Unzipping the barriers: Determining the role of the peritrophic matrix in *Trypanosoma brucei* migration through the midgut of *Glossina morsitans morsitans*

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by:

Clair Rose

April 2016

ABSTRACT

Unzipping the barriers: Determining the role of the peritrophic matrix in *Trypanosoma brucei* migration through the midgut of *Glossina morsitans morsitans*

By

Clair Rose

Tsetse flies serve as biological vectors for several species of African trypanosomes. These parasites undergo a life-cycle stage in both mammals and in the insect. Within the fly, it is proposed that for successful survival, proliferation and establishment of a midgut infection, trypanosomes must cross the tsetse fly peritrophic matrix (PM). Although having been well documented, the mechanism(s) of how trypanosomes are able to cross the PM is not well understood and the crossing event has never been visualised.

The peritrophic matrix is an acellular secretion that lines the guts of most insects and is mainly composed of chitin and glycoproteins. The functions of the PM involve facilitating digestion, and epithelial cell protection against ingested pathogens, toxins and against mechanical damage. In order to better understand the nature of the tsetse PM, dissected PMs from lab-reared *Glossina morsitans morsitans* were solubilised under stringent conditions and their protein composition determined using a mass spectrometry-based proteomics approach. It was found that the tsetse PM is comprised of just short of 300 proteins, including several proteins from the tsetse endosymbiont, *Sodalis glossinidius*. Tsetse PM proteins were classified into functional groups as peritrophins, enzymes, immune related proteins and other proteins. This study also includes the first report of three tsetse PM peritrophins (GMOY002708, GMOY007191 and GMOY011810), increasing the known repertoire of peritrophins from two to five. Peritrophins were then further classified into four main groups comprising simple, binary, complex and repetitive based on their structural organisation. Comparisons of these peritrophins in *Glossina morsitans morsitans* were carried out against orthologous proteins of several major *Glossina* vectors and two comparative non blood-feeders, *Musca domestica* and *Drosophila melanogaster*. The vast number of peritrophins in *M. domestica* and *D. melanogaster* compared to those in *Glossina*, and given the feeding habits of these flies, suggested that the PM peritrophins have a role in the protection of the midgut against ingested pathogens. Additionally, peritrophin silencing by RNAi provided information regarding the expression and regulation of peritrophins.

Multiple microscopy techniques showed the presence of trypanosomes within the layers of the PM which has previously been shown before. However, there was no evidence of

trypanosomes crossing the mature PM in the anterior midgut as judged by tissue tomography analysis and 3D reconstruction. Subsequent analysis of the tsetse proventriculus (PV; site of PM secretion) revealed that trypanosomes could be seen in this tissue both between and either side of the PM from as soon as 5 days after receiving a trypanosome-infected bloodmeal. This suggests that the immature PM within the PV provides a probable point of entry for parasites after differentiation into the procyclic life stage. Moreover, the opportunity for breaching of the PM must occur quite quickly after bloodmeal ingestion as refractory flies had already cleared the infection by 5 days post infection. Overall, the results from this thesis suggest that the tsetse PM: a) is structurally composed of a family of peritrophins that differ in complexity and confer resistance to degradation by trypanosomes in the midgut b) functions as a way to clear parasite infections by engulfing them before elimination in the hindgut; and c) the lifecycle of *T. brucei* may occur entirely (or in parallel with a developmental cycle in the ectoperitrophic space) within the PV of the fly where they are proposed to complete maturation before migration to the salivary glands.

DECLARATION

I hereby certify that this dissertation constitutes my own product, that where the language of others is set forth, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions or writings of another.

I declare that the dissertation describes original work that has not previously been presented for the award of any other degree of any institution.

Signed,

…….

Dedications

This thesis is dedicated to my two best mates

To my mum, Carol

1948-2013

R.I.P my guardian angel

Hope I have made you proud and continue to do so

To my Ellie

You make me proud every day

"The minute you think of giving up, think of the reason you held on for so long"

- Author unknown

Acknowledgements

To my supervisor Alvaro Acosta-Serrano, you had me at ‘sexy flies’! Thank you for taking a chance on me all those years ago when neither of us could understand each other. Now that my Scouse has been honed and I can understand your Spanglish, I wanted to say how grateful I am to you for everything you have done throughout my Ph.D., both professionally and personally.

To Mike Lehane, for your invaluable T.E.M insights and pearls of wisdom. You showed me we don’t all have to be born with silver spoons in our mouths to succeed.

To Lee and Dee, the terrible two! For the great discussions and advice but most of all for the great laughs. If the Ph.D. falls through at least I know I will have a secondary career to fall back on, even if it’s only you two in the audience laughing at the tales of woe that is my life.

To the greatest lab group there ever was past and present, for the good times, the laughs, the late nights and the lunches. For keeping me sane and grounded through some of the toughest times of my life and for making my time so enjoyable.

To the people who helped make completion of this project possible, Naomi, Ben, Carla, Dan, Ali and Marco.

To the LSTM for funding my studentship

And to Ellie. For being you.

Table of contents

List of Abbreviations

aa	Amino Acid
AAT	Animal African Trypanosomiasis
ADT	Asymmetrical Dividing Trypanosome
Agr	Antigen 5 related
AT	African Trypanosomiasis
BLAST	Basic Local Alignment Search Tool
BSF	Blood Stream Form
CBD	Chitin Binding Domain
cDNA	Complementary DNA
CI	Cuticular Intima
CID	Collision-Induced Dissociation
CJ	Cell Junction
Cp	Crop
CPAP	Cuticular Proteins Analogous to Peritrophins
CLSM	Confocal Laser Scanning Microscopy
CTLs	C-type Lectins
dsRNA	double-stranded RNA
dsRNase	double-stranded RNase
DABCO	1, 4-diazabicyclo-[2, 2, 2]-octane
DAPI	4', 6-diamidino-2-phenylindole
DTT	Dithiothreitol
EGF	Epidermal Growth Factor
ES	Ectoperitrophic Space
EST	Expressed Sequence Tag
FB	Fat Body
FG	Foregut
FM	Flight Muscle
GA	Glutaraldehyde
GAGs	Glycosaminoglycans
gDNA	Genomic DNA
GFP	Green Fluorescent Protein
GlcNAc	Poly β-(1, 4)-N-acetyl-D-glucosamine
GmmPro1	Proventriculin 1
GmmPro2	Proventriculin 2

GO	Gene Ontology
GPI	Glycophosphatidylinositol
HAT	Human African Trypanosomiasis
HG	Hindgut
HPE	Hours Post Eclosion
HRM	Haem-regulatory motifs
IGGI	International Glossina genome Initiative
IIM	Insect Intestinal Mucin
KD	Knockdown
LB	Lysogeny Broth
LC-MS/MS	Liquid chromatography tandem-mass spectrometry
LDL	Low-density lipoprotein
LN2	Liquid Nitrogen
MA	Microtubular axoneme
mAb	Monoclonal Antibody
MD	Mucin Domain
MG	Midgut
MV	Microvilli
NCBI	National Centre for Biotechnology Information
NFW	Nuclease Free Water
NGO	Non-Government Organisations
NTD	Neglected Tropical Disease
OMP	Outer Membrane Protein
Ov	Ovarioles
PAD	Peritrophin-A Domain
PAGE	Polyacrylamide Gel Electrophoresis
PBD	Peritrophin-B Domain
PBS	Phosphate Buffered Saline
PCD	Peritrophin-C Domain
PCF	Procyclic Form
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PE	Post Eclosion
PFA	Paraformaldehyde

