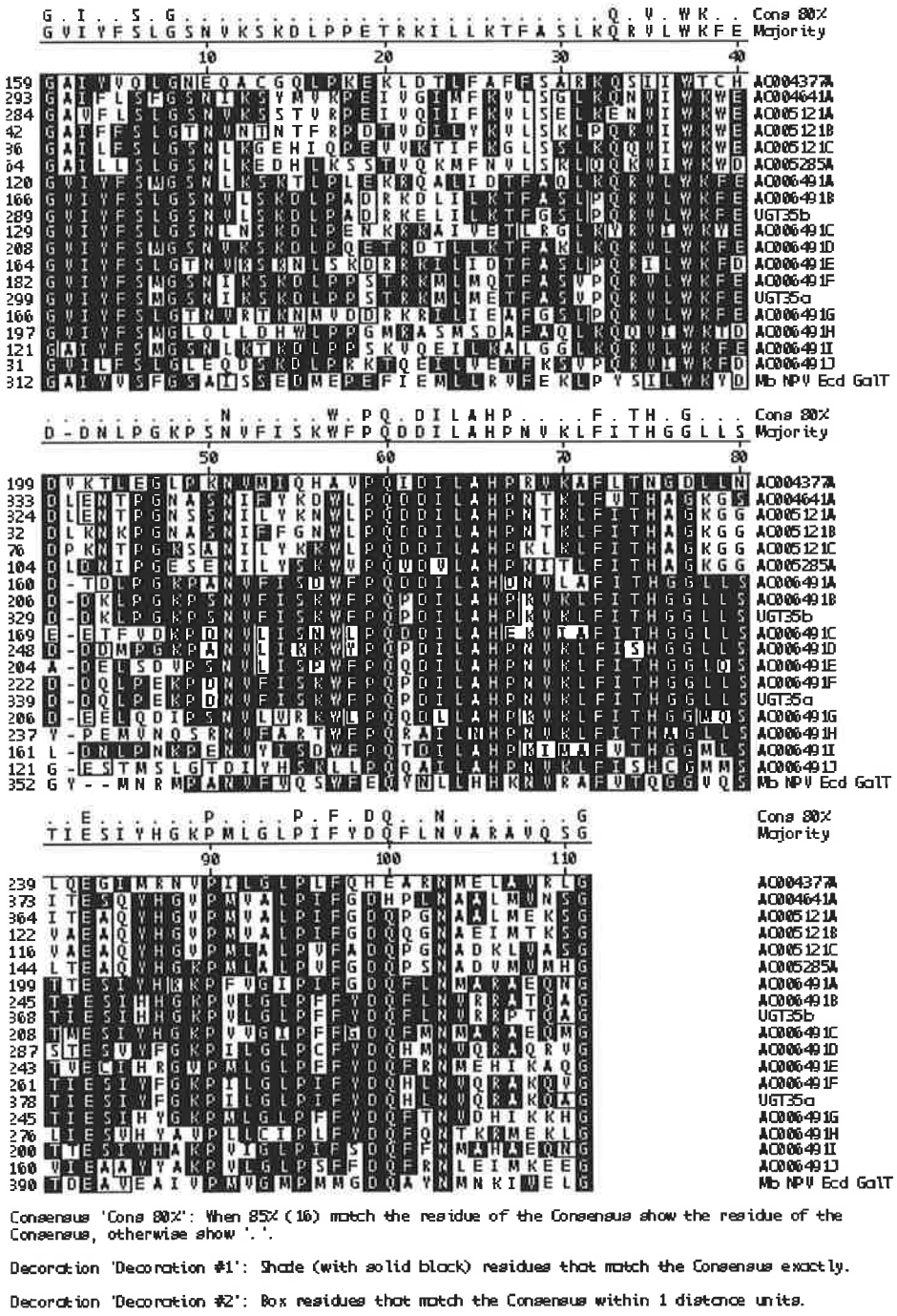

Fig. 2: Sequences of *Drosophila* ORFs that show homology with UDP-glucuronosyltransferases. The baculovirus ecdysone galactosetransferase is used as an insect member of the UGT family for comparison. UGT35a and UGT35b are two members of a family of 5 antennal-specific *Drosophila* UGTs that have been described recently (Wang *et al.,* 1999). They are almost identical to the ORFs above them (AC006491B and AC006491F) except for a few amino acid differences which might be due to polymorphisms. A consensus sequence, which is conserved amongst UGTs is present in all sequences of this alignment (Mackenzie *et al.,* 1997).

Table 1

Novel genomic fragments which contain ORFs with sequence homology to C-type lectins and their localisation in the genome.

SEQUENCE FRAGMENT	LOCALISATION ON THE GENOME	C-DNA OR EQUIVALENT GENOMIC SEQUENCE
Dgal-I		
AC004279A	24 D1- 24 D5	
AC004371A	24 D5- 24 D6	
AC004373A	24 F1- 24 F2	
AC004373B	24 F1- 24 F2	
AC004423A	29 C1- 29 C5	
AC004423B	29 C1- 29 C5	U 85758 (accessory gland protein)
AC004573A	21 C5- 21 D1	
AC004716A	22 C1- 22 C2	
AC004723A	28 C4- 28 D2	GH 24720
AC005149A	24 A1- 24 C5	
AC005889A	30 A3- 30 A6	
AC007082B	37 D	
AC007082C	37 D	
AC007083A	33 A	X94613*, AC007186, AC005116
AC007330A	46 C- 46 D	GH21870
AC007330B	46 C- 46 D	GH22861, CK 02422A: similar to Periplaneta lectin clot 8492 (incl. GH 10831) similar to Incilarin.

Note. In addition to the known galactose specific lectin and the furrowed gene product, six classes of cDNAs with homology to C-type lectins could be identified, some of which correspond to genomic sequences.

* Schmidt *et al.*, (1996)

Table 2

Chromosomal location of the genomic sequences from which the ORFs had been derived.

SEQUENCE FRAGMENT UGT	LOCALISATION ON THE GENOME
AC004641A	53 D1- 53 E2
AC005121A	36 B1- 36 B2
AC005121B	36 B1- 36 B2
AC005121C	36 B1- 36 B2
AC005285A	26 C1- 26 C2
AC006491A	86 D1- 86 E1
AC006491B	86 D1- 86 E1
AC006491C	86 D1- 86 E1
AC006491D	86 D1- 86 E1
AC006491E	86 D1- 86 E1
AC006491F	86 D1- 86 E1
AC006491G	86 D1- 86 E1
AC006491H	86 D1- 86 E1
AC006491I	86 D1- 86 E1
AC006491J	86 D1- 86 E1

An immune function for a glue-like *Drosophila* salivary protein

Introduction

Like vertebrates, insects use both soluble and cellular components to defend themselves against foreign intruders. Although cellular and humoral immune responses have traditionally been considered separately, most immune reactions *in vivo* entail elements from both branches of the immune system (Gillespie *et al.*, 1997; Elrod-Erickson *et al.*, 2000). An example of this is the coagulation of body fluids, which involves cellular and humoral procoagulant activity in both vertebrates and arthropods (Theopold *et al.*, 2002). Invertebrate hemolymph coagulation has been studied mostly in crustaceans (Hall *et al.*, 1999) and in the horseshoe crab (Iwanaga *et al.*, 1998). Although little is known about the coagulation of insect hemolymph, there is evidence for interaction between hemocytes and soluble proteins in several insect species (Theopold *et al.*, 2002; Bohn, 1986). In most insects, one of the key features of hemolymph clotting is the release of cellular material from hemocytes. This leads to the formation of an extracellular assemblage, which acts as the crystallization point for further crosslinking reactions, as well as a site where microbes may be immobilized and prevented from entering the body fluid (Gupta, 1984; Gupta, 1991). Biochemically, the material that is released from hemocytes is mostly composed of proteins with a high carbohydrate content and shows a positive reaction with polysaccharide specific stains and lectins (Gupta, 1991; Nappi and Silvers, 1984; Theopold and Schmidt, 1997; Hypsa and Grubhoffer, 1997).

Lectins had been previously used to identify and purify proteins from hemocytes or hemocyte-like cell lines in an attempt to identify coagulation components in insects (Theopold and Schmidt, 1997, Theopold *et al.*, 1996). Lectins with specificity for mucin-type glycodeterminants turned out to be particularly useful for this purpose. *Helix pomatia* lectin (HPL) is specific for N-acetylgalactosamine (GalNAc), the first monosaccharide added to proteins during O-linked glycosylation. One of the proteins identified using HPL is hemomucin, which was first isolated from *D. melanogaster*, and was subsequently found in other insects (Kinuthia *et al.*, 1999). In addition to hemocytes, hemomucin is expressed strongly in the gut and in follicle cells (Theopold *et al.*, 1996). In addition, three different hemomucin glycoforms were recently identified, one of which is strongly expressed in the salivary glands (Theopold *et al.*, 2001).

Peanut agglutinin (PNA) binds to core1, the determinant established by addition of galactose to GalNAc. There is little evidence for further extension of core1 in insects (März *et al.*, 1995). The attempts to isolate other O-glycosylated proteins from hemocytes using PNA led to the identification of a novel *Drosophila* hemocyte mucin. To my surprise, I found that this mucin is identical to a protein expressed in the salivary glands, where it is probably a component of salivary glue. Since dipteran salivary glands are related to silk glands of lepidopteran insects (both are labial glands), I decided to test whether silk proteins are expressed in both labial glands and immune tissues. Using sequence information available for two different lepidopteran fibroins (the main component of silk), I found both of them to be expressed in immune tissues, as shown by RT-PCR. This similarity across species leads to hypothesize a functional connection between the products of these different tissues.

Materials and Methods

Flies

Flies were kept on cornmeal/yeast food at 25°C with a 10/14 h light/dark cycle. *Ephestia kuehniella* larvae were fed on crushed oats and kept at the same light/dark cycle and temperature as the *Drosophila* cultures.

Hemocyte staining with lectin

Hemocytes were bled from individuals of the last larval instar. *Drosophila* larvae were torn to release a drop of hemolymph. *Ephestia* and *Galleria* caterpillars were bled by cutting off a proleg. In both cases, hemolymph was collected either into insect Ringer or into anticoagulant Ringer (Theopold and Schmidt, 1997), for Fig. 3 I-K, 10mM beta-mercaptoethanol was included). After allowing the cells to attach for a few minutes, FITC-conjugated PNA (50 µg/ml in anticoagulant Ringer including 0.05% Na-azide) was added. Intracellular staining was performed in the presence of 0.5% IGEPAL (Sigma). For testing bacterial binding to the clot, bacteria (*E. coli* Dh5a::gfplux) (Unge *et al.,* 1999) (kindly provided by J. Jansson, Södertörns Hökskolan) from an overnight culture were centrifuged, resuspended in an equal volume of anticoagulant Ringer and mixed with hemocytes on glass slides.

Electrophoretic techniques

SDS polyacrylamide-gel electrophoresis on a Mini-Protean II electrophoresis unit (Bio Rad) was performed essentially according to Laemmli (Laemmli, 1970).

Molecular weights were determined using prestained SeeBlue or Mark12 molecular weight markers (Novex). The proteins were blotted onto a nitrocellulose membrane (Amersham) as described by Theopold and Schmidt, 1997. Protein was loaded at ca. 5 µg/lane in Western blots or as indicated in the figure legends. The blotting efficiency was determined by staining the blot with Ponceau S. For Western-blots, peroxidase-conjugated PNA was used at a concentration of 1 µg/ml.

