Mucin-like proteins in *Drosophila* development

Zulfeqhar A. Syed
Abstract

Mucins are large and highly glycosylated proteins and major component of the mucus that coats the lining of epithelial organs. Mucins are characterized by the presence of extended regions rich in the amino acids Proline, Threonine and Serine (PTS domain), where the Serines and Threonines are O-glycosylated to form sugar-rich mucin domains. Mucins are classified into secreted gel-forming mucins and transmembrane mucins with possible signaling functions. The amino acid sequence of the PTS domains tends to be poorly conserved between species and different mucins. The goal of this thesis was to identify and study potential mucin-like proteins in *Drosophila melanogaster*. We devised a simple bioinformatic approach and developed a program that can identify PTS domains based on amino acid content. We thereby identified 36 mucins and mucin-related proteins. All proteins appear to be secreted, except for two that harbor a predicted transmembrane domain. Expression analysis at different stages of the *Drosophila* life cycle revealed that many mucins are expressed in the larval gut, consistent with a function in mucosal barrier formation. Interestingly, some of the mucins showed dynamic expression in different tubular organs during embryogenesis. Among these was Mur96B/Tenectin (Tnc) that was further studied to dissect its role in epithelial organ development. We found that Tnc is critical for diameter expansion of the developing hindgut. Tnc forms a transient matrix that fills the hindgut lumen and drives expansion in a dose-dependent manner, presumably by generating a luminal pressure. This study revealed a new mechanism in organ development, whereby the extent of lumen volume expansion can be regulated by the accumulation of single glycoprotein. In parallel to the bioinformatic approach, we identified a *Drosophila* protein that shares conserved domains with human SUSD2 and the non-mucin parts of human MUC4, called Mesh. We aimed to analyze Mesh function as a means to address the roles of these domains. Mesh was found to be is expressed in the digestive tract epithelium from mid-embryogenesis and throughout larval and adult life, localizing to the apical junction belt. Mesh is required for correct organization of the Scribble-complex, a main polarity complex conserved between fly and mammals, to prevent excess expansion of apical cell surface and for microvilli organization. The results demonstrate that mucin-like proteins, containing the PTS domains or other mucin-related domains, are essential for epithelial organ development in *Drosophila*.

**Keywords:** Mucins, PTS-domain, *Drosophila* development, Tube shape, Hindgut, Midgut, Malpighian tubules, Luminal matrix
PAPERS IN THIS THESIS

This thesis is based on the following papers, which will be referred to in the text by their roman number (I-III)


III. **Syed ZA**, Byri S, van Dijk-Härd, A Uv. Mesh is a lateral cell adhesion molecule required for apical cell membrane restriction in the *Drosophila* gut epithelium. *(Manuscript)*
Table of Contents

INTRODUCTION .................................................................................................................................1
AIM ..................................................................................................................................................2
DROSOPHILA AS A MODEL ORGANISM .......................................................................................3
  Transposons – a central tool in Drosophila research .................................................................4
  The UAS-GAL4 system for inducible gene expression ..............................................................5
  Gene silencing by RNA interference using the UAS-GAL4 system ........................................6
EPITHELIAL ORGANIZATION .....................................................................................................7
  Marginal zone ..........................................................................................................................9
  Adherens junctions ...............................................................................................................10
  Septate junctions ................................................................................................................11
  Polarity regulators .............................................................................................................12
DROSOPHILA DIGESTIVE TRACT ...............................................................................................14
MALPIGHIAN TUBULES .............................................................................................................16
GLYCOSYLATION .........................................................................................................................17
MUCINS ........................................................................................................................................18
  Gel forming mucins ..........................................................................................................19
  Transmembrane mucins .................................................................................................... 19
MUCIN-TYPE O-GLYCOSYLATION IN DROSOPHILA DEVELOPMENT ...................................20
RESULTS AND DISCUSSION ......................................................................................................23
PAPER I .......................................................................................................................................23
  Identification of mucins and mucin-related proteins in Drosophila ......................................23
  Drosophila mucins are expressed at different stages of life cycle .......................................26
  Drosophila mucins are expressed in developing epithelial organs .....................................27
PAPER II .....................................................................................................................................28
  Tenectin is an intraluminal protein required for diameter expansion of the hindgut ........28
  Tnc drives hindgut expansion in a dose-dependent manner ...............................................29
  Tnc is a component of O-glycosylated matrix in the hindgut lumen ................................30
  Model for Tnc-mediated tube dilation .............................................................................30
PAPER III ...................................................................................................................................32
  Mesh is expressed in the digestive tract and Malpighian tubules ......................................32
  Loss of mesh results in cysts in the Malpighian tubules .....................................................32
  Loss of mesh causes enlarged apical cell surfaces .............................................................33
  Mesh affects localization of the Scribble-complex to the sSJ region ................................34
CONCLUSIONS ..............................................................................................................................36
ACKNOWLEDGMENTS ..................................................................................................................37
REFERENCES ................................................................................................................................41
Abbreviations

ABP: Apical basal polarity
AJ: Adherens junction
AMOP: Adhesion-associated domain present in MUC4 and other proteins
API: Application Programming Interface
aPKC: atypical protein kinase C
Baz: Bazooka
BLAST: Basic Local Alignment Search Tool
Bub: Bubbles
Cdc42: Cell division cycle 42
CK: Cysteine Knot
Cont: Contactin
Cora: Coracle
Crb: Crumbs
Dlg: Discs Large
ECM: Extracellular matrix
EGF: Epidermal growth factor
ER: Endoplasmic reticulum
FasIII: Fasciclin III
FERM: 4.1/Ezrin/Radixin/Moesin
GalNAc: N-acetylgalactosamine
GlcNAc: N-acetylglucosamine
GlcNAcT: N-acetylglucosaminyltransferase
Gli: Gliotactin
GUI: Graphical user interface
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUK</td>
<td>Guanylate Kinase</td>
</tr>
<tr>
<td>Kune</td>
<td>Kune-kune</td>
</tr>
<tr>
<td>Lac</td>
<td>Lachesin</td>
</tr>
<tr>
<td>Lgl</td>
<td>Lethal giant larva</td>
</tr>
<tr>
<td>LRR</td>
<td>Leucine-rich repeat</td>
</tr>
<tr>
<td>MAGUK</td>
<td>Membrane associated guanylate kinase</td>
</tr>
<tr>
<td>Mega</td>
<td>Megatrachea</td>
</tr>
<tr>
<td>MZ</td>
<td>Marginal zone</td>
</tr>
<tr>
<td>NIDO</td>
<td>Extracellular domain of unknown function, found in nidogen (entactin) and hypothetical proteins</td>
</tr>
<tr>
<td>Nrg</td>
<td>Neuroglian</td>
</tr>
<tr>
<td>Nrx-IV</td>
<td>Neurexin IV</td>
</tr>
<tr>
<td>Par</td>
<td>Partition defective</td>
</tr>
<tr>
<td>Par3</td>
<td>Partition defective-3</td>
</tr>
<tr>
<td>Par6</td>
<td>Partition defective-6</td>
</tr>
<tr>
<td>Patj</td>
<td>Pals1-associated tight junction protein</td>
</tr>
<tr>
<td>PCP</td>
<td>Planar cell polarity</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDZ</td>
<td>Domain present in PSD-95, Dlg, and ZO-1/2.</td>
</tr>
<tr>
<td>PerA</td>
<td>Peritrophin-A</td>
</tr>
<tr>
<td>ppGalNAcT</td>
<td>Polypeptide N-acetylgalactosaminyltransferase</td>
</tr>
<tr>
<td>Pro</td>
<td>Proline</td>
</tr>
<tr>
<td>PTSPMiner</td>
<td>Proline Threonine Serine Pattern -Miner</td>
</tr>
<tr>
<td>Scrib</td>
<td>Scribble</td>
</tr>
<tr>
<td>Sdt</td>
<td>Stardust</td>
</tr>
<tr>
<td>SEA</td>
<td>Sea urchin sperm protein, enterokinase and agrin</td>
</tr>
</tbody>
</table>
### Mucin-like proteins in Drosophila development

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser</td>
<td>Serine</td>
</tr>
<tr>
<td>SH3</td>
<td>SRC Homology 3</td>
</tr>
<tr>
<td>Sinous</td>
<td>Sinu</td>
</tr>
<tr>
<td>SJ</td>
<td>Septate junctions</td>
</tr>
<tr>
<td>Ssk</td>
<td>Snake skin</td>
</tr>
<tr>
<td>SUSHI</td>
<td>Complement control protein (CCP) modules, or short consensus repeats (SCR)</td>
</tr>
<tr>
<td>TEM</td>
<td>Transmission electron microscopy</td>
</tr>
<tr>
<td>Thr</td>
<td>Threonine</td>
</tr>
<tr>
<td>Tnc</td>
<td>Tenectin</td>
</tr>
<tr>
<td>UAS</td>
<td>Upstream activating sequence</td>
</tr>
<tr>
<td>UDP-GalNAc</td>
<td>Uridine diphosphate $N$-acetyl-$\alpha$-galactosamine</td>
</tr>
<tr>
<td>Vari</td>
<td>Varicose</td>
</tr>
<tr>
<td>VWC</td>
<td>Von Willebrand factor C</td>
</tr>
<tr>
<td>VWD</td>
<td>Von Willebrand factor D</td>
</tr>
<tr>
<td>Yrt</td>
<td>Yurt</td>
</tr>
</tbody>
</table>
Introduction