PFR	Paraflagellar rod
PLP	Peritrophin-Like-Protein
PM	Peritrophic Matrix
PTM	Post-translational modification
PV	Proventriculus
PVDF	Polyvinylidene Difluoride
qPCR	Quantitative real-time PCR
RNAi	RNA interference
RPLC	Reverse-Phase Liquid Chromatography
ROS	Reactive Oxygen Species
SBSEM	Serial Block-face Scanning Electron Microscope
SDS	Sodium Dodecyl Sulphate
SEM	Scanning Electron Microscope
SERPIN	Serine Protease Inhibitor
SG	Salivary Glands
SM1	Salivary gland and Midgut peptide 1
Sub	Subpellicular corset
Tag5	Tsetse Antigen 5
TCH	Thiocarbohyrazide
TEM	Transmission Electron Microscope
TFA	Trifluoroacetic Acid
TFMS	Trifluorosulfuric Acid
Ts	Testes
VSG	Variable Surface Glycoprotein
WGA	Wheat Germ Agglutinin
WHO	World Health Organisation
WT	Wild Type

List of tables

List of Figures

Chapter 1. Introduction

1.1 General Introduction

Worldwide, there are 17 neglected tropical diseases (NTD's) resulting from four different causative pathogens; viruses, protozoa, bacteria and helminthes [1]. A number of these are transmitted by insect vectors affecting over a billion people per year, mainly in tropical and sub-tropical regions. Vector borne diseases are notoriously difficult to both manage and monitor due to the unsustainability of control techniques and the difficulties with epidemiological monitoring. Unlike many other vector borne NTD's, African trypanosomiasis (AT) occurs only where the vector is found and is restricted to 36 sub-Saharan African countries [2]. Transmitted by tsetse flies and caused by protozoan parasites, the African trypanosomiases are of constant impediment to the health and economic development in endemic regions of Sub-Saharan Africa [3]. Although the numbers of reported cases have dropped to below 10,000 per year since 2012 due to on-going vector control efforts, Human African Trypanosomiasis (HAT) still impacts approximately 70 million people over 1.55 million km^2 [4]. Indeed, this may be a gross under-estimation as many countries do not report cases due to countless people and communities having no access to healthcare. In addition, it is estimated that less than 1% of the 70 million people at risk are presently under surveillance for HAT. Currently, the pharmaceutical treatments available for humans are difficult to implement, have limited efficacy and are generating concerns about emerging parasite resistance to frontline treatments. In addition, livestock treatment for Animal African Trypanosomiasis (AAT) is rarely implemented and insufficiently maintained which creates a huge burden on the livelihood of Africa. As vaccine development towards trypanosomes is unfeasible due to the parasites variable surface glycoprotein (VSG) coat which gives it the ability to evade the mammalian immune system, new disease control ideas have shifted to investigating the vector-parasite interface rather than targeting parasite interactions within the mammalian host.

1.2 Importance of trypanosomiasis

The trypanosomiases consist of a group of significant animal and human diseases caused by parasitic, flagellated protozoans of the genus *Trypanosoma*. HAT, or sleeping sickness, causes a number of diseases in humans and accounts for approximately 9,000 deaths per year [5], with 30,000 people estimated to be infected. Although seemingly insignificant in comparison to other major infections, the geographically restricted outbreaks along the tsetse belt and threat of epidemics of the disease continue to have a devastating impact on the socioeconomic development of affected communities [6]. By the 1960's, African sleeping sickness was largely under control, with less than 5,000 cases reported across the

entire continent, due to effective vector management, mass screening, early diagnosis and swift treatment. However, a resurgence of the disease has occurred and recent epidemics have been strongly associated with political and civil unrest resulting in mass movement of communities into largely unpopulated areas [7]. This has had a significant impact on healthcare implementation, especially in the most remote areas, causing a huge decrease in population screening and treatment. The 3 major epidemics over the last century (1896-1906, 1920, 1970-1998) were in part controlled by groups of mobile teams implementing screening measures and mass drug treatment, after which surveillance was relaxed and there was a surge in cases. However, valiant efforts from WHO, nongovernmental organisations (NGO's) and donors responding to complex humanitarian emergencies have helped decrease cases of HAT by 73% between 2000 and 2012 and, as such, WHO has declared HAT as a target for elimination as a public health problem by 2020. Nonetheless, HAT remains a huge problem in Africa contributing to the major poverty and hardships faced by these communities.

Sleeping sickness is fatal if left untreated with progressive deterioration of mental ability until coma and eventually, death [8]**.** HAT is caused by *Trypanosoma brucei,* which exists as two morphologically identical subspecies; *T. b. gambiense* and *T. b. rhodesiense.* These parasites cause sleeping sickness in different parts of Africa (Fig.1.1) and lead to different clinical manifestations of the disease. *T. b. gambiense*, found in Central and West Africa, causes a chronic condition with progressive decline in health and is mainly human-human transmissible, whereas *T. b rhodesiense*, found in Southern and Eastern Africa, causes an acute infection in humans due to its higher rate of virulence [9] and is largely a zoonotic disease with multiple animal reservoirs. Uganda is the only country where both forms of the parasite are present [10], and although there have been no reports of people having being infected with both sub-species to date, the threat of these two species becoming sympatric is imminent due to *T. b rhodesiense* infected cattle migrating to Northern Uganda where at present only *T. b. gambiense* can be found. This situation is further exasperated by the buying and selling of infected livestock which are not always screened for parasites. African trypanosomes, mainly *T. congolense, T. vivax* and *T. b. brucei*, also cause disease in animals accounting for 3 million deaths of cattle every year [11]. AAT, also known as Nagana, are responsible for the majority of low meat production in Sub-Saharan Africa, hindering the progression of cattle farming over ten million square kilometres [12]. They also act as a reservoir for human infections, and there is a potential for zoonoses to occur where trypanosomes formerly non-infective to humans have the capacity to become fully transmissible between the two species.