N-terminal sequencing of p150

Protein extracts were separated on a denaturing 6% SDS-polyacrylamide gels according to Laemmli, 1970.

Proteins were transferred onto a polyvinyldifluoride membrane (PVDF, Millipore) as described by (Sambrook *et al.*, 1989). The PVDF membrane was washed twice in water, stained with 0.025% Coomassie Blue R-250 in 40% MeOH, and destained with 50% MeOH. Bands were excised, rinsed once with 0.1% trifluoroacetic acid (TFA), and then rinsed extensively with water. Automated Edman degradation was carried out using an Applied Biosystems 494 Procise Protein Sequencing System.

P150 - *E. coli* binding

For each binding assay, 150 3rd instar *Drosophila* larvae were bled into 100 ml of Ringer (100mM Tricine - HCl pH 6.8, 100mM NaCl, 40mM KCl, 15mM MgCl and 4mM $CaCl_2$, saturated with PTU). Centrifugation was carried out at 7000 rpm for 10 min at room temperature. The plasma was kept on ice, whereas the hemocyte pellet was lysed for 10 minutes in insect Ringer containing (0.5% IGEPAL, 0.5% aprotinin, and 1% beta-amino-n-caproic acid). Centrifugation was done at 14000

rpm for 20 min at 4°C. The hemocyte lysate together with the plasma were added to pre-washed *E. coli,* and were incubated end over end for an hour. After centrifugation at 14.000 rpm (4°C) for 10 min, the supernatant was kept and the bacterial pellet washed 6 times with Ringer to remove unbound proteins. After solubilization in loading buffer, samples were separated on 6% SDS-PAGE, blotted, and analyzed using peroxidase-conjugated PNA (1μg/ml).

RNA extraction

Total RNA was extracted using an RNeasy RNA kit from Qiagen following the supplier's instructions. DNase treatment of the RNA was performed according to the Quiagen protocol.

***In situ* hybridizations**

In situ hybridizations were performed as described (Theopold *et al.*, 1995).

Isolation of *Ephestia* ESTs

ESTs with similarity to *Venturia canescens* virus-like particles (VLPs) were isolated from a lambda gt11 *Ephestia kuehniella* library using a VLP specific antiserum as described (Theopold *et al.*, 1994). After subcloning, the clones were sequenced and analyzed by comparison to GenBank sequences using BlastX. The EST sequences were deposited into the dbEST database of GenBank with the accession numbers BG695827-695838.

RT-PCR

For designing primers specific for I71-7, the annotation method was used (Wright *et al.*, 1996). There are some discrepancies between the genomic sequence published by the BDGP (predicted gene CG7604) on one hand and (Wright *et al.*, 1996) on the other hand, most likely due to the repetitive structure of the gene. Nevertheless, both sequences contain the protein sequence were obtained by sequencing p150. The annotation of (Wright *et al.*, 1996) was followed because their sequencing was performed using exonuclease-generated clones, which facilitates localization of repetitive sequences.

For RT-PCR, 1 µg total RNA was reverse transcribed with 1 µg oligo (dT) 15 primer in a 20 µl reaction containing 1mM of each dNTP, 40 U of RNase inhibitor (Promega), and 20 U of reverse transcriptase (AMV reverse transcriptase, Promega) in the reaction buffer supplied by the manufacturer. RNA was denatured at 65°C for 5 min and chilled on ice before adding the other reaction components. The reaction was carried out at 42°C for 60 min, then heat inactivated at 95°C for 5 min The reaction was diluted to 50 µl and stored at -80°C until needed. Control reactions, without the addition of reverse transcriptase, were included in every experiment in order to discount the possibility of DNA contamination.

Relative quantitative PCR

PCR was used to determine the accumulation of mRNA transcripts. Primers for ubiquitin and RP49 were used to ensure that different PCRs contained equal amounts of cDNA template. All PCRs were carried out in a PC-960 microplate

Thermal Sequencer System (Corbette Research). Primers used for PCR amplification were as follows:

Ephestia fibroin: GTTGATTCCGTTTAGTCCAACC and

TTCGTCATCTTCAGCAGCTTC

Galleria fibroin: GATCTTGTGCTGTGCTCTGCA and

GCACTAAAGTCCCGTTCTCAA

Galleria calreticulin: CTTGCACTTAGAACGTTCCTACG and

CTCATGAAGATGCTGACTGTTCC

Ephestia ubiquitin: GCCGTTTGAATATTGAAGTCG and

TCAACAACCTGGGCAACTC

Drosophila RP49: CTTCATCCGCQCCAGTC and TCTCCTTGCGCTTCTTGG

Drosophila I71.7: GAGACACCAAGACCAATACATC and

GAGTAGTGCTTTCAGTACAAG

Drosophila SGS5: CCAGGAGACGAAAATCGAAG and

GGAATGGGGAATTCAGCAAC.

Reactions were set up in a 20 µl volume containing 2 µl 10 x Taq polymerase buffer (supplied with enzyme), 0.5 µM each of two specific primers, 200 µM of each dNTP, 2mM MgCl2 and 0.5 U Taq DNA polymerase (5 U/µl; Promega). Cycling conditions were: 5 min 94°C, followed by cycles of 94°C for 1 min/ 50°C for 1 min / 72°C for 1 min The cycle numbers are shown in the figure legends. PCR products were analyzed by agarose gel electrophoresis with 0.2 µg/ml ethidium bromide in both gel and buffer.

After analyzing PCR products on gels, reactions were repeated with adjusted amounts of template until the amplification products for the control (Rp49 and

ubiquitin) primers were equal in all samples. These adjusted amounts were subsequently used in all other PCRs.

Results

Glycoproteins are involved in coagulation of insect hemolymph, although the clotting process is not completely understood (Theopold *et al.*, 2002). The lepidopteran *Galleria* was one of the first model insects used to study this process (Ratcliffe and Rowley, 1979). Hemolymph bled from larvae will coagulate on glass slides (Gregoire, 1974), where strands of extracellular material are stained by PNA (Fig. 3 A). Following the classification of insect clot formation proposed by Gregoire (Gregoire, 1974), *Galleria* clots belong to class III; showing a mixture of fibrils and dot-like structures, although the fibrils are more predominant. When larvae were bled into a drop of bacterial suspension, bacteria were found closely associated with the strands and surfaces of hemocytes, and co-localized with PNA staining (Fig. 3 A-C). This suggested that PNA-reactive proteins might be important for microbial immobilization as well as clotting. In a second lepidopteran (*Ephestia kuehniella*), a pattern with prevalent dot-like structures was observed which also bound bacteria (Fig. 3 D and E). *Ephestia* clots are also classified as belonging to class III, but fibrillar components are less abundant than in *Galleria*. To have more molecular genetic tools available, I switched to *Drosophila* as a model system. *Drosophila* clots belong to Gregoire's class II, showing predominantly membranous structures. When *Drosophila* larvae were bled into a drop of bacterial suspension, the bacteria were seen associated with the PNA-staining extracellular material (Fig. 3 F-H), broadly resembling structures observed in *Galleria* and *Ephestia*. Under nonactivating conditions, the stained material is

seen in intracellular vesicles of different sizes (Fig. 3 I and K). Under the same conditions, no staining was observed in preparations of non-permeabilized cells (not shown).

In protein extracts from *Drosophila* hemocytes, PNA staining was concentrated in a major protein band of 150kDa (p150, Fig. 4 A). The same 150kDa band was detected in a gut lysate, and a strongly stained band was also seen in salivary glands. To investigate interactions between p150 and bacteria, bacterial binding experiments were performed using whole hemolymph lysates. A substantial fraction of p150 bound to bacteria (Fig. 4 B). This binding was shown to depend on both Ca^{2+} and sugar (Fig. 4 C). Inhibition was observed with galactose-containing sugars, but not with other sugars (Fig. 4 C) indicating that the PNA-binding carbohydrate determinant in p150 is necessary for binding bacteria. Thus these data point toward the involvement of a C-type lectin-like activity with specificity for galactose in p150 binding to bacteria. In contrast, two other mucins identified in the hemolymph lysate did not bind to bacteria (Fig. 4 C, lower part). Since one of the mucins (100kDa molecular mass) that did not bind bacteria is almost certainly hemomucin, binding to bacteria is not a general property of hemolymph mucins.

p150 is I71-7

This suggested that p150 has unique properties by forming aggregates with bacteria. This justified the molecular identification of the protein. I took advantage of the fact that p150 is strongly expressed in salivary glands to isolate sufficient amounts of material for protein sequencing from this tissue. The N-terminal protein sequence obtained was unique and identified p150 as I71-7 (Fig. 5 A). This is one of the proteins encoded by a locus located in the 71E late puff whose expression is

regulated by ecdysone (Wright *et al.*, 1996). In contrast to other members of this locus, which are late effector genes of the ecdysone-regulated cascade, I71-7 belongs to the intermolt genes, which are expressed earlier during metamorphosis. I71-7 shows similarity in sequence, amino-acid composition, and in its repetitive domain structure to *Drosophila* glue proteins, which are encoded by a set of intermolt genes and are necessary to attach the pupa to a dry substrate (Wright *et al.*, 1996). The predicted protein sequence identified I71-7 as a possible mucin, with 68 potential O-glycosylation (Hansen, 1998) and 10 potential N-glycosylation sites, numerous cysteines, and a repetitive structure, which is typical for many mucins (Fig. 5). The predicted molecular mass for the protein from the coding region is 42,950Da, indicating extensive posttranslational modifications to produce the mature protein of 150kDa. RT-PCR confirmed that I71-7 is expressed in salivary glands and hemocytes, just as PNA staining suggested for p150 (Fig. 5 B). The lack of signal in other tissues shows that I71-7 is not ubiquitously expressed. A second glue protein (SGS-5) was only expressed in the salivary glands, a pattern expected for a typical glue protein. It can be concluded that p150 is I71-7, and that this protein is expressed in both hemocytes and salivary (labial) glands. This pattern of expression, while surprising, is not unique, as hemomucin is also expressed in both *Drosophila* hemocytes and salivary glands (Theopold *et al.*, 1996; Theopold *et al.*, 2001).

To test whether immune proteins have been co-opted by the labial glands in other insects, the expression of one of the major products was examined (the fibroin heavy chain, F-hc) of labial (silk) glands in the two lepidopterans *Galleria* and *Ephestia*. The *Galleria* fibroin heavy chain sequence was known (Zurovec and

Sehnal, 2002), while in this laboratory several ESTs coding for the *Ephestia* fibroin heavy chain had been previously isolated (see material and methods).

RT-PCR results show that both *Galleria* and *Ephestia* F-hc are expressed in immune tissues in addition to their expression in silk glands (Fig. 6 A and B). While *Ephestia* F-hc is expressed in hemocytes (Fig. 6 C) similar to I71-7 and hemomucin in *Drosophila*, *Galleria* F-hc was not expressed in hemocytes, but in the fat body (Fig. 6 D), which is also an immune tissue.

Discussion

This chapter presents evidence for an immune function for some labial gland proteins. Previous observations suggested a dual expression of several proteins in both hemocytes and labial glands. One of these proteins is hemomucin, a cell surface mucin, which was first isolated from a *Drosophila* hemocyte cell-line and later found to be strongly expressed in salivary glands (Theopold *et al.*, 2001). In a screen for hemocyte proteins, several ESTs coding for fibroin heavy chain had previously been isolated and shown to be one of the major products in the silk glands from *Ephestia* (unpublished data).