The defining characteristic of metazoans is the presence of epithelial cells that are organized into multicellular tissues and organs. Epithelial cells cover the outer surface of the body and line all our vital internal organs. Several types of epithelia exist to fulfill important cellular and physiological functions, such as control and delivery of gases, nutrient exchange, secretion of enzymes, secretion of hormones and excretion of waste products. Most epithelia also serve an important function in protecting the underlying tissues from mechanical injury, harmful chemicals, invading microorganisms and in preventing excess loss of water by acting as selective and dynamic barriers between the internal compartments and the external environment. In its simplest form, organ epithelia consist of a layer of structurally and functionally similar epithelial cells. These can be wrapped into complex three-dimensional hollow structures, thereby generating tubular organs with diverse shape and size. During development, tubular primordia can arise by various mechanisms [1]. Once formed, however, the rudimentary tube generally has a small lumen that must grow in size to acquire appropriate dimensions to satisfy physiological demands. Acquiring characteristic shape and size is critical, as an obstructed or misshapen tube leads to compromised organ function. Consequently, defects in tube size are implicated in human diseases, such as vessel aneurysms and polycystic kidney disease [1,2].

Mucins are large, highly $O$-glycosylated proteins, and are the main component of the protective mucosa that lines the luminal surface of epithelial organs. A few studies have suggested that mucins not only function to protect epithelia from the external environment, but also have roles in epithelial organ development. These studies reported expression of mucins in human fetal organs, such as the gastrointestinal tract, respiratory tract, kidneys and male genital ducts [3-7]. Parallel studies that employed various lectins and antisera against mucin-type $O$-glycosylation to label glycans during animal development have shown that the lumen of growing epithelial organs are lined, or sometimes filled with glycan-rich components in a dynamic pattern as development proceeds. Examples include the developing rabbit kidney, embryonic chick lung, and most epithelial organs of the fruit fly [8-12]. Thus, some of the $O$-glycans detected during development might represent mucin-like proteins. The presence of luminal components
rich in O-glycans in different developing organs and from different species is intriguing and suggests important roles for such components during development.

**Aim**

The aim of this thesis is to identify mucin and mucin-related genes in *Drosophila* and characterize their potential involvement in epithelial organ development.
**Mucin-like proteins in Drosophila development**

**Drosophila as a model organism**

The fruit fly, *Drosophila melanogaster*, was first introduced as an invertebrate model organism to study classical genetics more than a century ago. In 1910, the discovery of the white mutation by Thomas Hunt Morgan and subsequent contribution by his graduate students kick-started the systematic use of *Drosophila* for genetic research. Since then, fly genetics has been successfully applied to study fields spanning from developmental biology to physiology, enriching our understanding of the genetic principles and molecular mechanisms underpinning biology [13]. Besides being a powerful model organism to study basic biology, *Drosophila* has over the years been widely used to study genetic components of various human pathologies, as it turns out that around 75% of the known human diseases genes have counterparts in fruit flies [14,15].

The enormous success of *Drosophila* as a model organism originates from the numerous practical advantages it has to offer. In addition to its powerful genetics and small size, fruit flies have a short generation time. *Drosophila* undergoes holometabolic development; this involves complete transformation of the immature larva, mostly lacking adult structures, into the adult fly (imago). The life cycle of *Drosophila* consists of six stages: embryo, 1\textsuperscript{st} instar larva, 2\textsuperscript{nd} instar larva, 3\textsuperscript{rd} instar larva, pupa and adult. The duration of the life cycle varies with temperature. At 25°C, embryonic development takes approximately 21-22 hours, after which the embryo hatches into a larva. The larva grows continuously, undergoing two molts from 1\textsuperscript{st} instar to 3\textsuperscript{rd} instar larva and a final molt into an immobile pupa over a period of 4 days. During the pupal stages, larval tissues undergo histolysis while new adult body structures are built in a process called metamorphosis. After 5 days of pupariation, adult flies eclose from the pupal case, and it takes up to 8 hours for the newly eclosed flies to attain sexual maturity. Overall, the life cycle of *Drosophila* takes ten days at 25°C.

*Drosophila* is relatively easy and cost-effective to maintain in the lab, facilitating high throughput experiments involving large numbers of different fly stocks. The fruit fly has three pairs of autosomal chromosomes and two sex chromosomes (X and Y). Recombination is confined only to female flies, which is a major advantage when performing fly genetics. *Drosophila* has many genetic tools in its arsenal, of which the most unique are the balancer chromosomes, which suppress recombination between
homologous chromosomes, and plethora of phenotypic markers. In addition, a multitude of transposon variants enable genetic strategies to manipulate a gene of interest at various temporal and spatial resolutions [16-18]. The constantly growing *Drosophila* research community and the collaborative effort to broaden the technical repertoire of *Drosophila* genetic tools make it a favourite organism for many researchers.

**Transposons – a central tool in *Drosophila* research**

Transposons are mobile genetic elements present in the genomes of most metazoans and have become an important tool in genome research [19]. Transposable elements provide powerful means to understand genome evolution and as tools for genetic manipulation. In general, transposable elements encode enzymes called transposases that mediate DNA cleavage and transposition of the transposable element in the genome. Transposable elements are commonly known as “jumping genes” and were discovered by geneticist Barbara McClintock in the 1940s while she was studying the pattern of pigmentation in maize. McClintock showed that the irregular pigmentation in maize was caused by genetic elements that transposed from one locus to another. For her ground breaking work on transposable elements, McClintock was awarded Nobel Prize in 1983 [20].

The *Drosophila P element* is one of the most widely used and best characterized eukaryotic transposons. *P elements* are thought to have entered *D. melanogaster* by horizontal transfer from another distantly related *Drosophila* species about 80 years ago [21]. They were first recognised as factors in P strains responsible for hybrid dysgenesis, and since then they have become widely used tools for studying gene function in *Drosophila* [22]. The *Drosophila P element* is a 2.9 kb DNA transposon that encodes a 87 kDa transposase protein, and transposition within the genome occurs by a cut-and-paste mechanism that requires approximately 150 bps of specific sequence at each end of the *P element* [23]. The sequences required for transposition includes 31 bps terminal inverted repeats, internal transposase-binding sites, and internal 11 bps inverted repeats [24-26]. *P elements* can either be autonomous, where they encode their own source of transposase needed for mobilisation, or non-autonomous, where an external source of transposase is required. Non-autonomous *P elements* in the form of engineered constructs were initially used for random gene disruptions [27]. Since then, *P elements* have been adapted and modified for different purposes of transgenesis, such as various types of gene tagging,
insertion of specific enzymatic target sites into the genome, and inducible gene expression
[16,28-31]. Other well-studied DNA transposons that have now been adapted for use in
Drosophila research are the piggyBac and Minos elements [32-35]. Genetic and
molecular data of all Drosophila genes including different transposon-induced alleles and
transgene information is made available by the FlyBase consortium [36].

The UAS-GAL4 system for inducible gene expression

Targeted gene expression is an important tool in the characterization of individual gene
function. The UAS-GAL4 system is an extremely useful tool for selective expression of
any cloned gene in a wide variety of cell and tissue specific patterns in Drosophila. This
binary system is based on the Saccharomyces cerevisiae transcriptional activator called
GAL4 that binds to specific DNA sequence called UASG (galactose upstream activating
sequence) and activates the transcription of linked genes [16,37]. The key feature of this
system is that the GAL4 gene and the UAS-target gene, both of which are introduced into
the fly genome by transposon-mediated integration, are initially separated into two
distinct transgenic lines. One strain expresses GAL4 under the control of a tissue specific
enhancer or promoter and is generally referred to as the driver line. The other strain
carries the gene of interest or reporter gene downstream of the UAS sequence. GAL4 has
no detrimental consequences in the fly, even at elevated levels, and the UAS transgenes
are largely silent in the absence of GAL4. When the fly strains are crossed to each other,
the combination of the two transgenes in the progeny of the cross results in GAL4-
binding to the UAS sequence and activation of the target gene (Figure 1). The progeny
can thus be conveniently analyzed to study the effects of directed gene expression. There
are wide selections of GAL4 driver lines available that express GAL4 in different cells
and tissues at different stages of development, or upon conditional induction, allowing for
targeted expression of a UAS transgene in a selected spatio-temporal manner.
Figure 1: The UAS-GAL4 system allows targeted expression of any cloned gene in a tissue-specific manner. This bipartite system utilizes the yeast transcriptional activator GAL4 to activate expression of a target gene fused downstream of UAS. This system consists of two distinct transgenic strains: a GAL4-driver line and a UAS-line. The GAL4-driver carries the GAL4 gene inserted into the Drosophila genome and expresses GAL4 under the control of nearby enhancers. In the progeny of a cross between these transgenic fly strains, GAL4 binds to the UAS sequences to activate expression of the linked gene in the cells where GAL4 is expressed. Figure adapted from [16]

Gene silencing by RNA interference using the UAS-GAL4 system

Ribonucleic acid interference (RNAi), also referred to as post-transcriptional gene silencing, is an important biological pathway in which double-stranded RNA (dsRNA) molecules induce sequence-specific inactivation of gene function. The phenomenon of RNAi was first discovered in the nematode Caenorhabditis elegans [38] and is an endogenous cellular mechanism used by most eukaryotes to regulate gene expression [39]. When dsRNA molecules are present a cells, it triggers the RNAi machinery, wherein dsRNA is cleaved into smaller fragments that are used to target homologous mRNA sequences for degradation, resulting in inactivation of gene expression [40]. In recent years, RNAi has developed into a powerful tool to manipulate gene expression, and is
used for many invertebrate and vertebrate model organisms and cell culture systems to probe gene function [41].