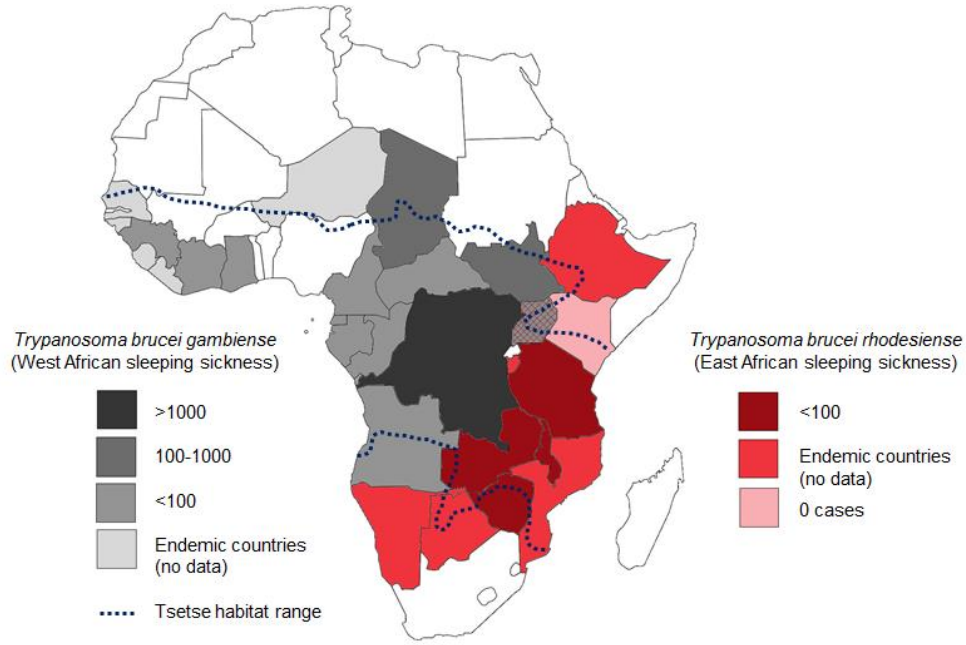


Figure 1.1. ***Distribution map of the two species of Trypanosomes (Trypanosoma brucei and T. b. rhodesiense) which cause HAT in Sub-Saharan Africa.*** *Colour intensities show the number of cases reported from each country per annum. Information taken from WHO for 2013. Created using Photoshop.*

T. b. gambiense and *T. b. rhodesiense* are transmitted to humans primarily via the bite of tsetse flies of the *Glossinidae* family, which act as biological vectors, although inadvertent transmission may also occur; for example through contaminated blood, needle pricks, or, rarely, through the placenta (congenital transmission). All species of *Glossina* can potentially transmit trypanosomes, although it is the species *G. morsitans, G. pallidipes, G. fuscipes fuscipes* and *G. palpalis* that are the most significant in terms of HAT and AAT transmission. Unlike other invertebrate vectors where usually only the female blood-feed such as those in the *Culicidae* and *Psychodidae* families, *Glossina spp.* are obligate heamatophagous insects, thus both sexes can be vectors of trypanosomiasis. Now that the annotation of the tsetse (*G. m. morsitans*) genome has been completed and other species genome annotation under way, we now have a comprehensive view of tsetse biology. Understanding how trypanosomes develop within the insect vector can lead to transmission blocking vaccines that can work alongside vector control methods and ultimately eliminate the disease from endemic areas.

1.3 Tsetse fly biology

The tsetse fly life-cycle is unusual for that of Dipterans, as females only need to mate once and are larviparous; i.e. they lay a single larva one at a time, as opposed to several thousand eggs. The first progeny is deposited 20–22 days post-adult eclosion, and If there is a plentiful food supply, larval deposition can be approximately every 9-12 days with the average female producing 8-12 single larva's each which is an unusually slow reproduction rate. After mating, the females store sperm in their spermathecae, and this sperm is released slowly over her lifetime in order to fertilise each egg. Each fertilised egg develops into a larva in the female's uterus where it is nourished by milk glands, with milk not dissimilar to human breast milk [13], until it reaches the third instar lava and is ready to be deposited. After deposition, the larva burrows into the ground where it develops for approximately 4-5 weeks depending on environmental factors. Lastly, the adult fly emerges, forcing its way up through the ground to the surface. The newly emerged adult will usually take its first blood meal after 24 hours and is now capable of becoming infected with trypanosomes from an infected host.

1.4 Trypanosome biology

African trypanosomes are unique within the kinetoplastids as they all express a variable surface glycoprotein (VSG) coat, which allows the parasite to avoid the mammalian immune system. This coat, of approximately 1 x 10^7 molecules [14], [15], poses a hindrance to vaccine development due to the high rate of antigenic variation which causes the expression of a different VSG coat during infection and allows successful evasion of the immune system. Although there are effective drugs available, the potentially dangerous side effects posed by these drugs and the parasite's increasing resistance to them make them unsuitable for use on a long-term basis [16], [17]. This, therefore, paves the way for alternative and fresh insights into the treatment of African trypanosomiasis and vector control. One such initiative involves preventing trypanosomes from developing within the tsetse midgut.

1.5 Development of *T. brucei* in the tsetse vector

All African trypanosomes belong to the clade of salivarian trypanosomes which are exclusively transmitted through tsetse bites. The life-cycle of trypanosomes is extremely complex as the digenic parasites are obliged to undergo different biochemical, morphological and physiological changes in each host. Although all species require

development within the tsetse vector, their route of migration and point of transmission can differ. For example, *T. vivax* develops within the mouthpart of tsetse and requires no further migration through the insect whereas *T. congolense* undergoes a series of directional migrations and developments through the fly until the parasites reach the mouthparts to become fully transmissible [18]. The only species capable of infecting humans are from the *T. brucei* complex. Consisting of three sub-species; *T. brucei brucei, T. b. gambiense* and *T. b rhodesiense*, only the latter two subpopulations have evolved the capability to infect humans whilst the former causes only animal infections. However, due to the extremely close phylogenetic relationship between all three sub-species and the sharing of fundamental features with the human infectious forms, *T. b. brucei* is often used as an experimental model for human infections and so an understanding of this parasites lifecycle is crucial. Here the emphasis is on the development within the fly as the parasites must undergo a complex ordered series of direction migrations and transitions in order to complete maturation (Fig. 1.2).