Here it is shown show dual expression in labial glands and hemocytes for a secretory *Drosophila* mucin, which is released from hemocytes upon bleeding and involved in subsequent clot formation. The novel mucin turned out to be identical to a previously described salivary gland protein (I71-7). The fact that glue proteins are also involved in immune reactions could constitute an interesting paradigm. Either the salivary gland proteins where co-opted by the immune system or, alternatively, the glue proteins evolved from immune proteins. Although it is difficult to decided at this stage, which of the two scenarios is correct, the fact that

the immune system precedes salivary gland functions, it is likely that the putative immune function of 171-7 is ancestral to the labial gland function. This led to revisit previous work on silk glands and confirm at a molecular level that fibroin heavy chain is expressed in both the labial glands and in immune tissues in *Ephestia* and *Galleria*.

In both *Drosophila* and *Galleria* I was able to observe entrapment of bacteria in the clot, which could be stained in both insects using PNA. The specific involvement of 171-7 was shown using bacterial binding assays (Fig. 4). In these assays, the interaction between bacteria and 171-7 was inhibited in the absence of Ca^{2+} and in the presence of sugars, which mimic the determinant recognized by PNA. Two other mucins present in hemolymph did not bind to bacteria. These data indicate that both carbohydrate determinants on 171-7 and a (most likely humoral) C-type lectin activity may participate in crosslinking bacteria to 171-7. Amongst the C-type lectins identified in the *Drosophila* genome, there are in fact several interesting candidates for clotting factors including hemolectin (Goto *et al.*, 2001), which shows domains with similarity to von Willebrand factor and a humoral lectin, which is in fact specific for galactose (Haq *et al.*, 1996). In lepidopteran hemolymph plasma PNA-like lectin-isomers are the most abundant lectins. Attaching bacteria to the clot prevents their dissemination into the circulation and may be important for immune defense.

While I was able to confirm that both *Ephestia* and *Galleria* fibroin proteins are expressed in immune tissues, differences should be noted between these two species. In *Ephestia* fibroin is expressed in labial glands and hemocytes (Gillespie *et al.*, 1997), like 171-7 in *Drosophila*, while in *Galleria*, fibroin is expressed by the fat body instead (Elrod-Erickson *et al.*, 2000). While both of these are fibroin

proteins, they show limited homology (only *Ephestia* fibroin contains short stretches of collagen-like repeats). Molecular differences are even more apparent when both fibroins are compared with I71-7. These differences in structure and expression might partly account for the morphological differences Gregoire described for clots (Gregoire, 1974) and are likely to produce matrices with different physical properties affecting adhesiveness, elasticity and tensile strength.

The biochemical basis for the dual usage of proteins in labial gland secretions and during hemolymph clotting lies in their tendency to easily precipitate from an aqueous solution and establish an extracellular assemblage. The same argument applies for the gut and for the follicle cells in the ovary, where components of extracellular matrix (the peritrophic membrane and the chorion respectively) have to harden quickly and where factors implied in immunity and coagulation are also detected (Theopold *et al.*, 1996).

In conclusion, these findings indicate that labial gland products are used in other insect tissues, where large assemblages, such as clots, glue or extracellular matrix has to be established quickly. This is particularly urgent during hemolymph coagulation both to avoid loss of fluid and prevent infections. In this context, it is natural to expect that hemolymph coagulation leads to entrapment of bacteria (see Fig. 3). I believe that this process relies on a number of components, which - like p150 - are expressed constitutively. In addition, tethering bacteria to the clot is expected to render them more accessible to other effector molecules of the immune system, including antimicrobial substances and phenoloxidase, which are activated at different levels. With a better molecular understanding of coagulation, it will be possible to characterize the specificity of the interaction between microorganisms

and clot components and to further our understanding of the interaction between hemolymph clotting and other immune responses.

Figures

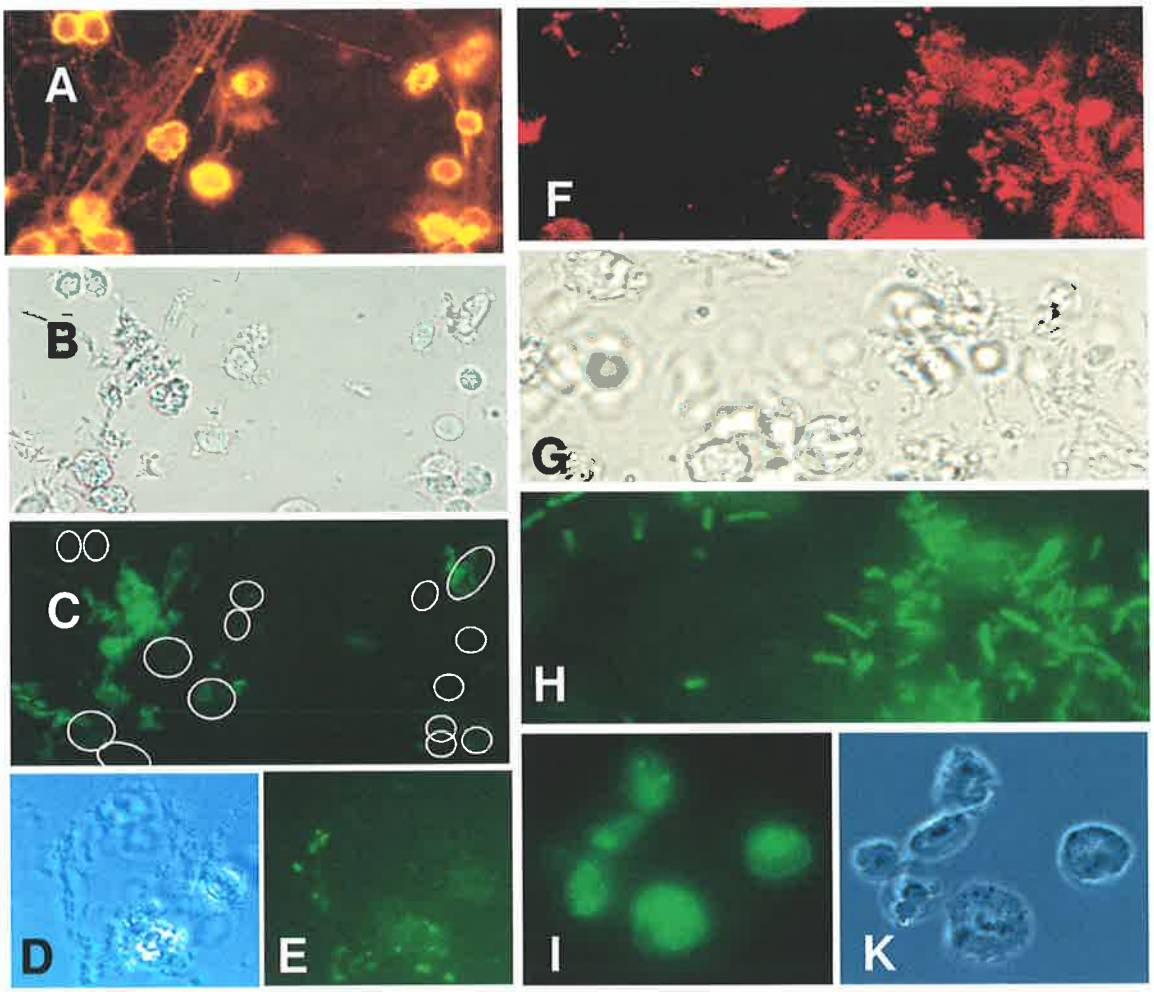

Fig. 3: Hemolymph clot formation in two lepidopteran species and *Drosophila* and entrapment of bacteria in the clot. Hemolymph from *Galleria mellonella* (A-C) *Ephestia kuehniella* (D and E), and *Drosophila melanogaster* (F-H) was bled onto a drop containing GFP-labeled gram-negative bacteria showing entrapment of bacteria in the fibrillar clot. Bacteria are visualized by green fluorescence in C, E, and H. The *Galleria* and *Drosophila* clots were visualized by labeling with TRITC-conjugated PNA in A and F. In I and K, *Drosophila* hemocytes were bled into anticoagulant Ringer and permeabilized to detect intracellular PNA-reactive material.

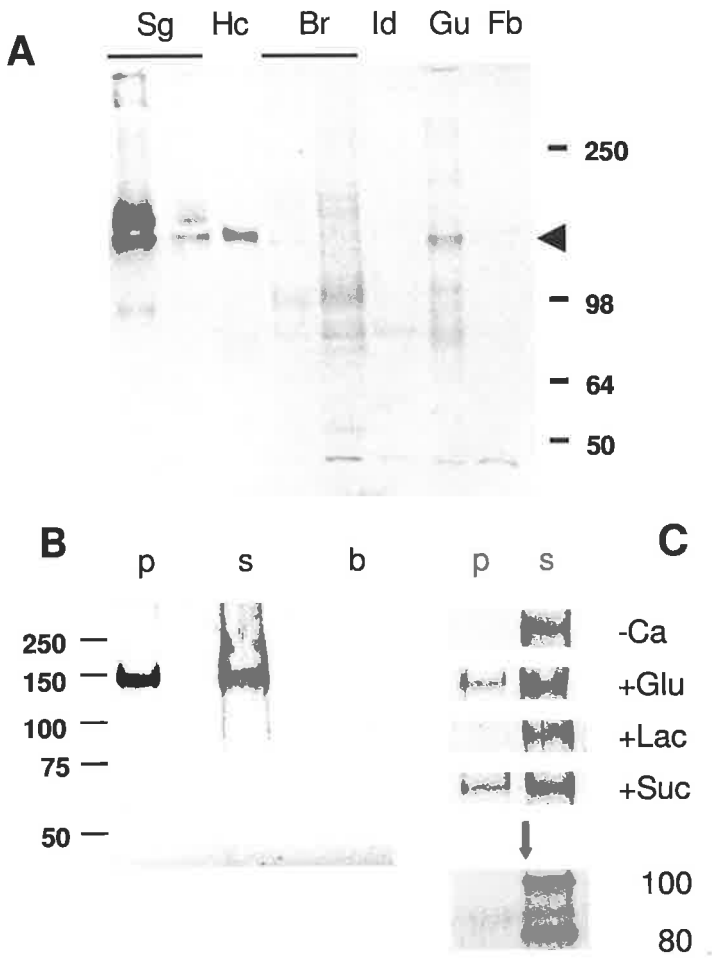

Fig. 4: p150 from hemocytes binds to bacteria. Lysates from different larval tissues were analyzed in a Western-blot using peroxidase-conjugated PNA (A). The amount loaded corresponds to 1 animal equivalent (AE) except for the Hc sample (10 AEs), for the silk gland sample (0.1 and 0.05AEs), and for the brain sample (1 and 2 AEs), (Sg=salivary gland, Hc=hemocytes, Br=brain, Id=imaginal discs, Gu=gut, Fb=fat body). B-C: bacteria were incubated with a hemolymph lysate washed extensively to remove unbound material and the bacterial pellet (p) as well as the supernatant (s) containing unbound material analyzed on a Western-blot developed with labeled PNA (B-C) or HPL (C, lower part). Bacteria alone were included as a control (b). In C, the binding was performed in Ca^{2+}-free buffer (-Ca) or in the presence of different sugars (glucose, lactose and sucrose, each at a conc. of 200 mMol). The blot from the incubation with sucrose was re-incubated with HPL to detect two additional hemolymph mucins of masses 100 and 80 kDa, which did not bind to bacteria. Incubation with HPL was performed for all blots were p150 binding to bacteria had been observed and in no case was any other mucin found binding to bacteria.