In *Drosophila*, RNAi can be induced either by injecting embryos with *in vitro* transcribed dsRNA before cellularization or by germline transformation of a transgene that expresses a hairpin-forming RNA sequence under the control of the UAS promoter [41]. RNAi mediated by injection has its limitation in that studies of gene function are restricted to embryonic development, and sometimes, maternal contribution may alter embryonic phenotype. Targeted expression of an RNAi transgene, using the UAS-GAL4 system offers several advantages over the injection approach. Depending on the GAL4-driver line used, cell-type, tissue-specific or developmental stage specific probing of gene function can be achieved. This approach allows spatio-temporal control of gene knockdown and has been extensively used in reverse genetics for rapid investigation of gene function. Transgenic-based RNAi is simple, as it requires only one fly cross, and communal efforts have been made to generate RNAi libraries covering most of the *Drosophila* protein-encoding genes [42]. In *Drosophila* cell culture systems, RNAi can be activated by adding dsRNA to the cell culture [43].

**Epithelial organization**

Organ epithelia are polarized, such that the apical surface faces the luminal space of an organ or the exterior of an organism. The basal domain faces the basement membrane or the underlying extracellular matrix (ECM) mediating cell-matrix adhesion [1]. This apical to basal polarization is manifested in each epithelial cell, which displays well-defined apical, basal and lateral membrane domains with different protein and lipid composition and oriented organization of cytoskeletal components and cytoplasmic organelles (Figure 2). The lateral cell domain faces neighbouring epithelial cells and connects the cells by means of structurally defined intercellular junctions [1,44]. Such asymmetric partition of along the apical to basal axis of cells confers specific structural and functional properties to epithelia and is generally referred to as apical-basal polarity (ABP) [2]. In addition to ABP, many epithelial tissues are polarized along the plane of the epithelium; this is known as planar cell polarity (PCP) [2]. The ability of epithelial cells to maintain their ABP is essential for preserving tissue integrity and, consequently, loss of cell polarity due to infection, diseases or genetic predisposition underlie many human pathologies [45,46].
Figure 2: Typical architecture of a simple epithelial tube. The tube wall is composed of epithelial cells that display apical-basal polarity. The apical surface faces the lumen, while the basal surface is exposed to the underlying basement membrane and interacts with the ECM. The lateral membrane faces neighbouring cells and possesses structurally defined intercellular junctions that provide cell-cell adhesion and a diffusion barrier between the apical and basolateral surfaces. Figure adapted from [2]

In *Drosophila*, the apical membrane domain encompasses the free apical surface and a narrow region of cell-cell contact at the most apical part of the lateral domain known as the marginal zone (MZ) [44]. The MZ corresponds to the position of the vertebrate tight junctions (TJ), and several apical polarity regulators contained within the MZ have mammalian homologues that are found to localize to the TJs [47,48]. Unlike the MZ, the TJs also provide a permeability seal that restricts free diffusion of ions and solutes across the paracellular space and forms a “fence” that separates the apical and basolateral domains [49]. Basal to the MZ lies the adherens junction (AJ) that provides strong cell-cell adhesion by forming a circumferential belt around the epithelial cell [44]. The epithelial barrier functions in *Drosophila* are mediated by the septate junctions (SJ), which are found basal to the AJs (Figure 3) [44].
Figure 3: Schematic presentation of the apical junctional complex in chordates (left) and Drosophila (right). The apical-most region of cell-cell contact is represented by the tight junction (TJ) in chordates and the marginal zone (MZ) in Drosophila. Both Drosophila and chordates exhibit adherens junctions (AJs) basal to the MZ and TJ, respectively. Below the AJs in Drosophila are specialized junctional structures known as the septate junction (SJ). SJs are characterized by ladder-like bridges when observed using electron microscopy, unlike TJs that appear as anastomosing intramembranous strands. Despite the morphological differences, both junctions are responsible for maintaining the diffusion barrier and separating the apical and basolateral domains. Desmosomes are indicated below the AJs in chordates, and are absent in Drosophila. Figure adapted from [44].

Marginal zone

Several apical polarity proteins localizes to the MZ. Traditionally, these have been defined into the Crumbs and Par (partitioning defective) polarity complexes, which function as apical determinants and regulators of epithelial cell polarity in Drosophila [50]. The Crumbs complex consists of Crumbs (Crb), Stardust (Sdt), Pals1-associated tight junction protein (Patj) and Lin-7 [50]. Crb is a large transmembrane protein with twenty-nine epidermal growth factor (EGF)-like domains and four Laminin A G-domains in the extracellular region [51]. The extracellular domain of Crb is involved in homophilic Crb-Crb interactions [52]. The transmembrane region is followed by a small highly conserved cytoplasmic region, which contains a functionally important binding site for the
4.1/Ezrin/Radixin/Moesin (FERM) domain and a C-terminal Postsynaptic density 95/Discs large/Zonula occludens-1 (PDZ) binding motif [51]. Crb is detected in all ectodermally derived epithelia from the time of gastrulation and confers apical membrane characteristics and promotes apical membrane growth [53-55]. Sdt is a membrane-associated guanylate kinase (MAGUK) protein belonging to the MPP/P55 (membrane protein palmitoylated) subfamily of MAGUKs [50]. It contains a PDZ domain, an SH3 (SRC homology 3) domain, a GUK (guanylate kinase) domain, a HOOK domain and two L27 domains. In addition, it contains the evolutionary conserved ECR1 and ECR2 domains at the N-terminus. The PDZ domain of Std binds to the cytoplasmic C-terminus of Crb [56], and the two L27 domains bind to Patj and Lin-7, respectively [57]. Loss of Sdt results in similar epithelial defects as seen with loss of Crb [51]. Patj has a single L27 domain and four PDZ domains and plays a minor role in epithelial polarization [58].

The Par complex consists of Bazooka (Baz, Partitioning defective-3 (Par3) in C. elegans), Partitioning defective-6 (Par6), atypical protein kinase C (aPKC) and the small GTPase Cdc42 [50]. Baz contains three PDZ domains and together with Par6, which has a single PDZ domain and a semi-CRIB domain, forms a complex with aPKC [59,60]. Baz has an early role in the formation of AJs [61]. aPKC is a serine/threonine kinase with several important targets that contribute to its role as an evolutionary conserved epithelial polarity determinant [50,62,63]. There are multiple interactions among the members of the Par complex and the Crb complex, and both these complexes are required for the establishment of ABP [50,63,64].

Adherens junctions

The adherens junctions (AJs) are cell-cell adhesion complexes located below the marginal zone [44]. They define the boundary between the apical and basolateral domains, and have multiple roles in development and homeostasis, including cell-cell adhesion, anchoring of the cytoskeleton to the plasmamembrane, signal transduction and transcriptional control [65]. Classical cadherins are the core components of AJs, and the assembly of AJs typically begins by homophilic cis- and trans-clustering of these transmembrane proteins [61]. The basic features of cadherins are the presence of extracellular cadherin repeats that mediate calcium dependent cell-cell adhesion and a highly conserved cytoplasmic tail that interacts with cytoplasmic proteins called catenins.
In *Drosophila*, members of the classical cadherin family include: DE-cadherin/Shotgun (DE-Cad), DN-cadherin (DN-Cad) and DN-cadherin2 (DN-Cad2) [61]. DE-Cad is the major epithelial cadherin. Its cytoplasmic tail binds to p120 catenin and beta-catenin (Armadillo), and beta-catenin binds to alpha-catenin that subsequently links to the actin filament, forming a circumferential actin belt around the cells [61].

**Septate junctions**

Septate junctions (SJs) are cell-cell junction complexes found basal to the adherens junctions [44]. After the establishment of ABP and AJs, these junctions can be detected by electron microscopy as ladder-like septa that span the intermembrane space [66]. The SJ strands meander along the lateral membrane forming a labyrinth-like structure that inhibits diffusion of molecules between the apical and basal domains [67,68]. They thereby provide a paracellular barrier and a fence function analogous to the vertebrate tight junctions, and mutational analyses of SJ proteins has revealed disruption of the paracellular barrier in dye permeation experiments [68].