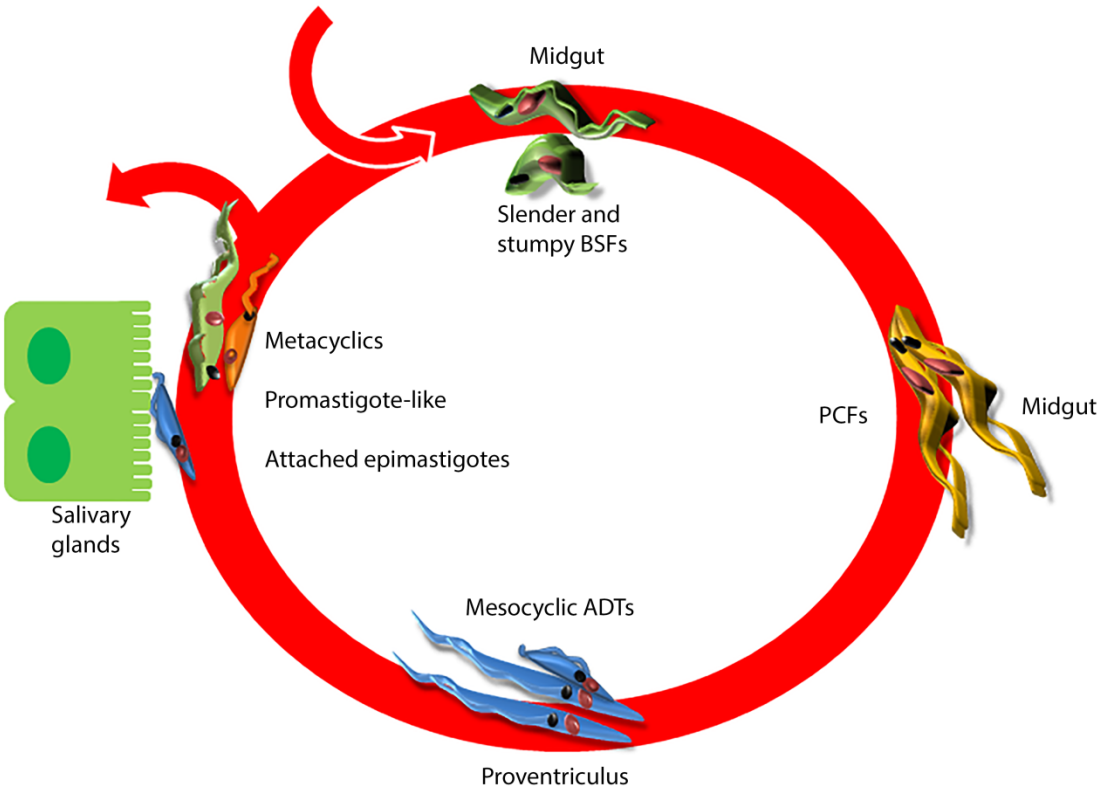


Figure 1.2. Schematic representing the lifecycle of T. brucei spp. in the tsetse. *When an infected bloodmeal is taken up, trypanosomes must undergo a series of migrations and developments in various tsetse tissues in order to complete their lifecycle and become transmissible to a new host.*

Multiple theories have existed regarding the possible migratory route trypanosomes have in the midgut of the fly (discussed further in Chapter 5), however, the following life-cycle is generally accepted. Upon feeding on an infective mammalian host, tsetse ingest “stumpy” bloodstream forms (BSF) with the blood meal. Within 24 hours post-infection of the fly, the parasites quickly transform into “slender” procyclics (PCF) within the midgut, lose their VSG coat and instead express procyclin molecules which enable them to survive in the hostile gut environment [19, 20]. Here PCFs resume cell division and undergo binary fission allowing establishment of a midgut infection. After 5-8 days *T. b. brucei* PCFs are able to breach the peritrophic matrix (PM) and invade the ectoperitrophic space (ES), where they shed the GPI anchored procyclin coats and differentiate into procyclic trypomastigote forms. This is an essential part of the parasite lifecycle as they are unable to survive in the midgut where they would be subjected to the tsetse digestive enzymes, hydrolases, reactive oxygen species, immune peptides and trypanolytic serum complement. ES invasion is also important for parasite replication enabling parasite proliferation. After 3-9 days in the ES, they then migrate through the proventriculus (possibly by crossing the PM a second time) where mesocyclic asymmetric dividing trypanosomes (ADTs) are present resulting in the generation of one long and one short epimastigote. The short epimastigotes continue migrating to the salivary glands where they adhere to the epithelial cells and undergo further asymmetric division to generate infectious metacyclic trypomastigotes which gain a variable antigen coat and are passed into the mammalian host upon the vector’s next bloodmeal [21-23]. Since this directional migration requires parasites to be in close contact with insect tissue, it has recently been proposed that observations of procyclic trypanosomes exhibiting social motility (SoMo) in vitro may be reflected in migration through the tsetse [24]. Moreover, it has been shown that only early PCFs display SoMo, suggesting this phenomenon is important in initial midgut establishment [25, 26]. Therefore, initial successful midgut establishment is dependent on the ability of parasites to passage from the midgut lumen into the ES in order to colonise it. Many flies can show midgut infections but not all will progress to salivary gland infections due to a bottleneck effect at each developmental stage [27]. The gut of the insect (and especially the PM) is the first major barrier to successful establishment that will result in salivary gland colonisation.

1.6 Insect gut

Insects are the most diverse and abundant animals in terms of species number, biomass and ecological niches. Their range of habitats have led to the evolution of different feeding habits and food exploitation which in turn effectively determines gut structure and function [28] (Fig. 1.3). Despite these differences however, the guts of most insect species at all life

stages consist of a continuous tube from mouth to anus. Simply, the alimentary canal of insects is separated into three main regions, each with a different function; the foregut, midgut and hindgut. Within the foregut (mouth, pharynx, oesophagus, crop and proventriculus), food is ingested, ground, partially digested and stored, before being transported to the midgut. Here, the food source is subject to the majority of the digestive enzymes where it is digested and absorbed [29]. Finally, indigestible material is excreted from the anus in the hindgut. This region is also the site of water regulation and salt uptake. Along the length of the entire gut, insects have adapted an environment that is able to defend against biotic and abiotic factors. Both the foregut (stomatodeum) and hindgut (proctodeum) have a cuticular lining known as the intima, which upon challenge with pathogens, is able to mount an immune response [30]. The intima is also shed and renewed in those insects that undergo moulting events. The midgut region doesn't contain a cuticular lining but is instead lined with replicating epithelial cells that are constantly renewed and replaced. In some insect species, the midgut is further protected by secretions from either the foregut/midgut junction or the entire midgut which form a semi-permeable, acellular structure called the peritrophic matrix (PM) which surrounds ingested food and compartmentalises it [31]. This ensures separation between the endo- and ectoperitrophic space keeping the ingested meal and any pathogens contained within it from coming into contact with gut epithelial cells. In many insects, the PM may also line the hindgut which is then excreted as multiple layers covering the fecal pellets [32, 33].