A

MKLTVVCLVVSFFLLHYAEHSDAC (1-24)

LEVIEKALGLQPCNEGGRNEHREPHRGGPGPVRSTRRRGRIPRRRETPRPIHHNTRERRHHTKTRKPRKPVPCITKR (25-101)

TEPPPVTDFTTRKSNPPCTCTESTTRKTNPTCTCTESTTKKTNPTCTCTESTTKKTNPTCTCTESTTPL (102-171)

TEPP VTDITTQKSNPPCTCTESTTRKTNPTCTCTESTTQKTNPTCTCTESTTKKTNPTCTCTESTTPL (172-240)

TEPP VTDITTQKSNPPCTCTESTTRKTNPTCTCTESTTRKTNPTCTCTESTTKKTNPTCTCTESTTPL (241-309)

TEPPVTDITTQKSNPPCTCTESTTQKIKSTSTTQG (309-342)

TEPPSTQKTLPPNPPSTKN (343-363)

TEPPNSTPPPEKTTRKPCGCSSSHPSGWNLAVL (364-393)

B

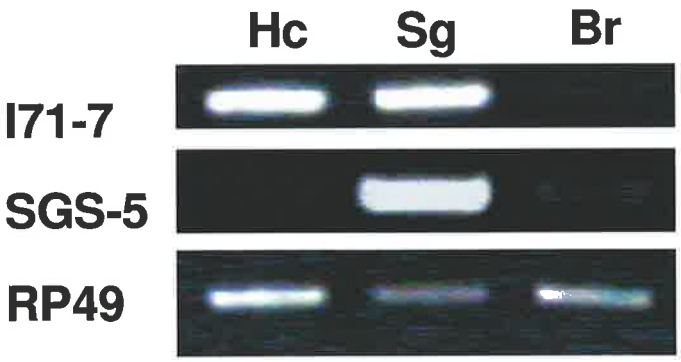

Fig. 5: p150 is I71-7. A: amino acid sequence of I71-7. The aminoterminal protein sequence obtained from p150 is underlined. An arg-rich domain and two types of triple-repeats are shown. B: RT-PCR detecting expression of I71-7 an unrelated second glue protein (SGS-5) and RP 49 as a loading control in RNA samples from hemocytes (Hc), salivary glands (Sg) and the brain (Br). For all samples, 30 PCR cycles were run.

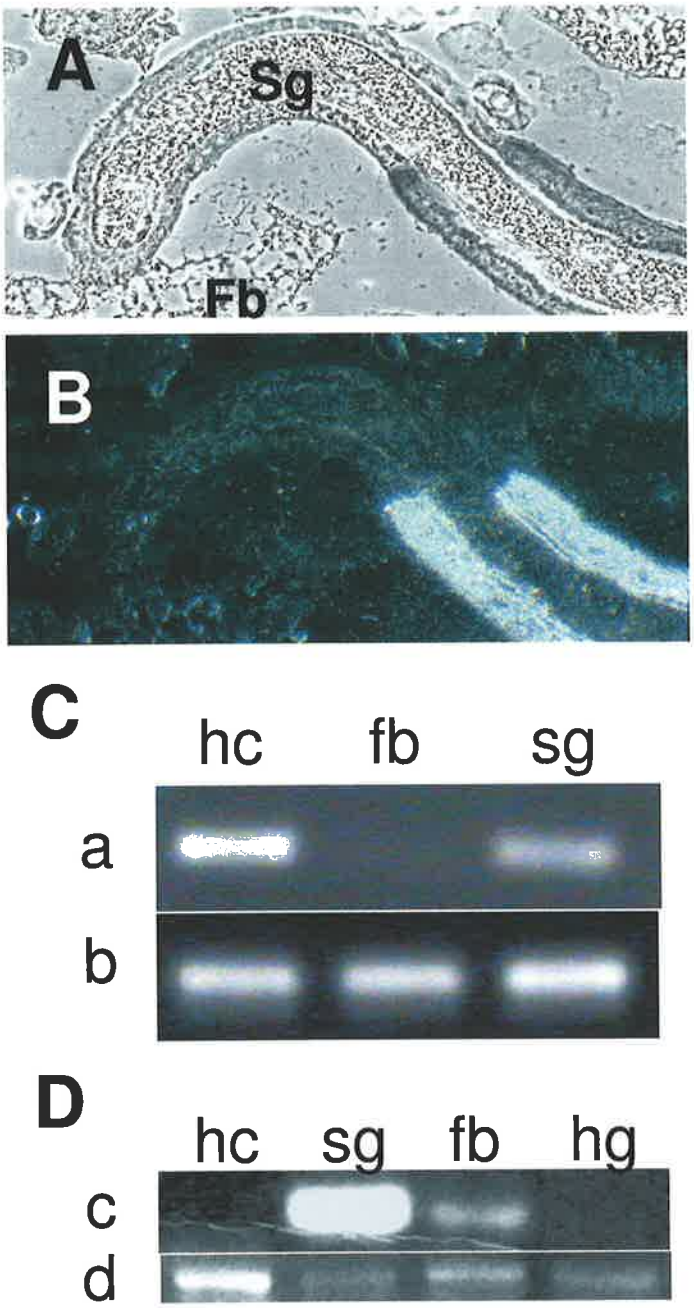

Fig. 6: Fibroinheavy chain from *Ephestia* and *Galleria* are expressed in immune tissues. A and B: Phase contrast (A) and *in situ* hybridization (B) using the insert from an EST coding for the *Ephestia* fibroin heavy chain. Strong hybridization was found to the posterior part of the silk glands (Sg, fat body=Fb). C and D: expression of the *Ephestia* and *Galleria* fibroin heavy chain was analyzed in the hindgut (hg), silk gland (sg), fat body (fb) and hemocytes (hc) using RT-PCR with primers for *Ephestia* fibroin (a: 35 PCR cycles) and ubiquitin (b: 40 cycles) as well as *Galleria* fibroin (c: 35 cycles), and calreticulin (accession No.AB063250: d: 27 cycles).

Animal and plant members of a gene family with similarity to alkaloid-synthesizing enzymes

Introduction

Amongst plant alkaloids, some monoterpenoid indole alkaloids have attracted particular interest due to their ability to act as toxins (strichnin), antimalarial (quinine), antineoplastic (vincristine and vinblastine) (Kutchan, 1995) or antipsychotic drugs (reserpine) (Bennet, 1998). This led to the identification of a number of core enzymes that are an essential part of their biosynthetic pathway, including tryptophane decarboxylase and strictosidine synthase (SS), two key enzymes in the production of monoterpenoid indole alkaloids (Kutchan *et al.*, 1988). SS has been isolated from two plant species, which are known for their pharmaceutical benefits, namely *Catharanthus roseus* (madagascar periwinkle) (McKnight *et al.*, 1990) and *Rauvolfia serpentina* (sarpagandha plant) (Kutchan *et al.*, 1988).

Members of this laboratory have previously isolated a cell-surface molecule (hemomucin) from a hemocyte-like *D. melanogaster* cell line (Theopold *et al.*, 1996). Hemomucin is composed of two domains, one with mucin-type repeats and the other with sequence similarity to SS. In addition, evidence for the existence of hemomucin in two other insect species has been reported the hymenopteran *Venturia canescens* (Kinuthia *et al.*, 1999) and the lepidopteran *G. mellonella* (Theopold and Schmidt, 1997). Using expression data and information obtained from genomic sequencing and EST projects, I identified a novel gene family with similarity to SS and hemomucin.

Materials and methods

Sequence smilarity searches

Sequences similarity searches were performed on the NCBI server or the BDGP server using the Blast algorithms (Altschul *et al.*, 1990). Sequences with with a significant similarity to the original sequence were downloaded into the Lasergene program package (DNASTAR Inc., Madison, WI) and further aligned using Megalign. Genomic regions were analysed using Genfinder. Chromosomal localisations for P1 clones and BACs were available from the BDGP/HHMI EST project (Hartl *et al.*, 1994; Kimmerly *et al.*,1996).

Insect cultures

D. melanogaster w118 flies were kept on cornmeal/yeast food at 25°C with a 10/14 h light/dark cycle.

Preparation of antisera

For the production of an antiserum against recombinant hemomucin a PCR amplified fragment covering amino acids 178-299 as in Theopold *et al.*,1996 or position 216-341 in Fig. 7 was expressed in the expression vector pQE32 (Qiagen). The resulting fusion protein was purified according to the instructions of the manufacturer and excised from a preparative polyacrylamide gel. Rabbits were immunized according to Harlow and Lane (Harlow and Lane, 1988) with approximately 10 µg protein/immunization.

Immunoblotting and Immunodetection of Proteins

The transfer of the separated proteins onto nitrocellulose filter was carried out in a Mini Trans-Blot cell (Bio-Rad) applying 200 mA / 70 V for about 1 h using transfer buffer containing 192mM glycine, 25mM Tris-base and 20% (v/v) methanol (Towbin *et al.*, 1979). After preincubation of the filters in first blocking solution (8% (w/v) non-fat milk powder and 0.02% sodium azide in PBS) for 1 h at room temperature with gentle agitation, the primary antiserum was added to a specific final concentration, and the incubation was continued for another 2 h. Then, the filters were washed three times, 10 min each, with gentle agitation: twice in PBS and once in 150mM NaCl, 50mM Tris-HCl pH 7.5.

Alkaline phosphatase-conjugated secondary anti-rabbit antibody (Pierce) diluted 1:10000 in second blocking solution (5% (w/v) non-fat milk powder, 0.02% sodium azide, 150mM NaCl, 50mM Tris-HCl pH 7.5) was added to the blot. The blots were incubated for 2 h at room temperature. After the incubation in secondary antiserum, alkaline phosphatase-conjugated anti-rabbit IgG (Pierce) in a 1:10000 dilution, the filters were washed twice for 10 min each in 150mM NaCl, 50mM Tris-HCl pH 7.5 with moderate shaking. Finally, for the visualisation of the protein bands, the blots were stained with NBT and BCIP as described by Sambrook *et al.* (1989).

Radiolabelling and Purification of DNA Probes

Ready-to-Go DNA labelling beads (Amrad Pharmacia Biotech, Boronia, Vic, Australia) were used according to the manufacturers instructions to radioactively label DNA probes. Briefly, the to be labeled DNA probe was first denatured and

then mixed with oligodeoxyribonucleotides of random sequence. These "random oligomers" anneal to random sites on the DNA and then serve as primers for DNA synthesis by a DNA polymerase. With [a-32^P]dCTP present during this synthesis, highly labeled DNA is generated.

Subsequently, to quantitatively remove the unincorporated [a-32^P]dCTPs from the DNA labelling reaction, ProbeQuant G-50 micro columns (Amrad Pharmacia Biotech, Boronia, Vic, Australia) were applied according to the supplier's instructions.