SJs are structurally and molecularly similar to vertebrate paranodal junctions, which are formed between axons and myelinating glial cells at the node of Ranvier [44]. SJs are classified into two types, pleated septate junctions (pSJs) and smooth septate junctions (sSJs). pSJs are found in most of the ectodermally derived epithelia and glial sheets, whereas sSJs are found in endodermally derived tissues [66]. The difference between the pSJs and sSJs is the arrangement of SJ-strands. The strands in pSJs form regular undulating "pleated" lines, whereas the strands in sSJs form linear bands [66]. pSJs are the most prominent junctions in *Drosophila*, and recent studies have identified several SJ components involved in epithelial barrier formation, such as Neurexin IV (Nrx-IV), Neuroglian (Nrg), Na+/K+-ATPase α- and β- subunit (ATPα and Nrv2), Gliotactin (Gli), Lachesin (Lac), Contactin (Cont), Coracle (Cora), Yurt (Yrt), Varicose (Vari), Disc Large (Dlg), Lethal giant larva (Lgl), Scribble (Scrib), and Fasciclin III (FasIII) [69-79]. In addition, three homologues of Claudin, a major component of vertebrate TJJs, are part of *Drosophila* SJs and are called Megatrachea (Mega), Sinous (Sinu), and Kune-kune (kune) [80-82]. Intact pSJs also appear to be required for correct apical secretion of chitin modifying enzymes [83]. In a recent study, it was found that many pSJ components (Nrx-IV, Nrg, ATPα, Nrv2, Sinu, Mega, and Vari) are highly mobile on the lateral membrane at
embryonic stage 12, when the SJs begin to form, but become rather immobile at stage 13, suggesting that they form a structural core of the SJs [84]. Loss of any of the core SJ components dramatically affects the mobility of the others, indicating interdependence for stable SJ complex formation. On the other hand, Lgl, Dlg and Scrib remain relatively mobile, also after stage 13, and are not considered to be part of the core complex [84].

Scrib is a membrane-associated scaffolding protein and belongs to the LAP (LRR and PDZ) protein family. It contains sixteen N-terminal Leucine-rich repeats (LRRs) and four PDZ domains [78]. Dlg is a member of the MAGUK super family of proteins, which has three PDZ domains, a SH3 domain and a GUK domain and acts as a scaffolding protein [77]. PDZ domains bind to short PDZ-binding motifs that are located in the C-terminus of target proteins, and help to anchor the target protein to the correct membrane domain. Both transmembrane and cytosolic proteins can be targeted to membrane complexes through PDZ interactions [85]. Lgl is a WD40 repeat protein and, unlike Dlg and Scrib, is not a scaffolding protein. Lgl has conserved phosphorylation sites that are critical for its localization and function [86]. Mutations in any of these components result in disruption of epithelial organization and expansion of the apical membrane [64].

In contrast to pSJs, relatively few components of the sSJs have been characterized. Some studies have reported the localization of αβ spectrin, Ankyrin and FasIII to the sSJs region in the midgut [87,88]. A better understanding of the components of the sSJs, and how the polarization machinery works in endoderm-derived organs is only now emerging. Two molecular components specific for the sSJs have recently been identified, called Snake skin (Ssk) and Mesh [89,90]. These were found to be expressed in endoderm-derived tissues like the midgut, proventriculus (out layer) and Malpighian tubules. Loss of either of these components results in a compromised paracellular barrier in the midgut and mis-localization of other sSJ-associated proteins, such as FasIII, Cora and Lgl, accompanied by larval lethality [89,90].

**Polarity regulators**

In an epithelial tissue, the shape and function of the constituting cells are completely dependent on their polarization. The separation of the distinct surfaces of the plasmamembrane prevents mixing of receptors, channels and transporters between the
domains. It also retains key protein complexes that are involved in protein sorting, recycling, trafficking, and signalling to distinct compartments [50]. Genetic studies in Drosophila and C.elegans have identified a set of conserved proteins that are involved in the establishment and maintenance of cell polarity [55,64,78,91-93]. These include, in addition to the members of the Crb, Par and Scrib complexes described above, components of the recently identified Yrt/Cora complex (Yurt, Cora, Nrx-IV and Na+/K+-ATPase) [94] and Partitioning defective-1 (Par-1), which localize to the basolateral cell domain [95]. These complexes act in a mutually antagonistic relationship to define the apical and basolateral domains.

The traditional description of the apical Crb and Par polarity complexes is now changing, as the components of the two complexes appear to interact in an interdependent apical protein network. For example, Crb can bring together Sdt, PatJ, PAR-6 and aPKC though its intracellular domain [50,96]. Polarization of different epithelial cell types can require different components of the polarization network and can occur by slightly different mechanisms, but a simplified mechanism of polarity establishment has been put forward [96]. In this model, the transmembrane protein Crb defines apical cell identity by localizing Sdt, PatJ, PAR-6 and aPKC to the MZ through protein interactions mediated by its intracellular domain. This network excludes the presence of Baz in the apical cell domain via aPKC-mediated phosphorylation of Baz. Baz interacts with components of the AJs, recruiting these to the region where Baz is enriched. The apical polarity regulators therefore restrict AJ formation in the apical direction. The basolateral polarity regulators prevent Baz from moving in the basal direction by Par-1-mediated phosphorylation of Baz. The apical and basal polarity proteins also antagonize each other. The apical polarity proteins exclude the Scrib complex from the apical cortex through phosphorylation of Lgl by aPKC. The Scrib complex in turn antagonizes the apical regulators, partly through interaction of Lgl with aPKC, to maintain the basolateral domain. The second basolateral polarity complex Yurt/Cora negatively regulates the activity of the Crb complex and stabilizes the basolateral membrane [94]. Thus, the apical and basal polarity regulators establish ABP by creating an equilibrium through mutually modulating each other activity (Figure 4) [96].
**Figure 4:** Epithelial polarity factors and their interactions. Positive feedback among members of the apical polarity determinants and mutual antagonism between the apical and basolateral determinants are required for the formation and maintenance of apical-basal polarity. *Figure adapted from [96]*

### Drosophila digestive tract

The *Drosophila* larval digestive tract is divided into three distinct anatomical regions: the foregut, midgut and hindgut. The foregut and hindgut are ectodermal in origin, while the midgut is derived from the endoderm and forms a secondary epithelium by undergoing a mesenchymal-to-epithelial transition during late embryogenesis [97,98]. The foregut arises from invagination of stomodeal cells from the anterior region of the blastoderm embryo. The posterior part of the foregut makes contacts with the anterior midgut, and through a series of cellular events, including cell division and cell shape changes, the posterior part of the foregut tube makes an inward movement and becomes surrounded by the anterior-most part of the midgut. This gives rise to a bulb-like three-layered structure called the proventriculus, where the outer layer is derived from the endoderm and the middle and inner layers are ectodermal in origin [99,100]. Thus, the proventriculus develops at the boundary of the foregut and midgut and serves as a valve regulating the...
passage of food into the midgut. The fully developed foregut consists of the atrium, pharynx, oesophagus and proventriculus [98-100].

The larval midgut is composed of two cell layers [97]. The outer visceral muscle layer is organized into circular and longitudinal muscles and is derived from visceral mesoderm, while the inner epithelial layer is derived from the endoderm [97,101]. The midgut arises from two spatially separated primordia at the anterior and posterior ends of the blastoderm embryo, called the anterior midgut primordium (amp) and the posterior midgut primordium (pmg) [102]. The visceral mesoderm derives from clusters of mesodermal cells in parasegment 2-13, which join together to form a continuous band of cells on each side of the embryo [102]. The formation of the midgut begins when the midgut primordia (amg and pmg) invaginate, lose their epithelial properties through epithelial-to-mesenchymal transition and start migrating along the bands of visceral mesoderm that serve as tracks [97,101]. The migrating primordia meet in the middle of the embryo, undergo a mesenchymal-to-epithelial change and fuse to form two bands of cells, that along with visceral mesoderm, extend ventrally and dorsally to wrap around the yolk to form a midgut tube. Depending on the interaction of the visceral mesoderm with the endoderm, the midgut tube generates three constrictions that subdivide the midgut into four lobes. The first lobe gives rise to the outer layer of the proventriculus and the four gastric caeca. The second, third and the fourth lobe develop into the anterior midgut, middle midgut and posterior midgut respectively [97,102].

The hindgut arises by invagination of a group cells, called the proctodeal primordium, at the posterior end of the blastoderm embryo [98]. The invagination elongates through convergent cell extension, accompanied by changes in cell size and cell shape to form a narrow, left-right asymmetric, shepherd’s crook shaped tube. A transcriptional hierarchy consisting of Drumstick (Drm), Lines (Lin) and Bowl controls hindgut patterning during tube elongation [103,104]. The elongated hindgut tube is divided into morphologically distinct sub domains. The anterior-most domain, called the small intestine, lies just posterior to the midgut and is followed by the large intestine and the posterior-most rectum. The large intestine is partitioned into ventral and dorsal domains by two lines of cells at each lateral side of the tube, referred to as ‘border cells’ [105]. The border cells also form circumferential rings at the border between the small and the large intestine and
between the large intestine and the rectum. During later stages of development, the hindgut grows by tube elongation and by luminal diameter expansion [104].