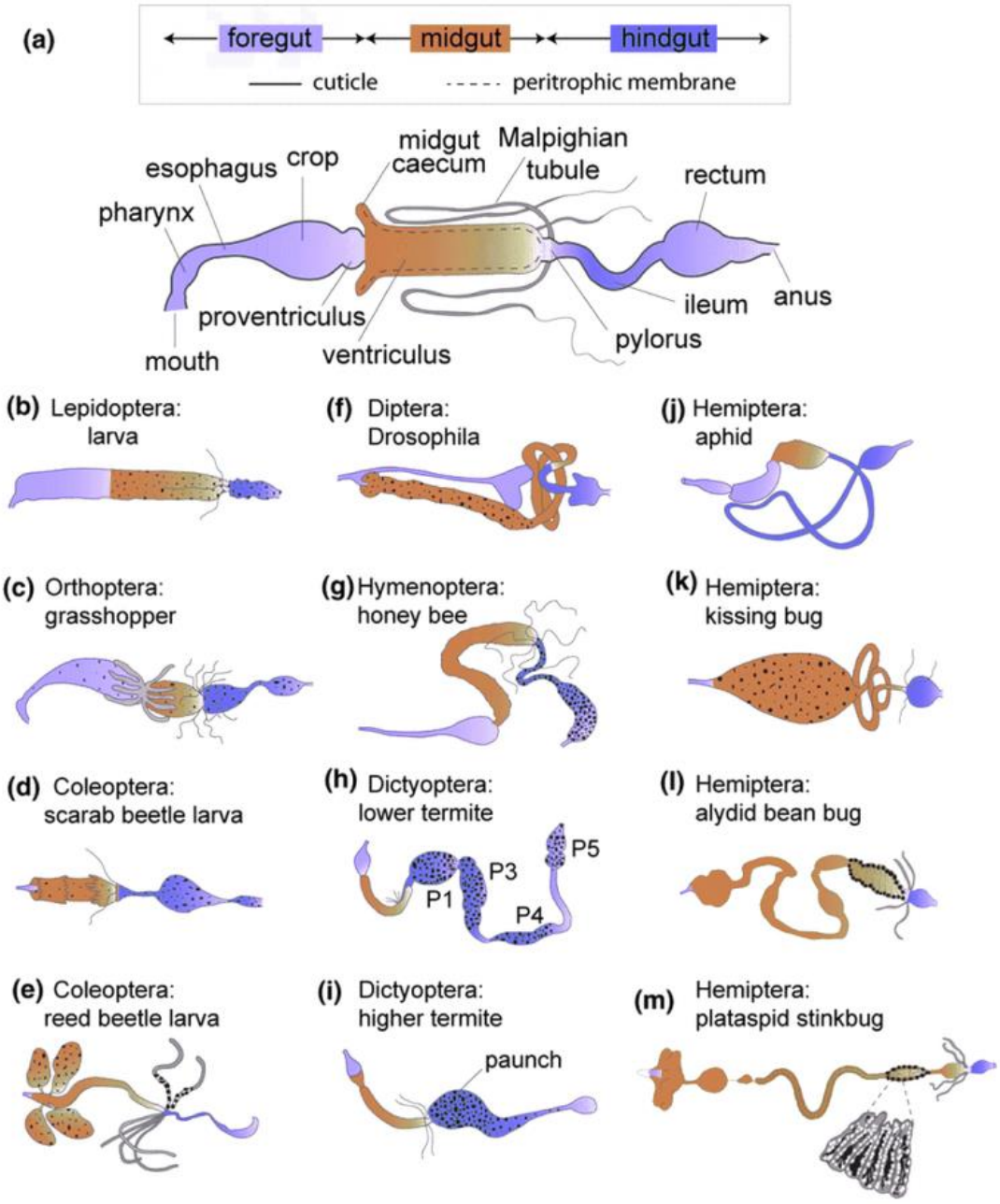


Figure 1.3. Insect guts. A: The generalised gut anatomy of insects. A solid line indicates cuticle and a dashed line indicates the peritrophic matrix. B: The larvae gut anatomy of Lepidopterans. C-M: The adult gut anatomy from several insect orders. Stipules specify the major residence of gut endosymbionts and in (M) shows a close-up of posterior midgut crypt cells and the localisation of gut microbiota within. Adapted from [34]

1.7 Insect Peritrophic Matrix

The peritrophic matrix (PM) is a non-cellular lining that surrounds the food bolos within most insect's midguts [35]. The function of the PM is somewhat disputed, but it is thought that it is similar to the mucous secretions of mammalian digestive tracts, in that it acts as a physical barrier to abrasive food particles, digestive enzymes and, to some extent, infective pathogens. The PM also acts as a biochemical barrier, retaining ingested toxins [35], [36]. There are two types of PM; type I which is secreted by the midgut epithelial cells and found in insects such as *Dictyoptera*, *Lepidoptera* and *Coleoptera*, and type II which is secreted by the proventriculus (PV=cardia), a group of specialised cells located in the anterior midgut [37], [38].

1.7.1 Tsetse Peritrophic Matrix

1.7.1.1 Production and Structure

Glossina spp. secrete a type II PM into the midgut as continuous concentric sleeves and ultrastructural and histochemical studies have shown it to be comprised of a 3 layered structure [39], with each one differing in thickness and composition. The type II PM is a highly organised structure consisting of chitin (poly β-(1,4)-*N*-acetyl-D-glucosamine [GlcNAc]), glycosaminoglycans (GAGS) and glycoproteins. It has been suggested that there is an abundance of *O*-linked glycans in a type II PM and this PTM is abundant on mucins and mucin-like peritrophins. There are many different types of glycoproteins with different functions; including mucins which act as secretory compounds to aid water retention and resist enzymatic proteolysis. Proteins which are bound to the PM have been shown to comprise of between 35-55% of the entire PM composition in many insects [40], and these PM associated proteins can be placed into one of four different protein classes according to the difficulty of which the protein is able to be extracted from the PM [41]. Structural proteins integral to the PM are grouped into the class III proteins and are collectively known as peritrophins.

1.7.1.2 Chitin

Chitin is one of the most abundant biological polysaccharides that exists. It is a polymer of β-(1, 4)-*N*-acetyl-D-glucosamine (GlcNAc) and reportedly constitutes between 3 and 13% of the PM [42], [43], although the chitin content of tsetse PMs is still unknown. It is, however suggested to be lower than that found in type I PMs and likely to mimic the low chitin quantity found in other type II PMs [44]. Parasites which are able to infect different insect species are known to synthesise chitinase in order to cross the PM, including *Plasmodium*

and *Leishmania spp.* [45], [46]. However, studies from the trypanosome genome have proved that *Trypanosoma brucei* do not produce chitinase [47], [48], suggesting a correlation with the low chitin content thought to comprise tsetse PMs.

1.7.1.3 Peritrophins

Although peritrophins are non-covalently linked to the PM, they are bound so tightly that they require harsh treatment to remove them. 8-M urea or 6-M guanidine hydrochloride is frequently used for this process; although mild detergents such as sodium dodecyl sulphate (SDS) can also be used [35]. The peritrophins are part of a large family of insect proteins that vary immensely in size, sequence and type and interact with irregular chitin fibres forming laminates of anionic glycosaminoglycans (GAGs). They are characterised by possessing one or more chitin binding domain (CBD) with conserved aromatic residues which promote the binding of chitin [49, 50]. There are 3 types of CBDs which are determined by the arrangement of conserved cysteine residues within the protein; Peritrophin A Domains (PAD) which have 6 conserved cysteine (C) residues in a consensus of $CX_{13\text{-}20}CX_5CX_{9\text{-}19}CX_{10\text{-}14}CX_{4\text{-}14}C$, Peritrophin B Domains (PBD) which have 8 conserved cysteine residues in the consensus of $CX_{12\text{-}13}CX_{20\text{-}21}CX_{10}CX_{12}CX_2CX_8CX_{7\text{-}12}C$ and Peritrophin C Domains (PCD) which have 10 conserved cysteine residues with the consensus $CX_{17}CX_{9–10}CX_{14}CX_9CX_{8–9}CX_{19}CX_{9–11}CX_{14}CX_{11}C$, where X is any amino acid other than cysteine. These cysteine residues mediate chitin binding interactions through formation of intra-disulphide bridges, and depending on secondary structure conformation, determine its affinity for chitin and resistance to proteolytic digestion [51]. The PADs are the most abundant type of CBD in the PM, where their function is suggested to help form multiple cross-links between chitin present in the PM and the peritrophin. However, proteins containing PADs are not always confined to the PM but can also be found in numerous other proteins such as chitinases, serine proteases [52], cuticular peritrophins and cuticular proteins analogous to peritrophins (CPAPs) [53] since these also have an association with chitin.