Hybridization conditions

The membrane was prehybridized for 1 h at 42°C with 15 ml hybridization buffer (5 x SSC; 50% formamide; 5 x Denhardt's-solution; 1% SDS; 100 µg/ml salmon sperm DNA). Meanwhile the radioactive labeled probe was heat for 3 min at 95°C and then cooled on ice.

The prehybridization buffer was discarded and the membrane incubated overnight at 42°C with fresh hybridization buffer and the probe. The hybridization buffer was discarded and the membrane washed twice with 2 x SSC, 0.1% SDS at 65°C and once with 0.1 x SSC, 0.1 x SDS at 42°C.

The wet membrane was exposed overnight at-80°C to X-ray films.

Northern blots

The multiple-tissue Northern blot was purchased from Invitrogen and used for hybridization experiments as described above.

Results

Novel strictosidine synthase and hemomucin

In an attempt to identify genes that are related to hemomucin, the hemomucin amino acid sequence was used to run comparisons with genome sequences obtained from sequencing projects of the major model organisms, including *D. melanogaster*, *Arabidopsis thaliana, Caenorhabditis elegans*, *Homo sapiens* and *Mus musculus*. In order to reduce background due to the mucin domain, which shows sequence similarity with a large number of glycoproteins, only the domain with similarity to SS was used for comparison. Interestingly, one additional ORF with similarity to SS could be identified in *Drosophila* (Fig. 7). Further inspection of the genomic region comprising this sequence revealed no ORFs with mucin-type character. Using Genefinder, a complete ORF could be assembled from the genomic sequence, which also did not contain any mucin-like sequences. The novel gene is localised to the same chromosomal position as hemomucin (98F). Because of the sequence similarity to SS and the absence of a mucin domain, the gene coding for the second member of this *D. melanogaster* family was named strictosidine synthase-like 2 (DmSSl2). The ORF predicts a product of approximately 46kDa molecular mass. Since the sequence conservation between hemomucin and DmSSl2 is strong enough to expect antibody cross-reactions, Western-blot analyses using an antiserum produced against recombinant hemomucin was performed. Since structural prediction programs predicted DmSSl2 to be secreted (Nielsen *et al.*, 1999; Nakai and Horton, 1999), hemolymph samples for hemomucin-like proteins were analysed. A labeled band of the predicted molecular mass (and different from the 100kDa of hemomucin) was detected, in a

blot developed with the pre-immune serum (Fig. 8) was completely absent. It can be argued that DmSSl2 is antigenically related to hemomucin and present in *Drosophila* hemolymph. It does not contain the mucin domain and differs in subcellular localisation from hemomucin, which is a cell surface protein. DmSSl2 also lacks two of the three N-linked glycosylation sites present in hemomucin and is expected to be less glycosylated or not glycosylated at all.

In order to identify members of the SS family from other organisms, sequence searches were performed on several organisms with genomic and/or EST sequencing projects. In *A. thaliana*, 13 members could be identified, two with the highest similarity to SS from *Catharanthus roseus* (AtSSl11 and AtSSl12) and other less related members (Fig. 9 A and B). Amongst these, four sequences seem more closely related to hemomucin than to SS (AtSSl4-7). Three sequences, derived from soybean (*Glycine max*) ESTs, were long enough to allow an alignment and were included in the analysis. Additional members were identified from *C. elegans* and both mouse and human. None of the newly identified members contain a mucin domain as with hemomucin, which seems to be unique to insects. Since SS has been grouped together with paraoxonases and gluconolactonases into a superfamily based on sequence similarity, human paraoxonase was also included in the alignment (Kobayashi *et al.*, 1998). In a phylogenetic analysis, all SS-like sequences are more related to each other than to paraoxonase (Fig. 9 B). The majority of the plant sequences group together and are separate from the animal sequences. Sequences AtSSl4-7 from *A. thaliana*, and one of the soybean sequences (GmSSl3) are an exception as they show higher similarity to the animal members than to other plant members of the family even from *A. thaliana*.

Since the EST sequencing projects are performed on tissue-specific libraries, a preliminary indication of the specificity of expression can be obtained from the number of ESTs obtained from one particular tissue. Judged from this distribution, high expression for the human SS-like EST is expected in the brain (a majority of the ESTs in the clot with hemomucin similarity are from the brain, see: http://www.ncbi.nlm.nih.gov/UniGene/clust.cgi?ORG=Hs&CID=22391). In order to confirm high expression in the brain, a human multiple tissue Northern blot was probed with one of the ESTs (clone image 258726), confirming that the strongest expression amongst the tissues chosen is in fact in the brain (Fig. 10). The chromosomal localisation of the human SSL-sequence is to chromosome 20 between regions 11:21 and 11:23.

Discussion

Evidence for the existence of a novel gene family with similarity to the plant enzyme strictosidine synthase, one of the key enzymes in plant alkaloid biosynthesis is shown using bioinformatics approaches. Since secondary metabolites like indole alkaloids are widely distributed in plants but less often found in animals, the existence of members of this family in a number of animal species came as a surprise. It could be speculated that the animal members of the family have a different function than SS. The fact that the *Arabidopsis* members fall in two groups, with one group being more similar to hemomucin than other plant genes, indicates that , SSl proteins may also perform more than one function in plants and are involved in more than one biochemical pathway. The existence of at least 13 members of the SSl family in *Arabidopsis* is in agreement with the identification of multiple isoforms of the enzyme in *C. roseus* (Dewaal *et al.*, 1995)

together with the observation that this plant is known to contain over 100 different monoterpenoid indole alkaloids (Kutchan, 1995). Each of these SSl family members may differ in their enzymatic activities and/or substrate specificities for the intermediate products of secondary metabolism. In addition, some genes might differ in their expression and/or tissue-specificity some of which were shown to be induced by fungal elicitors and jasmonate in the case of SS from *C. roseus*. The selective pressure for the production of a variety of phytoalexins may have lead to the duplication of an ancestral SSl gene in Arabidopsis and other plants (like soybean). In *Arabidopsis*, multiple gene duplication events have lead to the 13 observed members of the family (for example AtSSl8 and 9 and AtSSl4-7 are closely linked and have arisen by gene duplication).

The gene coding for the next enzyme in the biosynthetic pathway leading to indole alkaloids, strictosidine glucosidase (Geerlings *et al.*, 2000) also seems to be part of a multigene family in *Arabidopsis*. The most intensely studied members of this family are myrosinases, which are believed to be part of the plant's defense against insects and possibly pathogens (Rask *et al.*, 2000). In contrast to that and similar to the situation with SS, only one gene with significant similarity to strictosidine glucosidase in *Drosophila* could be identified (unpublished results). Strictosidine itself and the deglucosylation product of strictosidine could be shown to have antibacterial activity but seem to lack the antifeedant activity of further downstream products in the biosynthetic pathway for indole alkaloids (Luijendijk *et al.*, 1996). As for insects, hemomucin, the previously identified member from *Drosophila* seems to be exceptional in that it contains a mucin-like domain attached to the SS-like domain. Therefore in *Drosophila* and possibly in other insects, there seem to exist two members of the family, one on the cell-surface and a second one which is

present in hemolymph. At this stage, it can only be speculated what the substrates for SSl members in insects might be, but it is interesting to note that a number of components of the phenoloxidase cascade (like dopamine) show structural similarity with indole ring-containing substances. The enzyme upstream of SS in the biosynthetic pathway for monoterpene indole alkaloids (tryptophane decarboxylase) shows high sequence similarity with the enzymes dopa decarboxylase and the product of the *l(2)amd* gene, which are known to be involved in both the generation of neurotransmitter substances (like dopamine and serotonin) and central components of the phenoloxidase cascade (Wright, 1996; Wang and Marsh, 1995). The *amd* product is also expressed in the lymph glands suggesting a possible role in immunity (Wang and Marsh, 1995). In agreement with a possible function of hemomucin in defense reactions, which may include antimicrobial activity and/or protein crosslinking, it was previously shown to be involved in hemolymph coagulation (Theopold and Schmidt, 1997).

An interesting finding is the fact that the human member of the SSl family shows the strongest expression in the brain, the very organ some indole alkaloids like reserpine act upon by interfering with the synaptic transmission mediated by dopamine (Bennet, 1998). Learning about the function of the human SSl protein(s) might therefore help to understand the effect of these alkaloids in the brain. It will also be interesting to analyze a possible function of the insect genes in the nervous system and to look into a possible correlation between neurological defects in the genomic and mutations in the corresponding gene. With the availability of the sequence information presented here it is now possible to address these questions.

Figures

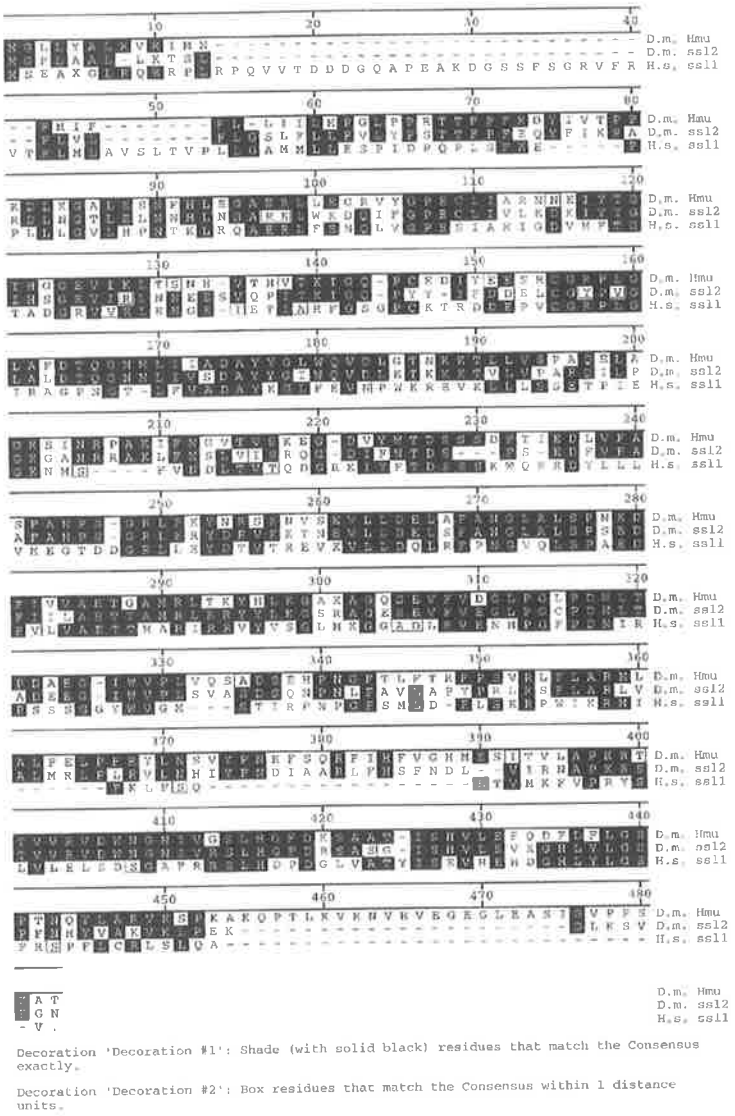

Fig. 7: Sequence alignment of 3 members of the SSl family. *Drosophila* (D. m.) and human (H.s.), members of the SS family were aligned using the Megalign program.