**Malpighian tubules**

The Malpighian tubules are the excretory organs of *Drosophila*, and are functionally equivalent to the vertebrate kidney. The larval Malpighian tubules consist of two pairs of single-cell layered epithelial tubes that originate from the hindgut during embryogenesis [98,106]. The development of the Malpighian tubules in *Drosophila* involves successive morphogenetic events including (a) cell specification and eversion of the tubule primordia (b) cell proliferation (c) cell rearrangement and tube elongation and (d) cell differentiation [106].

Malpighian tubule cells are specified by interactions between the midgut and the hindgut and depends on the zinc-finger transcription factor Krüppel (Kr) and the homeodomain-containing protein Cut [107]. In the presence of the Kr and Cut transcriptional regulators, four clusters of Malpighian tubule primordial cells start to bud from the hindgut. One cell from each bud is selected to become a tip cell by lateral inhibition, specified by the Notch pathway. This tip cell secretes EGF and promotes cell division in its neighbouring cells [107,108]. When cell division ceases, the tubules are short with 8-12 cells encircling the lumen. These cells make up the main cell type of the Malpighian tubules, known as the principal cells (PC) [106]. Subsequent tubule growth and elongation occurs largely by cell rearrangements and cell intercalation that decreases the number of cells at the circumference of the tubules. As the tubules extend, they undergo stereotypic path-finding through the body cavity with the anterior tubules moving forward towards the thorax region and the posterior tubule protruding along either side of the hindgut [106]. During this phase of tube elongation, a population of cells from the caudal mesoderm incorporate into the tubule, by undergoing a mesenchymal-to-epithelial transition, and differentiate into a physiologically distinct cell type known as stellate cells (SC) [109]. As the tubules develop, SC are progressively integrated into the epithelium between the PC [106]. The SCs become apical-basal polarized once they are incorporated into the tubules. At the end of embryogenesis, the Malpighian tubules have attained their final architecture with an extensive increase in length and a narrow lumen with two cells at the circumference. Before hatching, precipitates of uric acid are visible in the tubule lumen, indicating the
onset of excretory activity. The excretory function of the Malpighian tubules relies on the combined function of PCs and SCs [106].

**Glycosylation**

Post-translational modifications are of critical importance to the function of an expressed protein [110]. Two of the most abundant forms of posttranslational modifications that involve carbohydrates are *N*- and *O*-linked glycosylation, distinguished by their glycosidic linkages to amino acid side chains [111]. Glycosylation results in the addition of sugar groups to the protein, and takes place in the lumen of the endoplasmic reticulum (ER) and the Golgi complex [111].

*N*-linked glycosylation is initiated in the ER, and further processing takes place in the Golgi complex. *N*-linked glycans are characterized by being linked to the amide nitrogen atom in the side chain of Asparagine (Asn). An asparagine residue can accept an oligosaccharide only if the residue is part of an Asn-X-Ser or Asn-X-Thr consensus sequence, where X can be any amino acid except proline [112]. *N*-linked glycoproteins acquire their initial sugars from Dolichol donors in the ER [112]. All *N*-linked oligosaccharides have in common a oligosaccharide core, consisting of three glucoses, nine mannoses and two *N*-acetylglucosamine (GlcNAc) residues (Glc₃Man₉GlcNAc₂), which serve as the foundation for a wide variety of *N*-linked oligosaccharides that are categorized into High-mannose type, Complex type and Hybrid type [112]. The final oligosaccharide structure acquired on the mature glycoprotein is dictated by the action of different glycosyltransferases and glycosidases residing in the ER and Golgi complex [112].

*O*-linked glycosylation takes place in the Golgi complex [113]. *O*-linked glycans are linked to the oxygen atom in the side chain of Serine (Ser) or Threonine (Thr). Unlike *N*-glycosylation, *O*-glycosylation does not begin with the transfer of an oligosaccharide from a Dolichol precursor, but with the addition of a single monosaccharide [113]. Mucin type *O*-glycosylation is initiated by the enzymatic addition of a *N*-acytylgalactosamine (GalNAc) residue to the side chain of Ser or Thr by the UDP-GalNAcT:polypeptide *N*-acytyl-galactosaminyltransferase (ppGalNAcTs), referred to as GalNAc transferases in mammals and PGANTs in *Drosophila* (EC 2.4.1.41) to generate the Tn-antigen (GalNAc-
α-1-O-Ser/Thr) [114,115]. Subsequent elongation by transferases yields eight distinct core structures, which can be further elongated or modified by Sialylation, Sulfatation, Acetylation, Fucosylation, and Polylactosamine-extension to build hundreds of different O-glycan chains [113,114]. O-glycans with O-linkages to Ser or Thr other than GalNacA includes O-linked fucose, glucose, mannose, xylose and GlcNAc. A large family of ppGalNAcTs exists, which indicates redundancy in the activity of these enzymes and spatio-temporal expression and substrate preference [114-116].

Mucins

Mucins are large and highly glycosylated multifunctional proteins found on the surface of epithelial tissues lining the respiratory, digestive and urinogenital tracts [117,118]. Mucins are the major component of the mucus that protects underlying epithelial cells from infection, dehydration and physiological or chemical injury [119]. A common structure in mucins is a protein backbone termed “apomucin”, which is decked with a large number of O-linked oligosaccharides and a few N-glycan chains [118]. Apomucins contain variable numbers of tandem repeats that are particularly rich in amino acids Ser and Thr whose hydroxyl groups will become O-linked with oligosaccharides. These tandem repeat regions are called PTS domains (Proline, Threonine and Serine) or mucin-domains. The O-linked oligosaccharides account for up to 80% of the molecular mass of the mucin and results in a highly extended and rigid structure of the mucin [118]. They have high water holding capacity, and are therefore largely responsible for the viscous nature of mucus [119]. The PTS domains are not conserved between species and can vary from one mucin to another [120,121]. The heavily glycosylated mucin domains adapt an outstretched conformation, best described as a “bottle brush”, where the stalk represents the protein backbone and the bristles are represented by oligosaccharide chains [111].

Mucins have been subdivided into gel-forming and membrane-bound forms. In humans, there are nine membrane-bound mucins (MUC1, MUC3, MUC4, MUC12, MUC13, MUC15, MUC16, MUC17 and MUC20) and five secreted gel-forming mucins (MUC2, MUC5B, MUC5AC, MUC6 and MUC 19 [118].
Gel forming mucins

A characteristic of gel-forming mucins is the capacity of monomers to form polymeric structures. Secreted mucins are produced by specialized cells, generally referred to as “goblet cells” [119]. The secreted mucins contribute to the formation of a physical barrier that protects epithelial cells lining the respiratory, urinogenital and gastrointestinal tracts [118]. Gel-forming mucins have several VWD (Von Willebrand factor-D) and VWC (Von Willebrand factor-C) domains flanking the mucin domains. They also harbour cysteine-rich regions named ‘CK’ domains (Cystine Knot) at their C-terminal ends [118]. MUC2, a major gel-forming mucin of the colon forms dimers via its C-terminal and trimmers via its N-terminal that leads to a polymeric structure [122-124].

Transmembrane mucins

Transmembrane mucins are present along the apical surface of epithelial cells. The human transmembrane mucins are characterized by either a SEA (sea urchin sperm protein, enterokinase and agrin) domain or a special variant of the VWD (Von Willebrand factor D) domain that is lacking cysteines. From amino- to carboxyl ends, the overall structure of membrane-bound mucins exhibits three main regions: (I) An extracellular domain, which carries the mucin domain and extends far from the surface of the cell, (II) a type I transmembrane domain that spans the lipid bilayer layer and (III) a short cytoplasmic tail (Figure 5B) [118]. Several of the human transmembrane mucins are known or predicted to be cleaved in their SEA or in VWD domain to yield two peptides that remain attached by non-covalent forces [125-127]. The cytoplasmic tails of some of the transmembrane mucins have been implicated in different cell signalling events [128,129].
**Typical example of a gel-forming mucin (A) and a transmembrane mucin (B).** Gel-forming mucins are synthesized in specialized cells, known as goblet cells that are characterized by large mucin-packed secretory granules. Upon regulatory signals or stimulation, mucin granules are released and can expand up to 1000-fold on hydration. Gel-forming mucins form large polymers through oligomerization/multimerization, which are held together with numerous disulphide bonds. Transmembrane mucins are expressed at the apical cell surface and appear cleaved at SEA/vWD domains into amino- and carboxy-terminal subunits that are held together by non-covalent forces. The N-terminal subunit harbours highly glycosylated mucin-domains that are tethered to the C-terminal transmembrane subunit. The mucin-domains extend far from the cell membrane into the glycocalyx. (Protein backbones are shown in brown and oligosaccharides in green). *Figure adapted from [130]*