1.7.1.4 Insect Intestinal Mucins

Peritrophins contain at least one CBD but can also contain multiple copies of CBDs, with a combination of either the same or different CBD type. They may also contain mucin domains (MDs), in which case they are more commonly referred to as insect intestinal mucins (IIMs) [54, 55]. The amino acid sequence of mucin domains are rich in threonines,

serines and prolines which result in high levels of predicted *O*-linked glycosylation and may aid in water retention and tensile strength of the PM. Mucins have been implicated in providing protection in insect midguts and other tissues in an analogous way to vertebrate mucins. They are also known for their ability to bind pathogens, thereby mediating parasite-midgut interactions and inhibiting infection establishment [56, 57]. Oligosaccharides on the side chains of mucin have been shown to mimic pathogen cell surface receptors and are able to competitively bind pathogens [58]. Once bound, these pathogens are susceptible to the digestive processes in the lumen and to the natural turnover rate and renewal of the PM. Although there have so far been only 2 *Glossina* peritrophins identified that contain mucin domains [59], it is not yet understood if these have a role in mediating trypanosome infection in the fly.

1.7.1.5 Function

The main function of the tsetse PM is to compartmentalise ingested blood and protect the midgut epithelium from damage caused by the process of blood meal digestion. It also acts as a protective barrier to pathogens ingested with food and it has been suggested that insects which feed on a sterile diet typically possess no PM or a type II PM whereas those possessing a type I feed on diets more likely to contain pathogenic micro-organisms thus a type II PM is a more efficient barrier to pathogens [35, 60].

1.7.1.6 Permeability

The porous nature of the tsetse PM enables it to act as a semi-permeable barrier, allowing the movement of digestive enzymes from the midgut epithelial cells into the lumen and the reverse movement of digested food and other small molecules into the ectoperitrophic space [35]. The pore size in tsetse (~9nm) indicates a degree of selectivity to certain molecules and the arrangement and type of GAGs give it a fixed charged density [33, 36]. Differences in the packing of the polysaccharides, affected by pH and the ionic environment, determine the PM permeability by altering the ionic interactions between the PM and solute. This was shown in tsetse where an increase in calcium ions in the midgut altered the porosity of the PM causing an increase in permeability to alkaline phosphatase [61]. GAG properties also change under different pH values [62, 63] and could alter PM permeability by influencing the distribution of ionic charges on the polysaccharides which would have implications in tsetse as different regions of the midgut have different pH values.

1.8 Recent findings on the functions of peritrophic matrices and implications for novel vector control strategies and transmission blocking vaccines

In recent years, experimental evidence based on the analyses of insects with a type I and type II PM has been provided for the following major PM functions; the facilitation of digestion by the compartmentalisation of digestive enzymes and/or the products of digestion, the neutralisation of ingested toxins through binding or retention, and protection from pathogens including parasites, bacteria and viruses. It has been shown that PM formation and digestion is a factor in successful *Leishmania* development in various sandfly species. In competent vectors, the PM formation was slower as was the activation of trypsin in comparison to refractory species [64]. RNAi analysis of two peritrophic matrix proteins in the red flour beetle, *Tribolium castaneum*, indicated that the proteins were important in maintaining the structural integrity of the PM as indicated by its leakiness after knockdown. The authors also suggested these proteins are implicated in the formation of a PM permeability gradient which facilitates digestion [65]. The PMs from some phytophagous (plant feeders) insects are able to neutralise phenolic compounds, mainly through sequestration and elimination of toxins, ultrafiltration or through antioxidant activity [66]. Similarly, in some heamatophagous insects, the PM has been shown to be involved in detoxification of the toxic by-products of blood digestion. An IIM (AeIMUC1) from the PM of *Aedes aegypti* was shown to be expressed in larvae in response to heavy metal exposure [67]. The same protein was subsequently found to bind haem in the midguts of adult females during haemoglobin digestion. This binding via multiple haem-binding sites promotes haem aggregation thereby significantly lowering the toxic effects [68], [69]. RNAi of AeIMUC1 was shown to significantly reduce the number of mature oocysts in *Aedes aegypti*, possibly through the parasites increased exposure to haem. The PM in the fruit fly, *Drosophila melanogaster,* has been shown to protect the midgut against infections from ingested entomopathogenic bacteria. One of the PM components found to be responsible for this protection was drosocrystallin, an essential protein in normal PM formation. Mutant flies defective in drosocrystallin presented with decreased survival rates after oral infections due to a decrease in PM width and increase in PM permeability. In addition, the flies were also more susceptible to bacterial extracts containing Monalysin, a pore-forming toxin produced by *Pseudomonas entomophila*, indicating a multiple protective role of the PM in *Drosophila* [70]. A similar study in tsetse, however, showed the reverse to be true. When the PM in tsetse was compromised through RNAi of structural peritrophins, the fly was able to successfully clear orally ingested bacteria due to the earlier presence of attacin, an antimicrobial peptide. When challenged with trypanosomes however, flies with a compromised PM were more susceptible to infection. The authors suggested therefore that

the PM is not a physical obstruction to infection but rather acts as a barrier which regulates the flies' immunological responses [71]. Given the protective role of the PM against pathogens and its involvement in digestion and development processes, novel control strategy ideas have recently shifted to insect management through targeting of the PM. Reducing the numbers of the tsetse vector through the use of traps is one of the most effective ways of bringing outbreaks of HAT under control [72], due to the slow reproductive rate of the fly. Identification of new genetic targets could provide better and more cost efficient methods of controlling tsetse numbers. Moreover, identification of targets that inhibit the successful establishment of trypanosomes in the tsetse midgut may provide protection to humans and animal through the development of a transmission blocking vaccine.