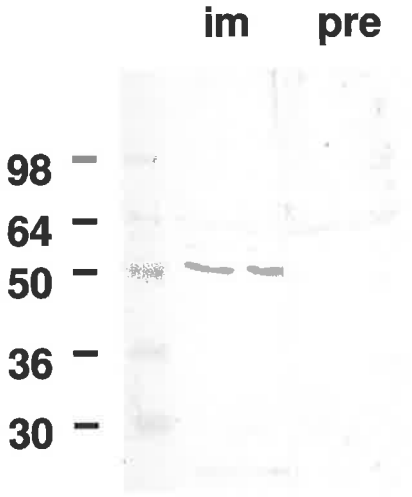

Fig. 8: A candidate for DmSSl2 in *Drosophila* hemolymph. Two different concentrations of *Drosophila* hemolymph (equivalent to the hemolymph of 40 and 20 3rd instar larvae) were analysed with an antiserum specific for hemomucin and the pre-immune serum as a control. A band of the size expected for DmSSl2 was only detected with the immune-serum.

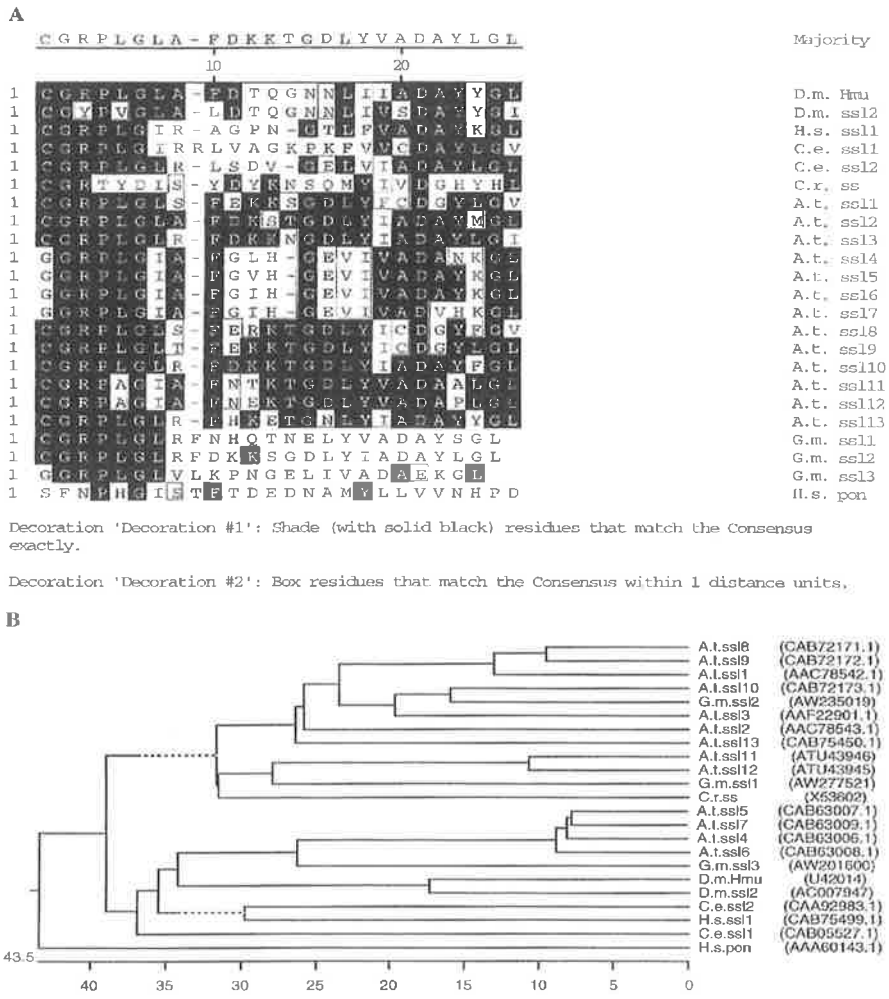

Fig. 9: Relationship between members of the SSl family. The same sequences as in Fig. 7 and sequences from *C.elegans* (C.s.), *C.roseus*, (C.r.) strictosidine synthase as well as *Arabidopsis* sequences were analysed for their possible phylogenetic relationship (unbalanced display, where branch distances correspond to sequence divergence) using the CLUSTAL program as part of Megalign. One of the conserved parts (position 155-179 in Fig. 7) is shown in (A). Some of the sequences are identical in this part but differ otherwise. The phylogenetic tree in (B) is based on the sequence between position 116-281 in Fig. 7.

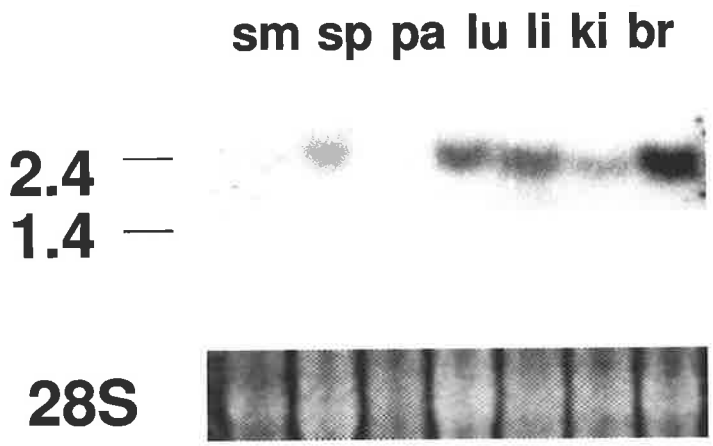

Fig. 10: Expression pattern of the human SSl member. Total RNA from brain (br), kidney (ki), liver (li), lung (lu), pancreas (pa), spleen (sp) and skeletal muscle (sm) were analysed using the EST coding for the human SSl protein as a probe. The 28S RNA is shown as an internal control for the amount loaded.

References

Adams, J. C. and Watt, F. M. (1993) Regulation of development and differentiation by the extracellular matrix. *Development,* **117,** 1183-98.

Akbar, A. N., Savill, J., Gombert, W., Bofill, M., Borthwick, N. J., Whitelaw, F., Grundy, J., Janossy, G. and Salmon, M. (1994) The specific recognition by macrophages of CD8+,CD45RO+ T cells undergoing apoptosis: a mechanism for T cell clearance during resolution of viral infections. *J Exp Med,* **180,** 1943-7.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment tools. *J. Mol. Biol.*, **215**, 403-410.

Anderson, K. V., Jurgens, G. and Nusslein-Volhard, C. (1985) Establishment of dorsal ventral polarity in the *Drosophila* embryo: Genetic studies on the role of the Toll gene products. *Cell,* **42,** 779-789.

Andersson, K. and Steiner, H. (1987) Structure and properties of protein P4, the major bacteria-inducible protein in pupae of *Hyalophora cecropia. Insect Biochem.,* **17,** 133-140.

Ashida, M. and Yamazaki, H. I. (1990) Biochemistry of the phenoloxidase system in insects: With special reference to its activation. In *Moulting and Metamorphosis,* eds. Ohnishi, E. and Isizaki. H. (Japan Sci. Soc. Press, Tokio und Springer-Verlag Berlin, Heidelberg) 239-265.

Barwig, B. (1985) Isolation and characterization of plasma coagulogen (PC) of the cockroach *Leucophaea maderae* (Blatteria). *J. Comp. Physiol. (B),* **155,** 135-44.

Bastiani, M. J., Harrelson, A. L., Snow, P. M. and Goodman, C. S. (1987) Expression of fasciclin I and II glycoproteins on subsets of axon pathways during neuronal development in the grasshopper. *Cell,* **48,** 745-55.

Bennett, M. R. (1998) Monoamineric synapses and schizophrenia - 45 years of neuroleptics. *Journal of Psychopharmacology,* **12,** 289-304.

Bettencourt, R., Lanz-Mendoza, H., Lindquist, K. R. and Faye, I. (1997) Cell adhesion properties of hemolin, an insect immune protein in the Ig. *Eur. J. Biochem.,* **250,** 630-637.

Biron, C. A., Byron, K. S. and Sullivan, J. L. (1989) Severe herpesvirus infections in an adolescent without natural killer cells. *N Engl J Med,* **320,** 1731-5.

Bohn, H. (1986) in *Immunity in invertebrates*, ed. Brehelin, M. (Springer-Verlag, Heidelberg) 188-207.

Boman, H. G. (1998) Gene-Encoded Peptide Antibiotics and the Concept of Innate Immunity - an Update Review. *Scandinavian Journal of Immunology,* **48,** 15-25.

Botto, M., Dell'Agnola, C., Bygrave, A. E., Thompson, E. M., Cook, H. T., Petry, F., Loos, M., Pandolfi, P. P. and Walport, M. J. (1998) Homozygous C1q deficiency causes glomerulonephritis associated with multiple apoptotic bodies. *Nat Genet,* **19,** 56-9.

Brehelin, M. (1979) Hemolymph coagulation in Locusta migratoria: evidence for a functional equivalent of fibrinogen. *Comp. Biochem. Physiol.,* **62,** 329-334.

Breton, C. and Imberty, A. (1999) Structure/function studies of glycosyltransferases. *Curr Opin Struct Biol,* **9,** 563-71.

Bulet, P., Dimarcq, J., Hetru, C., Lagueux, M., Charlet, M., Hegy, G., Van Dorsselaer A. and Hoffman, J. (1993) A novel inducible antibacterial peptide of

Drosophila carries an O-glycosylated substitution. *The Journal of Biological Chemistry*, **268**, 14893-7.

Butters, T. D. and Hughes, R. C. (1978) Lectin binding to mosquito Aedes aegyptii and human KB cells: structural comparisons of membrane oligosaccharides. *Carbohydr Res.*, **61**, 159-68.

Chasan, R. and Anderson, K. V. (1989) The role of easter, an apparent serine protease, in organizing the dorsal-ventral pattern of the *Drosophila* embryo. *Cell,* **56,** 391-400.

Davidson, D., Bretthauer, R. and Castellino, F. (1991) alpha-Mannosidase-catalyzed trimming of high-mannose glycans in noninfected and baculovirus-infected Spodoptera frugiperda cells (IPLB-SF-21AE). A possible contributing regulatory mechanism for assembly of complex-type oligosaccharides in infected cells.Biochemistry. *Biochemistry*, **30** , 9811-15.

De Luca, V. and Laflamme, P. (2001) The expanding universe of alkaloid biosynthesis. *Curr Opin Plant Biol.*, **4**, 225-33.

DeLotto, R. and Spierer, P. (1986) A gene required for the specification of dorsal-ventral pattern in *Drosophila* appears to encode a serine protease. *Nature,* **323,** 688-92.

Dewaal, A., Meijer, A. H. and Verpoorte, R. (1995) Strictosidine synthase from *Catharanthus roseus* - purification and characterisation of multiple isoforms. *Biochem. J.*, **306**, 571-580.

Doolittle, R. F. and Riley, M. (1990) The amino-terminal sequence of lobster fibrinogen reveals common ancestry with vitellogenins. *Biochem Biophys Res Commun,* **167,** 16-9.

Duvic B. and Brehelin M. (1998) Two major proteins from locust plasma are involved in coagulation and are specifically precipitated by laminarin, a beta-1,3-glucan. *Insect Biochem Mol Biol.*, **28**, 959-67.

Elrod-Erickson, M., Mishra, S. and Schneider, D. (2000) Interactions between the cellular and humoral immune responses in *Drosophila*. *Curr Biol.*, **10,** 781-4.