**Mucin-type O-glycosylation in *Drosophila* development**

In *Drosophila*, mucin-type O-glycosylation is initiated by PGANTs. There are at least twelve putative genes in the *Drosophila* genome encoding PGANTs, out of which nine have been demonstrated to have enzymatic activity *in vitro* [131,132]. Structurally, members of this family are type II transmembrane proteins, consisting of a short cytoplasmic tail at the N-terminus that is tethered to the Golgi membrane by means of a transmembrane domain and a highly conserved catalytic domain at the C-terminus that lies within the Golgi lumen [131,132]. Biochemical studies have shown that members of the PGANT family have a hierarchy of enzymatic activity [132]. Similar to mammalian...
GalNAc transferases, PGANTs have been categorized into two groups based on their activity. One group consists of enzymes that catalyse the initial addition of GalNAc to unmodified peptides (peptide transferases), while the other group of enzymes act on previously glycosylated substrates that contain GalNAc residues (glycopeptide transferases) [132]. The initial addition of GalNAc to selected Ser/Thr residues on the protein backbone by PGANTs results in the so-called Tn-antigen, which can be further extended by addition of galactose by core 1 $\beta_1,3$-galactosyltransferase (C1GalT1) to form a core 1 structure, called the T-antigen [133]. Unlike mammalian mucin-type O-glycans, which have several arrays of high-order O-glycan structures, Drosophila O-glycans tend to be shorter and less extended, and they mainly consist of Tn-antigens and T-antigens [134,135]. Expression analysis of individual PGANTs has revealed highly dynamic spatio-temporal and frequently overlapping patterns of expression during Drosophila embryogenesis [136]. This dynamic expression of PGANTs indicates a specific requirement of O-glycans in diverse tissues and at various stages of development. Indeed, labelling of Drosophila embryos with lectins and an antibody against the Tn-antigen has shown the presence of mucin-type O-glycans in most of the developing embryonic tissues [8-10]. In particular, O-glycans were predominantly found along the luminal and apical surfaces of epithelial tubes of the salivary glands, developing gut and the tracheal system [8].

The first evidence implicating mucin-type O-glycosylation in Drosophila development was demonstrated by the observation that one of the members of the PGANT family, pgant35A, is recessive lethal [131,137]. Subsequently, it was found that loss of pgant35A was associated with an altered tracheal tube morphology, accompanied by mislocalization of SJ proteins and a compromised paracellular barrier [138]. Loss of pgant35A also resulted in reduced levels of Crb and tracheal luminal components (the 2A12 antigen and O-glycans), suggesting a role for pgant35A in trafficking of apical and luminal components during tracheal development [138]. pgant35A is also expressed in the developing salivary glands and hindgut [136], but the irregular tube morphology and cell polarity defects seen in pgant35A mutants were restricted to the tracheal system. This could be due to functional redundancy among PGANT isoforms [138].

Specific roles of O-glycosylation in Drosophila have also been demonstrated for integrin mediated cell adhesion during wing development [139]. Loss of pgant3 results in
reduction of $O$-glycans along the basal surface of the larval wing imaginal discs, causing irregular adhesion of the two epithelial cell layers that will ultimately form the adult wing blade [139]. This aberrant adhesion was evident in localized blisters in the adult wing soon after eclosion. A combination of bioinformatics and in vitro glycosylation assays showed that PGANT3 glycosylates the integrin-binding ECM protein Tiggrin that is normally secreted into the basement membrane [139]. This was further confirmed by immunoprecipitation and genetic interaction experiments. In $pgant3$ mutants, reduced $O$-glycosylation of Tiggrin was observed, and it was proposed that $O$-glycans found on Tiggrin could affect some aspects of integrin-ECM adhesion and also could influence protein stability, secretion and binding interactions [139]. Additional roles of $pgant3$ in secretion were further demonstrated in a $Drosophila$ cell culture system, in which RNAi against $pgant3$ resulted in altered Golgi structure and reduced secretion of a reporter construct [140].

In a recent study, tissue-specific knockdown of multiple members of the PGANT family using RNAi identified $pgant4$, $pgant5$, $pgant7$ and $CG30463$ to be essential in various developing organs [141]. Loss of $pgant5$ was found to cause altered copper cell morphology (disorganized apical microvilli), reduced levels of $O$-glycans along the apical and luminal surfaces of the copper cells and defects in larval midgut acidification [141]. Copper cells are specialized cup-shaped cells found in the $Drosophila$ midgut and are responsible for gut acidification [142]. Although no target substrate for PGANT5 was identified, this study suggested the possibility that PGANT5 could be responsible for glycosylation of components essential for localization of ion transporters or proteins involved in organizing apical polarity in copper cells [141]. Together, these studies point to important roles for mucin-type $O$-glycosylation in $Drosophila$ development.
Results and Discussion

Paper I

Mucins are a large family of heavily $O$-glycosylated proteins and are the major components of the protective mucosal surfaces lining several vital organs of the body. Malfunction of this protective mucosal-barrier can lead to infections, acute or chronic inflammation and development of cancer [117,118]. Despite their importance in human pathologies, there is limited knowledge about the mechanisms regulating mucin expression and glycosylation. Moreover, it has been difficult to study mucins in relation to disease development, largely due to their physical and biochemical properties. One way to approach these questions is to address them in simpler invertebrate model systems, such as *Drosophila*, from which parallels to vertebrates can be drawn. The main aim of the study in Paper I was to identify mucin or mucin-like proteins in *Drosophila* and describe their expression pattern during development from embryo to adult.

*Identification of mucins and mucin-related proteins in Drosophila*

The PTS domains found in mucins tend to be poorly conserved, and there is no general consensus sequence defined to predict mucin-type $O$-glycosylation [143]. Identification of PTS domains using sequence similarity methods like BLAST is unreliable. However, statistical studies using experimentally verified $O$-glycosylation sites have led to development of sophisticated algorithms to predict mucin-type $O$-glycosylation [143-145].

To identify mucins in *Drosophila*, we devised a simple bioinformatic strategy that targets PTS repeats. To accomplish this, we developed a program called PTSPMiner that was used to find PTS domains in the predicted *Drosophila* proteome. PTSPMiner is developed in Java programming language and utilizes BioJava API. The first step in the program is implemented to calculate the total frequency of the amino acids Ser, Thr and Pro in a predicted protein, and the second step identifies the number of amino acid tandem repeats in the sequence. When applying cut-offs for Ser and Thr content $> 25\%$ and number of repeats $> 4$, forty-two proteins encoded by different genes were identified (Figure 6). PTSPMiner allows visualization of PTS repeats, by highlighting Pro, Thr and Ser in
different colours (Figure 7). By manual analysis, we found that nine of the proteins lacked PTS repeats and instead contained other types of repeats. These were excluded from further analysis. In addition, we included three mucin-like proteins that were identified through homology searches for mucin-associated domains. These were not picked by PTSPMiner because their Ser and Thr contents were below threshold (Figure 6).

In order to name the thirty-six identified proteins, we adopted a simple nomenclature: Proteins in which the PTS domain(s) constitute more than 30% of the protein length were termed Mucins (Muc), and proteins with PTS domains constituting less than 30% of the protein, or where the Ser and Thr-rich regions contained no Pro, were termed mucin-related proteins (Mur) (Figure 6). This nomenclature was followed by the cytological position of the encoding gene.

In our further analysis, we focused on fifteen mucins and eight mucin-related proteins (Figure 7). Of these, two mucins and two mucin-related proteins had no predicted signal sequence or transmembrane domains, which might be due to inaccurate prediction of the encoding genes. None of the identified mucins contain a transmembrane domain, nor did they harbour a vWD, SEA or CK domain found in human mucins. However, other conserved protein domains involved in protein interactions, such as the vWC domain, chitin binding Per-A domains and EGF-like domains were found both in the identified mucins and mucin-related proteins. Moreover, three of the identified mucins contain cysteines within their PTS domains, and this is also observed in the PTS domains of gel-forming mucins in *Xenopus tropicalis* [121].

Figure 6: Schematic workflow of the devised strategy for the identification of mucins in *Drosophila*.
A previous bioinformatic study to identify mucins lead to the development of two approaches that were implemented in two programs called PTSPRED and MPRED [145]. PTSPRED identifies regions in a protein sequence that shows high content of Ser, Thr and Pro, while MPRED uses a statistical model called the hidden Markov model to make probabilistic predictions of whether an amino acid sequence conforms to a mucin domain. PTSPRED and MPRED were used to identify mucin domains in several different species, including *Fugu rubripes* [145], chicken [146] and *Drosophila* [121]. PTSPMiner was able to identify most of the mucin-domain containing *Drosophila* proteins identified by PTSPRED and MPRED [121], with the exception of a few that either lack repetitive nature or had low Ser and Thr-content. Like PTSPRED, PTSPMiner is based on amino acid compositional bias, but it differs in some architectural elements such as GUI, programming language and platform independence. The GUI allows for easy manual inspection for repeat regions with different colouring schemes and convenient file input.
and output handling (Figure 7). PTSPMiner, being written in Java, makes it platform-independent, enabling it to run on multiple operating systems.

**Drosophila mucins are expressed at different stages of life cycle**

To gain further insight into possible functions of the identified mucins and mucin-related proteins, we analysed their expression pattern at different stages of development from embryo to adult. We performed reverse transcription PCR on RNA extracts from different developmental stages and on dissected organs from third instar larva, and found that several of the mucins and mucin-related proteins were dynamically expressed during the fly life cycle and showed tissue-specific expression. Three of the mucins were exclusively expressed either at embryonic stage (Muc30E), larval stage (Muc68D) or adult stage (Mur11Da).