1.9 Aims and objectives of this study

As reviewed above, the tsetse PM plays a critical role in many physiological aspects of the fly including bloodmeal compartmentalisation and digestion. It is also the first line of defence against ingested trypanosomes. The overall aim of this study was to determine the role of the PM in successful establishment of trypanosomes in the midgut of tsetse, with focus on the structure and function of peritrophins and their role in trypanosome migration using a multi-faceted approach;

1. To provide a comprehensive identification of the proteins that comprises the tsetse PM (Chapter 2).
2. To identify novel peritrophins within the PM of *Glossina morsitans morsitans* (Chapter 2).
3. To compare identified peritrophins of *G. m. morsitans* to analogous genes in four other major tsetse vectors and a comparative non blood-feeding species, *Musca domestica* (Chapter 3)
4. To provide structural analysis of the major PM peritrophins and to expand the understanding of peritrophins containing Peritrophin-A Domains (PADs) in Dipterans (Chapter 3).
5. To evaluate RNAi as a way of inhibiting expression of selected peritrophins and the effect it has on protein expression (Chapter 4).
6. To re-examine the migration of trypanosomes through the tsetse PM, using advanced, multiple microscopy approaches (Chapter 5).

Chapter 2. An investigation into the protein composition of the teneral *Glossina morsitans morsitans* peritrophic matrix.

The journey begins……..

2.1 Introduction[1]

The concept of blocking trypanosome development within its tsetse host has been underexplored, primarily due to a lack of understanding the molecular events involved in the vector-parasite interactions and also difficulties in accessing an established colony of tsetse flies needed to implement such studies. Due to evidence of emerging parasite resistance to the current frontline therapeutics [73], mammalian toxicity to current drug treatments and no working vaccine, new disease transmission control ideas have shifted to investigating the vector-parasite interface rather than targeting parasite interactions within the mammalian host.

For successful transmission to occur, salivarian trypanosomes must overcome many immunological and physical barriers to undergo a complex migration and development in the fly. Once ingested with a bloodmeal, the bloodstream form transforms in the midgut lumen into the procyclic stage within 1-2 days post-ingestion. After a successful differentiation into procyclics, the parasites then must avoid the proteolytic attack of tsetse digestive enzymes, reactive oxygen species [74], immune peptides [75] and serum complement [76]. They do this by escaping to the ectoperitrophic space (ES) in the midgut thereby breaching the peritrophic matrix (PM), an acellular secretion that lines the midgut of many insects and could be present in more than one life stage [35], [77]. After establishing an infection in the ES, the trypanosomes then colonise the proventriculus (PV) or cardia, where they continue to develop into long epimastigotes, which then cross the PM for a second time *en route* to the salivary glands. More details on the cycle of African trypanosomes in the fly are described in Chapter 1.

In general, insect PMs are believed to be multi-functional and several roles have been proposed for this structure. Most functions depend on the selective permeability of the PM, but it is generally accepted that this tissue is analogous to the mucous secretions of mammalian digestive tracts [35, 77, 78], in that it acts as a physical barrier to abrasive food particles and digestive enzymes. It has also been demonstrated that the PM acts as a biochemical barrier retaining ingested toxins [68, 69, 79], thereby preventing cell damage and lethality to the insect. Perhaps more importantly, insect PMs impose physical barriers that prevent pathogens from reaching the midgut epithelium as demonstrated in mosquito studies [80-82], and more recently, shown in two publications in *Drosophila* and *Glossina* [70, 83].

There are two types of insect PMs described: type I and type II. Many heamatophagous adult diptera and important parasite vectors such as sand flies and mosquitoes possess a type I, which is secreted once from the midgut epithelial cells. Tsetse produce a type II PM,

[1] Parts of this chapter has been published in PLoS NTD; DOI: 10.1371/journal.pntd.0002691

which is present prior to taking a bloodmeal and is continually secreted by a specialised group of cells in the PV (Chapter 5). Electron microscopy, in combination with cytochemistry and lectin binding approaches, revealed that adult tsetse possess a highly organized, 3-layered PM (~340 nm thick) composed of glycosaminoglycans (GAGs), glycoproteins of unidentified nature and chitin (poly β-(1,4)-*N*-acetyl-D-glucosamine [GlcNAc]) fibers [78, 84]. In addition, very little is known on its overall protein composition and there is limited knowledge of the number of peritrophins that compose the tsetse PM. Until now, only Proventriculin 1 (GmmPro1) and Proventriculin 2 (GmmPro2) have been identified as putative components of the tsetse PM since these proteins are produced exclusively in the PM-secreting PV [85]. These putative peritrophins have barely been characterised, however, it is known that GmmPro2 is upregulated in susceptible tsetse lines (salmon flies) [86]. More recently, it was shown that RNAi induced knockdown of GmmPro2 increases the flies' survival to oral bacterial infections by altering the integrity of the PM and allowing the bacteria to come into contact with the antimicrobial peptide attacin much earlier than if the PM was intact [71].

Peritrophins are structural PM proteins that are characterised by containing at least one chitin binding domain (CBD) that in turn have several conserved aromatic residues [50]. These CBDs interact with and bind chitin fibres present in the PM and other chitin containing proteins, which effectively influence PM tensile strength, elasticity and porosity, whilst the aromatic residues may bind carbohydrates. Peritrophins can also possess one or more mucin domains, reflecting the fact that they are believed to have evolved from mucins with the acquisition of CBDs [50, 87]. These mucin domains possibly act as secretory compounds that aid water retention and resist enzymatic proteolysis.

The teneral tsetse PM is the only partial physical barrier to trypanosome infection in the tsetse midgut and modifications to the PM as the fly ages may lead to a complete barrier to infection [88]. There is good evidence using electron microscopy that trypanosomes penetrate the tsetse PM [89, 90]. However, this process must be dependent on the activity of PM-degrading enzymes since the pores in the tsetse PM are approximately 9 nm in size, which are too small for procyclic trypanosomes (several microns long) to pass through [61, 91]. It is possible that proteins integral to the tsetse PM are important in infection establishment considering that parasites of other invertebrates secrete hydrolytic enzymes to degrade PM proteins in their respective hosts. In addition, for crossing or penetration to occur, there must first be adherence or attachment to specific ligands or receptors. Many pathogens such as bacteria and helminths have developed a number of mechanisms for attachment to their target cell prior to invasion. These include the development of fimbria, pilli, pseudopodia and, for helminthic parasites, suckers and cutting appendages. In addition to these structural modifications pathogens and parasites can also secrete

glycoconjugates to aid cell entry including adhesion molecules and cellular receptors 67. For example, *Plasmodium* ookinetes express on their surface the enzyme enolase which specifically binds plasminogen present in the blood. This interaction causes plasmin to become activated which facilitates the invasion of the midgut epithelium by the parasite. It is also suggested that enolase plays a dual role in that it acts as a ligand for midgut epithelial receptors as it contains an SM1-like domain, a peptide shown to bind tightly to the midgut luminal surface [92].

To understand if trypanosomes use such strategies as these described, a thorough revision of the composition and structure of the tsetse PM is required. Using a proteomics approach, this study provides the first insight into the overall protein content of the tsetse PM in an effort to understand, at the molecular level, the events involving trypanosome migration within the tsetse vector.