Fearon, D. T. and Locksley, R. M. (1996) The instructive role of innate immunity in the acquired immune response. *Science,* **272,** 50-3.

Fuller, G. M. and Doolittle, R. F. (1971) Studies of invertebrate fibrinogen. II. Transformation of lobster fibrinogen into fibrin. *Biochemistry,* **10,** 1311-5.

Gabius, H.-J. (1997) Animal lectins. *European Journal of Biochemistry,* **243,** 543-576.

Geerlings, A., Ibanez, M. M. L., Memelink, J., van der Heijden, R. and Verpoorte, R. (2000) Molecular cloning and analysis of strictosidine beta-d-glucosidase, an enzyme in terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *J.Biol. Chem.*, **275**, 3051-3056.

Geisow, M. J. (1991) The coming of age of glycobiology. *Trends Biotechnol,* **9,** 221-5.

Gillespie, J. P., Kanost, M. R. and Trenczek, T. (1997) Biological mediators of insect immunity. *Annu. Rev. Entomol.*, **42,** 611-643.

Goto, A., Kumagai, T., Kumagai, C., Hirose, J., Narita, H., Mori, H., Kadowaki, T., Beck, K. and Kitagawa, Y. (2001) A *Drosophila* haemocyte-specific protein, hemolectin, similar to human von Willebrand factor. *Biochem J.,* **359,** 99-108.

Götz, P. and Boman, H. G. (1985) In Comprehensive Insect Physiology, Biochemistry and Pharmacology; Volume 3: Integument, Respiration and

Circulation, eds. Kerkut, G. A. and Gilbert, L. I. (Pergamon Press, Oxford) 453-485.

Gregoire, C. (1974) in *The Physiology of Insecta*, ed. Rockstein, M. (Academic Press, New York) Vol. 5, 309-360.

Gum, JR Jr.(1995) Human mucin glycoproteins: varied structures predict diverse properties and specific functions. *Biochem Soc Trans.*, **23**, 795-9.

Gupta, A. P. (1984) in *Comprehensive Insect Physiology, Biochemistry and Pharmacology*, eds. Kerku, G. A. an Gilbert, L. I. (Pergamon Press, Oxford) Vol. 3, 401-451

Gupta, A. P. (1991) In *Immunology of Insects and other Arthropods,* ed. Gupta, A. P. (CRC Press, Boca Raton, Florida) 19-118.

Hall, M., Wang, R., van Antwerpen, R., Sottrup-Jensen, L. and Söderhäll, K. (1999) The crayfish plasma clotting protein: a vitellogenin-related protein responsible for clot formation in crustacean blood. *Proc Natl Acad Sci (USA),* **96,** 1965-70.

Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L. and Brunak, S. (1998) *Glycoconj J,* **15,** 115-30.

Haq, S., Kubo, T., Kurata, S., Kobayashi, A. and Natori, S. (1996) Purification, characterization and cDNA cloning of a galactose-specific C- type lectin from *Drosophila melanogaster*. *Jour. Biol. Chem.*, **271**, 20213-20218.

Harlow, E., and Lane, D. (1988) *Antibodies: A Laboratoy Manual.* (Cold Spring Harbor: Cold Spring Harbor Laboratory).

Hartl, D. L., Nurminski, D. I., Jones, R. W. and Lozovskaya, E. R. (1994) Genome structure and evolution in *Drosophila*: applications of the framework P1 map. *Proc. Natl. Acad. Sci. (USA)*, **91**, 6824-6829.

Hoffmann J. A., Kafatos F. C., Janeway C. A. and Ezekowitz R. A. (1999) Phylogenetic perspectives in innate immunity. *Science*, **284**, 1313-8.

Hultmark, D. (1993) Immune reactions in *Drosophila* and other insects, a model for innate immunity. *Trends Genet.*, **9,** 178-183.

Hypsa, V. and Grubhoffer, L. (1997) Two hemocyte populations in *Triatoma infestans*: ultrastructural and lectin-binding characterization. *Folia Parasitol,* **44,** 62-70.

Iketani, M. and Morishima, I. (1993) Induction of antibacterial protein synthesis by soluble peptidoglycan in isolated fat body from larvae of the silkworm, *Bombyx mori. Insect Biochem. Molec. Biol.,* **23,** 913-917.

Iwanaga, S. (1993) The limulus clotting reaction. *Current Opinion in Immunity,* **5,** 74-82.

Iwanaga, S., Kawabata, S. and Muta, T. (1998) New types of factors and defense molecules found in horseshoe crab hemolymph - their structures and functions. *Journal of Biochemistry,* **123,** 1-15.

Janeway, C. A. (1989) Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb Symp Quant Biol,* **54,** 1-13.

Janeway, C. A. (1994) The role of microbial pattern recognition in self:non-self discrimination. *Phylogenetic Perspectives in Immunity: The Insect Host Defence.*

Jiang, H. B. and Kanost, M. R. (2000) The clip-domain family of serine proteinases in arthropods. *Insect Biochemistry and Molecular Biology,* **30,** 95-105.

Jomori, T. and Natori, S. (1992) Function of the lipopolysaccharide-binding protein of *Periplaneta americana* as an opsonin. *FEBS Lett.,* **296,** 283-286.

Kaltner, H. and Stierstorfer, B. (1998) Animal lectins as cell adhesion molecules. *Acta Anatomica,* **161,** 162-179.

Kaltner, H. and Stierstorfer, B. (1998) Animal lectins as cell adhesion molecules. *Acta Anat. (Basel),* **161**, 162-79.

Kanost, M. R., Zepp, M. K., Ladendorff, N. E. and Andersson, L. A. (1994) Isolation and characterization of a hemocyte aggregation inhibitor from hemolymph of *Manduca sexta* larvae. *Arch. Insect Biochem. Physiol.,* **27,** 123-136.

Kapitonov, D. and Yu, R. K. (1999) Conserved domains of glycosyltransferases. *Glycobiology,* **9,** 961-78.

Kawasaki, K., Kubo, H. and Natori, S. (1993) A novel role of *Periplaneta lectin* as an opsonin to recognize 2-keto-3-deoxy-octonate residues of bacterial lipopolysaccharides. *Comp. biochem. Physiol. B,* **106,** 675-680.

Kawasaki, K., Kubo, T. and Natori, S. (1996) Presence of the *Periplaneta lectin*-related protein family in the American cockroach *Periplaneta americana. Insect Biochemistry and Molecular Biology*, **26**, 355-364.

Kerscher, S., Albert, S., Wucherpfennig, D., Heisenberg, M. and Schneuwly, S. (1995) Molecular and genetic analysis of the *Drosophila* mas-1 (mannosidase-1) gene which encodes a glycoprotein processing alpha 1,2-mannosidase. *Developmental Biology*, **168**, 613-26.

Kimmerly, W., Stultz, K., Lewis, S., Lewis, K., Lustre, V., Romero, R., Benke, J., Sun, D., Shirley, G., Martin, C. and Palazzolo, M. (1996) A P-1 based physical map of the *Drosophila* euchromatic genome. *Genome Research*, **6**, 414-430.

Kinuthia, W., Li, D., Schmidt, O. and Theopold, U. (1999) Is the surface of endoparasitic wasp eggs and larvae covered by a limited coagulation reaction?. *J. Insect Physio.*, **45**, 501-506.

Kobayashi, M., Shinohara, M., Sakoh, C., Kataoka, M. and Shimizu, S. (1998) Lactone-ring-cleaving enzyme: genetic analysis, novel RNA editing, and evolutionary implications. *Proc. Nat. Acad. Sci. (USA),* **95**, 12787-12792.

Komano, H. and Natori, S. (1985) Participation of *Sarcophaga peregrina* humoral lectin in the lysis of sheep red blood cells injected into the abdominal cavity of larvae. *Dev Comp Immunol,* **9,** 31-40.

Kopacek, P., Hall, M. and Söderhäll, K. (1993) Characterization of a clotting protein, isolated from plasma of the fresh water crayfish *Pacifastacus leniusculus*. *European J. of Biochem.,* **213,** 591-597.

Kutchan, T. M. (1995) Alkaloid biosynthesis - the basis for metabolic engineering of medicinal plants. *The Plant Cell,* **7,** 1059-1070.

Kutchan, T. M., Hampp, N., Lottspeich, F., Beyreuther, K. and Zenk, M. H. (1988) The cDNA clone for strictosidine synthase from *Rauwolfia serpentina*. DNA sequence determination and expression in *Escherichia coli*. *FEBS Lett.,* **237**, 40-4.

Lackie, A. M. (1988) Immune mechanisms in insects. *Parasitology Today,* **4,** 98-105.

Lackie, A. M. (ed.) (1986) Hemocytic and Humoral Immunity in Arthropods.

Ladendorff, N. E. and Kanost, M. R. (1991) Bacteria-induced protein P4 (hemolin) from *Manduca sexta* : a member of the immunoglobulin superfamily which can inhibit hemocyte aggregation. *Arch. Insect Biochem. Physiol.,* **18,** 285-300.

Laderach, D., Bach, J. F. and Koutouzov, S. (1998) Nucleosomes inhibit phagocytosis of apoptotic thymocytes by peritoneal macrophages from MRL+/+ lupus-prone mice. *J Leukoc Biol,* **64,** 774-80.

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. *Nature*, **227**, 680-685.

Lanier, L. (1998) NK cell receptors. *Annu Rev Immunol.*, **16**, 359-93.

Lanz-Mendoza, H., Bettencourt, R., Fabbri, M. and Faye, I. (1996) Regulation of the immune response: the effect of hemolin on the cellular immune mechanisms. *Cell. Immunol.,* **169,** 47-54.

Lavine, M. D. and Strand M. R. (2001) Surface characteristics of foreign targets that elicit an encapsulation response by the moth *Pseudoplusia includens. J Insect Physiol.,* **47**, 965-974.

Leshko-Lindsay, L. A and Corces, V. G. (1997) The role of selectins in *Drosophila* eye and bristle development. *Development.*, **124**, 169-80.

Luijendijk, T. J. C., Vandermeijden, E. and Verpoorte, R. (1996) Involvement of strictosidine as a defensive chemical in *Catharanthus roseus. J. Chem. Ecol.,* **22**, 1355-1366.

Mackenzie, P. I., Owens, I. S., Burchell, B., Bock, K. W., Bairoch, A., Belanger, A., Fournel-Gigleux, S., Green, M., Hum, D. W., Iyanagi, T., Lancet, D., Louisot, P., Magdalou, J., Chowdhury, J. R., Ritter, J. K., Schachter, H., Tephly, T. R., Tipton K. F. and Nebert, D. W. (1997) The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics*. **7**, 255-69.

Mann, P. L. and Waterman, R. E. (1998) Glycocoding as an information management system in emryonic development. *Acta Anatomica,* **161,** 153-161.

März, L., Altmann, F., Staudacher, E. and Kubelka, V. (1995) In *Glycoproteins*, eds. Montreuil, J., Vliegenthart, J. F. G., Schachter, H. and Van Deenen, L. L. M. (Elsevier, Amsterdam) Vol. 1, 543-563.

McKnight, T. D., Roessner, C. A., Davagupta, R., Scott, A. I. and Nessler, C. L. (1990) Nucleotide sequence of cDNA encoding the vacuolar protein strictosidine synthase from *Cantharanthus roseus*. *Nucleic Acids Research,* **18,** 4939.