In *Drosophila*, many epithelia are protected by an apical chitinous cuticle, such as the epidermis, the tracheal system and parts of the digestive system (foregut and hindgut). If *Drosophila* mucins were to have similar physiological functions as vertebrate mucins, their primary site of expression should be within cuticle-free organs. We found that a majority of the mucins and mucin-related proteins were expressed in cuticle-free organs, including the salivary glands, digestive tract and Malpighian tubules of third instar larvae. The midgut is protected on the luminal side by a non-cellular apical matrix known as the Peritrophic matrix (PM) [147]. The PM plays the role of a physical barrier and consists of a scaffold of chitin fibres embedded with glycosylated and most often chitin-binding proteins (Peritrophins) [147], and is regarded to be functionally similar to vertebrate mucosa [148]. The mucins and mucin-related proteins detected in the digestive tract could potentially be components of the PM. Some of these have similar domains to Invertebrate Intestinal Mucin (IIM), a PM protein of *Trichoplusia ni* [149].

In addition to the digestive tract, the salivary glands showed prominent mucin expression. Two of the mucins expressed in salivary glands, Muc25B/Sgs1 and Muc68Cb/Sgs3, were previously reported to belong to the salivary gland secretion (Sgs) family of proteins. These are secreted towards the end of third instar larva to produce a sticky secretion by which the larvae attach themselves to a solid surface prior to pupa formation [150]. The other mucins that are expressed in the salivary glands might be glue proteins or have
Mucin-like proteins in Drosophila development

protective roles, and further characterization has to be carried out to address their function.

**Drosophila mucins are expressed in developing epithelial organs**

An interesting finding was that many of the mucins and mucin-related proteins were expressed during embryogenesis. In order to identify their expression pattern, we performed whole mount RNA *in situ* hybridization. This revealed that mucins and mucin-related proteins were expressed in the developing salivary glands, midgut, foregut, hindgut and trachea, as well as in the proventriculus and epidermis during late stages of embryogenesis. One of the mucins, Muc30E, was exclusively expressed in the extraembryonic amniosera and its expression was ceased after dorsal closure. This selective expression indicates a specific role for this mucin during development, and further characterization is required to understand its function. The observed expression patterns of mucins and mucin-related proteins parallel the previously reported expression of PGANTs [136] and detection of O-glycans in the lumen and along the apical surface of developing organs [8]. It is therefore possible that the identified mucins and mucin-related proteins represent novel components of the apical O-glycan-rich matrices seen in developing epithelial organs. During tracheal development, a transient chitin-based luminal matrix is required for shaping the tube [151,152], and earlier studies have shown that one of the mucin-related proteins, Mur96/Tnc, is present in the lumen and along the apical surface of the developing trachea, foregut and hindgut [153]. It is an exciting possibility that glycosylated proteins like mucins could form luminal matrices needed to shape epithelial organs.
Paper II

It was intriguing that many of the mucins and mucin-related proteins were expressed during embryogenesis. To study a possible involvement of mucin in embryonic epithelial organ development, we choose to analyse the function of Mur96/Tnc in epithelial tubes like the foregut, hindgut and trachea.

Tenectin is an intraluminal protein required for diameter expansion of the hindgut

Tnc is a large mucin-related protein consisting of 2788 amino acids. The deduced protein sequence reveals that Tnc is secreted and harbours two PTS domains flanked by five vWC domains. The domain organization of Tnc resembles that of secreted gel-forming mucins, in which dense O-glycosylated PTS domains are separated by cysteine-rich von Willebrand factor domains that mediate polymerization [119]. Confocal imaging of wild type embryos labelled with antisera against Tnc showed Tnc to be secreted at the apical surface in the developing hindgut, foregut and tracheal dorsal trunks, consistent with the previously reported expression of Tnc in these tissues [153]. In addition, Tnc was present in the proventriculus, salivary gland ducts and in the lumen of dorsal vessel. In all tissues analysed, Tnc localized to the apical surface and spanned the entire lumen, thus behaving like a secreted intraluminal component.

To address a possible function of Tnc in epithelial organ development, tnc mutants were generated and the loss of function allele called tnc13c was analysed. Labelling of tnc mutant embryos with the apical marker Crb to visualize the organ lumens revealed that tnc mutants have an unusually narrow hindgut lumen. At the end of embryonic stage 13, the hindgut layout is established with the partitioning of the hindgut into distinct compartments (small intestine (Si), large intestine (Li) and rectum) [154]. From stage 14 onwards, the hindgut tube (Si and Li) grows by tube elongation and diameter expansion. Morphometric analysis of the hindgut (Si and Li) in wild type and tnc mutants at stage 14 and stage 16 indicated that the volume of the hindgut lumen in tnc mutants was at least 50% less than in the wild type, while tube length was slightly longer than normal. The narrow hindgut lumen in tnc mutants was accompanied by detectable cellular changes in the epithelium. Staining of wild type and tnc mutant embryos with DE-cad to highlight the apical cell circumference showed that tnc mutants have reduced apical cell domains,
and this was particularly evident in Si. Examination of the apical cell circumference in Li of tnc mutants further revealed aberrations in cell arrangement, coherent with the narrow diameter. A reduction in outer tube diameter was also observed in tnc mutants. These observations show that Tnc is required for hindgut tube dilations, both at the apical and basal surface, by mechanisms associated with changes in cell shape and slight cell rearrangements. We could not find that epithelial polarity or patterning of the hindgut tube was affected in the tnc mutants.

**Tnc drives hindgut expansion in a dose-dependent manner**

Expression analysis of Tnc in the hindgut showed that Tnc gradually accumulates in the lumen from stage 13 onwards, and by stage 16, Tnc was abundant in the lumen with stronger expression in Si than in Li. The period of accumulation of Tnc in the lumen, together with the severe reduction of Si diameter compared to Li and the stronger expression of tnc in Si versus Li suggested that the degree of diameter expansion might be connected to the levels of Tnc in the lumen. To test this possibility, tnc was overexpressed in the hindgut at varying levels using different GAL4-driver lines. Over-expression of tnc caused excessive dilation of the hindgut lumen, and over-expression of tnc at higher levels in Si than in Li resulted in an elevated increase in Si diameter compared to Li. The excessive lumen dilation associated with Tnc over-expression was accompanied by an increase in outer tube diameter and enlarged apical circumferences of the hindgut epithelial cell. These observations contrasted the defects seen in the hindgut of tnc mutants and indicate that Tnc drives hindgut lumen dilation depending on the pattern and levels of expression. Interestingly, ectopic tnc expression in other epithelial organs also lead to excessive tube dilation and increase in apical cell circumference, similar to the effects of Tnc over expression in the hindgut.

When Tnc was over-expressed in only the dorsal compartment of Li, this resulted in enlarged apical cell circumferences in both the dorsal and ventral compartments, suggesting that Tnc acts non-cell autonomously. Moreover, when tnc was expressed in the epidermis, to test its effect on epithelial organization in a non-luminal context, no anomalies or changes in cell shape were seen in epidermis. Together these results argue that Tnc promotes tube dilation by exerting an internal luminal pressure.
**Tnc is a component of O-glycosylated matrix in the hindgut lumen**

Tnc carries two large PTS domains, which could potentially serve as substrates for O-glycosylation. When protein extracts from embryos and larvae were probed on western blots, Tnc was detected as high molecular weight species substantially larger than the predicted molecular mass of 290 kDa. This migration pattern of Tnc would be consistent with Tnc carrying post-translational modifications. To test if Tnc might be glycosylated, embryonic extracts were treated with deglycosylation enzymes. Treatment with O-glycanase or with O-glycanase together with three other enzymes, Sialidase, β(1-4) Galactosidase and β-N-Acetylglucosaminidase, caused slightly faster migration of Tnc, while N-glycanase treatment did not affect the migration, indicating that Tnc carries mucin-type O-glycans. If Tnc carries mucin-type O-glycans, it should contribute to detectable O-glycans in the hindgut lumen. To address this, both wild type and tnc mutant embryos were labelled with an antibody that detects the Tn-antigen. Wild type embryos showed strong staining along the luminal surface and in intraluminal compartment of the hindgut, whereas tnc mutants displayed reduced intraluminal Tn-staining. Staining with *Vicia villosa* lectin (VVA) that also recognizes the Tn-antigen showed a similar reduced intraluminal VVA staining in the tnc mutant hindgut. Both Tnc and Tn-staining showed punctate, and sometimes discontinuous, staining in the hindgut lumen with formaldehyde fixation. It has earlier been demonstrated that ethanol-based fixation methods best preserve the texture of glycan-rich matrices [155], and when embryos were fixed with the ethanol-based Clark’s fixative, Tnc appeared as a striated structure that filled the entire hindgut lumen. Tnc therefore appears to be an important carrier of mucin-type O-glycans, forming a glycan-rich matrix in the hindgut lumen.