Medzhitov, R. and Janeway, C. A. Jr. (1997) Innate immunity: impact on the adaptive immune response. *Curr Opin Immunol.,* **9**, 4-9.

Muta, T., Hashimoto, R., Miyata, T., Nishimura, H., Toh, Y. and Iwanaga, S. (1990) Proclotting enzyme from horseshoe crab hemocytes. cDNA cloning, disulfide locations, and subcellular localization. *J Biol Chem,* **265,** 22426-33.

Muta, T., Miyata, T., Misumi, Y., Tokunaga, F., Nakamura, T., Toh, Y., Ikehara, Y. and Iwanaga, S. (1991) Limulus factor C. An endotoxin-sensitive serine protease zymogen with a mosaic structure of complement-like, epidermal growth factor-like, and lectin-like domains. *J. of Biol. Chem.,* **266,** 6554-6561.

Muta, T., Nakamura, T., Hashimoto, R., Morita, T. and Iwanaga, S. (1993) Limulus proclotting enzyme. *Methods Enzymology,* **223,** 352-358.

Muta, T., Seki, N., Takaki, Y., Hashimoto, R., Oda, T. and Iwanaga, S. (1995) Purified horseshoe crab factor G: reconstitution and characterisation of the (1-3)-ß-D-glucan-sensitive serine protease cascade. *J. of Biol. Chem.,* **270,** 892-897.

Muta, T. and Iwanaga, S. (1996) The role of hemolymph coagulation in innate immunity. *Current Opinion in Immunology,* **8,** 41-47.

Nakai, K., and Horton, P. (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localisation. *Trends Biochem. Sc.,* **24**, 34-35.

Nappi, A. J. and Silvers, M. (1984) Cell surface changes associated with cellular immune reactions in *Drosophila*. *Science,* **225,** 1166-1168.

Nielsen, H., Brunak, S., and von Heijne, G. (1999) Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Prot. Engineering,* **12,** 3-9.

Pearson, A. (1996) Scavenger receptors in innate immunity. *Current Opinion in Immunology*, **8,** 20-8.

Pech, L. L. and Strand, M. R. (1996) Granular cells are required for encapsulation of foreign targets by insect haemocytes. *J. Cell Sci.,* **109,** 2053-2060.

Rademacher, T. W., Parekh, R. B. and Dwek, R. A. (1988) Glycobiology. *Annu Rev Biochem,* **57,** 785-838.

Raff, M. C. (1992) Social controls on cell survival and cell death. *Nature,* **356,** 397-400.

Rask, L., Andreasson, E., Ekbom, B., Eriksson, S., Pontoppidan, B. and Meijer, J. (2000) Myrosinase: gene family evolution and herbivore defense in Brassicaceae. *Plant Mol. Biol.,* **42,** 93-113.

Ratcliffe, N. A. and Rowley, A. F. (1979) in *Insect Hemocytes*, ed. Gupta, A. P. (Cambridge University Press, Cambridge), 331-414.

Ratcliffe, N. A. (1993a) In *Parasites and Pathogens of Insects*, eds. Beckage, N. E., Thompson, S. N. and Federici, B. A. (Academic Press, San Diego), 267-304.

Ratcliffe, N. A. (1993b) Cellular defense responses of insects: unresolved problems. *Parasites and Pathogens of Insects,* 267-304.

Rizki, R. M. and Rizki, T. M. (1990) Encapsulation of parasitoid eggs in phenoloxidase-deficient mutants of. *J. Insect Physiol.,* **36,** 523-529.

Roth, J., Kempf, A., Reuter, G., Schauer, R. and Gehring, W. (1992) Occurrence of sialic acids in *Drosophila melanogaster*. *Science*, **256**, 673-5.

Saito, T., Kawabata, S., Shigenaga, T., Takayenoki, Y. and Cho, J. (1995) A novel big defensin identified in horseshoe crab hemocytes: isolation, amino acid sequence and antibacterial activity. *J Biochem (Tokyo),* **117,** 1131-37.

Salt, G. (1965) Experimental studies in insect parasitism XIII. The haemocytic reaction. *Proc. R. Soc., Ser. B,* **162,** 303-318.

Salt, G. (1970) Experimental studies in insect parasitism XV. The means of resistance. *Proc. R. Soc., Ser. B,* **176,** 105-114.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual* Second Edition, ed. Cold Spring Harbor (Cold Spring Harbor Laboratory).

Savill, J. (1997a) Apoptosis in resolution of inflammation. *J Leukoc Biol,* **61,** 375-80.

Savill, J. (1997b) Recognition and phagocytosis of cells undergoing apoptosis. *Br Med Bull,* **53,** 491-508.

Schmidt, A., Hollmann, M. and Schafer, U. (1996) A newly identified Minute locus, M(2)32D, encodes the ribosomal protein L9 in *Drosophila melanogaster*. *Mol Gen Genet.,* **251**, 381-7.

Shimizu, Y. and Shaw, S. (1993) Mucins in the mainstream. *Nature,* **366,** 630-631.

Smith, C. L. and DeLotto, R. (1992) A common domain within the proenzyme regions of the *Drosophila* snake and easter proteins and *Tachypleus* proclotting enzyme defines a new subfamily of serine proteases. *Protein Sci,* **1,** 1225-6.

Söderhäll, K., Johansson, M. W. and Crenius, L. (1994) In *Phylogenetic Prespectives in Immunity: The insect host defense,* eds. Hoffmann, J. A., Janeway, C. A. and Natori, S. (Boca Raton: CRC Press) 97-104.

Söderhäll, K. and Cerenius, L. (1998) Role of the prophenoloxidase-activating system in invertebrate immunity. *Current Opinion in Immunology,* **10,** 23-28.

Sugumaran, M. and Kanost, M. R. (1993) Regulation of insect hemolymph phenoloxidase. *Parasites and Pathogens of Insects,* 317-342.

Sun, S. C., Lindström, I., Boman, H. G., Faye, I. and Schmidt, O. (1990) Hemolin: an insect-immune protein belonging to the immunoglobulin. *Science,* **250,** 1729-1732.

Super, M., Gillies, S. D., Foley, S., Sastry, K., Schweinle, J. E., Silverman, V. J. and Ezekowitz, R. A. (1992) Distinct and overlapping functions of allelic forms of human mannose binding protein. *Nat Genet,* **2,** 50-5.

Theopold, U., Krause, E. and Schmidt, O. (1994) Cloning of a VLP-protein coding gene from a parasitoid wasp *Venturia canescens. Arch. Ins. Biochem. Physiol.,.* **26,** 137-145.

Theopold, U., Pintér, M., Daffre, S., Reyselius, Y., Friedrich, P., Nässel, D. R. and Hultmark, D. (1995) *CalpA*, a *Drosophila* calpain homolog specifically expressed in a small set of nerve, midgut, and blood cells. *Mol. Cell. Biol.*, **15**, 824-834.

Theopold, U., Samakovlis, C., Erdjument-Bromage, H., Dillon, N., Axelsson, B., Schmidt, O., Tempst, P. and Hultmark, D. (1996) *Helix pomatia* lectin, an inducer of *Drosophila* immune response, binds to hemomucin, a novel surface mucin. *J. Biol. Chem.,* **271,** 12708-12715.

Theopold, U. and Schmidt, O. (1997) *Helix pomatia* lectin and annexin V, two molecular probes for insect microparticles: Possible involvement in hemolymph coagulation. *J. Insect Physiol.*, **43**, 667-674.

Theopold, U., Dorian, C. and Schmidt, O. (2001) Changes in glycosylation during *Drosophila* development. The influence of ecdysone on hemomucin isoforms. *Insect Biochemistry and Molecular Biology,* **31,** 189-197.

Theopold, U., Li, D., Fabbri, M., Scherfer, C. and Schmidt, O. (2002) The coagulation of insect hemolymph. *Cell Mol Life Sci,* **59,** 363-72.

Toh, Y., Mizutani A., Tokunaga, F., Muta, T. and Iwanaga, S. (1991) Morphology of the granular hemocytes of the Japanese crab *Tachypleus tridentatus* and immunocytochemical localisation of clotting factors and antimicrobial substances. *Cell Tissue Research,* **266,** 137-147.

Towbin, H., Staehelin, T., and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci (USA)*, **76**, 4350-4.

Unanue, E. R. (1984) Antigen-presenting function of the macrophage. *Annu Rev Immunol,* **2,** 395-428.

Unge, A., Tombolini, R., Molbak, L. and Jansson, J. K. (1999) Simultaneous monitoring of cell number and metabolic activity of specific bacterial populations with a dual gfp-luxAB marker system. *Appl Environ Microbiol,* **65,** 813-21

Unkeless, J. C. and Jin, J. (1997) Inhibitory receptors, ITIM sequences and phosphatases. *Curr Opin Immunol.*, **9**, 338-43.

Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology,* **3,** 97-130.

Wago, H. (1980a) Humoral factors promoting the adhesive properties of the granular cells. *Cell. Immunol.*, **54,** 155-169.

Wago, H. (1980b) Observation of the early processes of phagocytosis by the granular cells. *Appl. Entomol. Zool.*, **15,** 489-491.

Wang, D. and Marsh, J. L. (1995) Developmental regulation of the alpha methyldopa hypersensitive gene of *Drosophila melanogaster. Dev.Biol.*, **168,** 598-612.

Wheeler, M. B., Stuart, G. S. and Hapner, K. D. (1993) Agglutinin mediated opsonization of fungal blastopores in *Melanoplus differentialis. J. Insect Physiol.*, **39,** 477-483.

Wolfner, M. F., Harada, H. A., Bertram, M. J., Stelick, T. J., Kraus, K. W., Kalb, J. M., Lung, Y. O., Neubaum, D. M., Park, M. and Tram, U. (1997) New genes for male accessory gland proteins in *Drosophila melanogaster. Insect Biochemistry and Molecular Biology*, **27**, 825-834.

Wright, L. G., Chen, T., Thummel, C. S. and Guild, G. M. (1996) Molecular characterization of the 71E late puff in *Drosophila melanogaster* reveals a family of novel genes. *J Mol Biol*, **255**, 387-400.

Wright, S. D. (1991) Multiple receptors for endotoxin. *Curr Opin Immunol,* **3,** 83-90.

Wright, T. R. F. (1996) The Wilhelmine E. Key 1992 invitational lecture. Phenotypic analysis of the dopa decarboxylase gene cluster mutants in *Drosophila melanogaster. J. Heredity,* **87**, 175-190.

Wyllie, A. H., Kerr, J. F. and Currie, A. R. (1980) Cell death: the significance of apoptosis. *Int Rev Cytol,* **68,** 251-306.

Yuasa, H. J., Furuta, E., Nakamura, A. and Takagi, T. (1998) Cloning and sequencing of three C-type lectins from body surface mucus of the land slug, Incilaria fruhstorferi. *Comp. Biochem. Physiol. B, Comp. Biochem.*, **119**, 479-484.

Zhao, L. and Kanost, M. R. (1996) In search of a function for hemolin, a hemolymph protein from the. *J. Insect Physiol.*, **42,** 73-79.

Zurovec, M. and Sehnal, F. (2002) Unique molecular architecture of silk fibroin in the waxmoth, Galleria mellonella. *J Biol Chem,* **8,** 8.