**Model for Tnc-mediated tube dilation**

Our results indicated that Tnc has biochemical characteristics similar to secreted gel-forming mucins and forms a striated structure that fills the hindgut lumen during tube expansion. Overexpression of Tnc showed that Tnc is able to drive tube dilation in dose-dependent manner, and that this lumen dilation is associated with an increase in apical cell circumferences and an increase in both inner and outer tube diameter. Based on these observations, a model for Tnc can be presented in which Tnc drives tube dilation by generating a luminal pressure upon its secretion into the lumen, and thus the extent of
lumen volume expansion would directly relate to the amount of Tnc in the lumen. This represents a new mechanism of shaping epithelial tubes.

The proposed function of Tnc would be analogous to that of a luminal hydrostatic pressure, which can cause an increase in lumen volume [156,157]. This is believed to occur in the initial inflation of the brain ventricle lumen [158,159] and Kupffer’s vesicle in zebrafish [160]. However, Tnc can also cause local tube dilation, possibly because Tnc is relatively immobile in the lumen and forms an expanding matrix at the site of secretion. This differs from the effect of hydrostatic pressure, where the force acts uniformly on the tube wall.

The function of Tnc differs from that of luminal chitin matrix in trachea. The chitinous matrix does not drive lumen dilation, but appears to provide a scaffold for uniform diameter expansion [151,152], while Tnc is needed to increase the luminal volume. Tnc function is also distinct from that Eye Shut in the fly retina [161] and CD34 sialomucin Podocalyxin [162] that are required for initiating lumen formation, as Tnc is not required for opening a lumen cavity, but for subsequent diameter expansion. It is intriguing how a luminal pressure generated by Tnc is perceived by the epithelium to cause epithelial remodelling. Recent studies examining tube elongation in salivary glands have shown that apical membrane dynamics and membrane-cytoskeleton interactions are critical for facilitating cell shape changes [163,164]. Another recent study proposed that epithelial relaxation, mediated by the activity of myosin phosphate facilitates inflation of the brain ventricular lumen during zebrafish development [165]. Further investigation of the Drosophila hindgut might provide insights into the molecular basis of tube wall plasticity in response to increase in luminal pressure.
Conclusions

In this thesis work potential Drosophila mucins and mucin-related proteins were identified in a genome wide screen. Drosophila mucins lack the conserved domains associated with vertebrate mucins such SEA, vWD and CK domains, but have potential protein- and chitin-interaction domains that could mediate multimerization and apical matrix formation.

Expression analysis of Drosophila mucin-like proteins revealed two interesting findings. First, many mucins were expressed in the digestive tract, suggesting that they contribute to a protective apical matrix analogous to vertebrate mucosa. Second, several of the mucins and mucin-related proteins showed dynamic and tissue-specific expression during embryogenesis, implicating a role for mucins in organ development. Characterisation of one of the mucin-related proteins, Tnc, shows that it fills the lumen of developing hindgut and drives tube diameter expansion in a dose-dependent manner. We believe that this represents a new mechanism of shaping epithelial tubes, where a glycan-rich luminal matrix generates a mechanical force on the surrounding epithelium.

Although conserved domains associated with vertebrate mucins are absent in Drosophila mucins, many are found in other Drosophila proteins. One of these is called Mesh and contains several protein domains found in human MUC4. As a first step to understand the possible roles of these domains, we have shown that Mesh is important for epithelial integrity.

Further studies of the identified mucins and mucin-related proteins should enrich our understanding of mucin biology, both in physiological settings and during development, and possibly reveal new mechanisms of mucin function.
Acknowledgments

I would like to take this opportunity to thank everybody who has contributed directly or indirectly to this thesis work, and without their kind help and support this would not have been possible.

Iris Härd for introducing me to the world of mucins and giving me this opportunity. You have given me enormous freedom to explore and develop my scientific skills. Discussing scientific and non-scientific matters with you has always been fun and interesting. Thank you so much for your unlimited support throughout my studies and your motivation, mentorship and friendship is invaluable to me.

Anne Uv for her unlimited support and taking me into her lab. Throughout my time in lab, your energy and kind nature have created relaxed and very inspiring scientific environment. I admire your commitment, and your endless discussion about science has always been interesting and fun. Above all you have been such great friend, mentor and sport. I thoroughly enjoyed my time in lab with you and all this made my PhD a nice experience.

I thank you both for your invaluable scientific guidance and constructive comments on my scientific writing.

A special thanks to Torleif Härd for the PhD position and support.

Thanks to Gunnar Hansson for being my co-supervisor and his inputs. Thanks to all the group members for letting me borrow reagents and equipment.

I would like to thank all the past and present members of the A.Uv and I. Härd lab.

Thank you Simin for sharing your extensive scientific experience, kind nature and friendship. It was lot of fun sharing lab bench with you.

Kate, thank so much for proofreading my thesis and all the laughs. I enjoy your sense of humor and conversations about gretchen and entities that grind our gears.
Hamid (hello jaamess!!!), It was fun working with you and being part of akbar brothers tea club. Thank you so much for your friendship and all the help and good food. Good luck with your future ventures!!

Anna Tonning, for always smiling and helping me out during my early lab days, even though you were busy preparing for your thesis. Erika, thank you for the good time in lab and sharing jokes and laughs about the akbar brothers. Tina, thanks for the all talks and discussions during our early days in lab.

Fredrik and Erik (the cool dude), thanks for the good time and fun being part of the akbar brothers. I enjoyed our lunch conversations and thanks for the fishing advice!! Good luck in your future endeavors. Moin, Banaz, Jonathan and Mengistu thanks for your time and help and good luck with your studies. Jonathan, thank you for introducing me to one of my favorite pizzeria.

Thanks to Levent and his group members Sashi and Srujuna for letting me use their equipment. Sashi and Srujuna, thank you both for your friendship. Sashi thanks for the good time and for being my fishing buddy. I had lot fun with you during our fishing trips and you are good sport, keep it up. Good luck with your studies!!!

Cinna, thank you so much for the laughs, lunches to Hani and for sharing your interest in electronics. Good luck with your studies!!!

Arun, you are truly a nice guy!! Thank you for the good time and interesting conversations and sharing your views. Keep up the good nature and I wish you all the best for your studies.

Ali (Gotiya sahib), thank you so much for all the jokes and help. It was really great hanging out with you and thank you for being such a good friend. Good luck with your new job!!!

Bernard, it was great having you in the lab and thank you so much for all the interesting scientific discussions and collaboration. I have never seen someone with such a cool moves on the dance floor.
Michaela, Gosa and Mehrnaz from Homa’s group and Katarina, thank you all for the good time in lab and interesting lab discussions.

Thanks to Chandra and all his group members (Tanmoy, Gaurav, Kankadeb, Sanhita, Arif, Shantilal, Mathieu) for all the laughs and discussions. Erik, Babak and Zhiyuan thanks for interesting lunch conversations and discussing cool stuff!!!

Thanks to Julia and Maria at the Cellular imaging core facility for such a fantastic course and answering all my annoying questions and help.

Bengt, Yvonne and Canita thank you so much for such an interesting EM course.

Thanks to the administration at Biomedicine, Carina Ejdeholm, Ulrica Lilja, Andrea Frateschi, Carina Petersson for the all their help.

I would thank all my friends in India, Vipul, Aman, Parvez, Tabrez, Hamed, Bushra, Ruby, Razi, Masood, Feroze, Sainath, Kiran, Murli, Arif Bhai, Rafi, Salman and Asjad. You all have been great and we had a good time during our schooling, college and university days. Good luck with your jobs and research.

To my parents for their endless love and support, Abba and Ammi you’re the best and thank you so much for your prayers and motivation. I miss you Dada, Dadi and Nana, Nani.

To my brother, Jeelani, you have been such a great inspiration and role model for me. Thank you so much for introducing me to music, books and all the cool stuff a younger brother could wish for. Thanks a lot bhaiyya!!, I couldn’t ask for anything more!!

My brother, Rabbani, you are such great sport; we had so much fun during our schooling. Thank you for supporting me in whatever I plan to do. Salma appa and Farha, my adorable sisters!!! Thank you for your love, caring and support and teaching me how to read and write urdu. Your Biryaniis are delicious and thank you so much for teaching me how to cook!!!!!
Thanks to my brother-in-law Mohiuddin and Malik and Sister-in-law Ruth and Sara for your caring, affection and motivation. A Special thanks to Sunita’s parents, you have been amazing and thank you so much for your love and support.

My two lovely niece, Gousia, Nazia and newphews Rafic, Toufeeq, Levi, Faiz and Ryan, I love you all and promise to have loads of fun when I visit you.

My beloved Sunita, for her endless love, support and encouragement. I admire your great patience and sharing all the moments of happiness and difficulties. Thank you for sharing with me your passion for science. I love you!!!
Mucin-like proteins in Drosophila development

References


Mucin-like proteins in Drosophila development


Mucin-like proteins in Drosophila development


Mucin-like proteins in Drosophila development


Mucin-like proteins in Drosophila development


146. Lang T, Hansson GC, Samuelsson T (2006) An inventory of mucin genes in the chicken genome shows that the mucin domain of Muc13 is encoded by multiple exons and that ovomucin is part of a locus of related gel-forming mucins. BMC Genomics 7: 197.