2.2 Materials and Methods

2.2.1 Tsetse fly maintenance and dissection of peritrophic matrices

Glossina morsitans morsitans (Westwood) were taken from an established colony at the Liverpool School of Tropical Medicine, which was maintained on sterile, defribinated horse blood (TCS Biosciences) at a relative humidity of 65-75% and an ambient temperature of 27°C ± 2°C. Experimental flies where collected at <24 hours post eclosion where they were briefly chilled at 4°C for initial sorting and kept in a 12 hour light and dark cycle in the same conditions as the colony until they were 72 hours old. Initial analyses included teneral (young, unfed) and fed male and female adults with the subsequent proteomic analysis carried out on teneral males. PMs were dissected in sterile, chilled phosphate buffered saline solution (PBS, 140 mM NaCl, 1 mM KCl, 6 mM phosphate buffer, pH 7.4), transferred to 1.5 ml microcentrifuge tubes containing 200 µL of sterile PBS and centrifuged at 18,400x*g* for 5 minutes at 4°C. The supernatant was removed and the remaining PM pellet was washed three times in ice-cold distilled water for 10 minutes each at 18,400x*g* (to remove excess salts, non-adhered bacteria and midgut contaminants) then snap frozen and kept at -80°C until needed.

2.2.2 Mild acid hydrolysis of *G. m. morsitans* PM proteins

200 µL of 0.3% triflouroacetic acid (TFA) was added to PM samples from males and females at each time point. They were then boiled at 100 °C for 30 minutes before being frozen in liquid nitrogen. Each sample was then lyophilized overnight and each pellet subsequently re-suspended in 15 µL 2 x Laemmli with equal parts distilled water. Each sample was boiled at 100 °C for 10 minutes before being loaded onto a 15% acrylamide, 1.5 mm mini-gel.

2.2.3 Chemical de-glycosylation of *G. m. morsitans* PM proteins using TFMS

PM material from male flies at 48hrs old were washed sufficiently in distilled water and dehydrated by lyophilisation for 24 hrs. Keeping the samples on ice, 50 µL trifluoromethanesulfonic acid (TFMS) was added according to the protocol for the deglycosylation of N-linked and O-linked glycoproteins (Sigma GlycoProfile IV Chemical Deglycosylation Kit, PP0510). Neutralisation using 150 µL 60% pyridine was performed at 5, 15, 30 and 60 minutes and RNase B supplied from the kit was used as a positive control for deglycosylation. Samples were dried in a speed vacuum for 24 hrs and proteins were

recovered using a 3 kDa, 6.5 mL Amigon ultrafilter in order to remove the salt precipitate. Proteins were subsequently resuspended in 2 x Laemmli buffer and equal parts 0.5 M DTT before being loaded into a 15% acrylamide, 1.5 mm mini-gel. 1D SDS-PAGE was carried out for 3 hours at 80V.

2.2.4 Solubilisation of PM proteins using Urea/SDS

PMs from ~150 24 hr old tsetse were thawed and re-suspended in 150 µL of 50 mM Tris-HCl (pH 6.8), containing 8 M urea, 3% SDS and 50 mM Dithiothreitol (DTT). The sample was then sonicated in a sonicating ice-cold water bath 3 times for 5 minutes each and PM proteins precipitated with trichloroacetic acid (TCA)-acetone. Briefly, the PM suspension was mixed with 100% ice-cold acetone and 100% TCA (1:8:1, V/V/V respectively) and kept at -20°C for 1 hour [93]. After precipitation, the sample was centrifuged at 12,400 x *g* for 15 minutes at 4°C, the supernatant discarded, and the protein pellet was washed twice with 1 ml ice-cold acetone. After the last wash, the remaining acetone was allowed to evaporate at room temperature, and the protein pellet was then re-dissolved in distilled water, mixed with Laemmli buffer [94], and heated for 10 minutes at 95^{o}C. In a separate experiment, 150 PMs were extracted and solubilized in urea buffer as described above, and then processed for in-solution tryptic digestion as described below.

2.2.5 1D Polyacrylamide Gel Electrophoresis and staining with Coomassie Brilliant Blue G-250 for proteomic analysis

The PM protein preparation was fractionated on a NuPAGE (Invitrogen) precast 4-12% gel Tris-Bis gradient gel according to the manufacturer's recommendations. The gel was fixed overnight and the proteins were stained with colloidal Coomassie Blue G-250 (Sigma) as described by Neuhoff [95], to allow sensitive visualization and destaining of proteins prior to mass spectrometry analysis.

2.2.6 Western blotting

Approximately 10 µg/lane of a preparative PM protein urea extract were fractionated on a 12% SDS-PAGE and then transferred onto BioTrace polyvinylidene diflouride (PVDF) membrane at 90 V for 30 minutes. The membrane was then incubated overnight at 4°C in blocking buffer (PBS/0.1% (v/v) Tween 20/5% (w/v) skimmed milk powder), containing

0.05% (w/v) sodium azide to prevent bacterial growth. After several washes in washing buffer (PBS/0.1% (w/v) Tween 20), separate membrane strips (containing equal amounts of protein) were probed for 1 hour at room temperature with either anti-tsetse or anti-bacterial primary antibodies: 1) mAb 4A2 (mouse anti-Proventriculin 2 (Pro2)) 1:25 dilution, 2) mAb TBRP/247 (mouse anti-EP procyclin) 1:10 dilution, 3) mAb 3B2 (mouse anti-lectin) 1:2 dilution 4) mAb 1H1 (mouse anti-symbiont GroEL) 1:20 dilution, and 5) polyclonal rabbit anti-*Sodalis glossinidus* 1:10,000 dilution. All antibodies were a generous gift from Prof. Terry Pearson (University of Victoria, Canada). After several washes, the strips were incubated with a 1:50,000 dilution of secondary antibody (goat anti-mouse IgG (antibodies 1-4), or mouse anti-rabbit, Thermoscientific (antibody 5) (all conjugated to horse radish peroxidase (HRPO)) at room temperature for 1 hour. After several washes, the strips were incubated with SuperSignal West Dura (Pierce, UK) peroxidase buffer and luminol/enhancer solution at a 1:1 ratio, and developed by chemiluminescence, which continued for up to 3 hours.

2.2.7 Tryptic digestion

2.2.7.1 In-gel digestion

Excised gel plugs were destained in 50% acetonitrile/25 mM ammonium bicarbonate (pH ~8), reduced for 30 minutes at 37°C with 10 mM dithiothreitol (Sigma) in 50 mM ammonium bicarbonate and alkylated with 55 mM iodoacetamide (Sigma) in 50 mM ammonium bicarbonate for 30 minutes in the dark at room temperature. Gel plugs were washed for 15 minutes in 50 mM ammonium bicarbonate and dehydrated with 100% acetonitrile. Acetonitrile was removed and the gel plugs rehydrated with 0.01 µg/µL proteomic grade trypsin (Sigma) in 50 mM ammonium bicarbonate. Digestion was performed overnight at 37°C. Peptides were extracted from the gel plugs using successive 15 minute incubations of 2% (v/v) acetonitrile, 1% (v/v) formic acid. Peptide extracts were pooled and reduced to dryness using a centrifugal evaporator (Jouan RC10-22), and re-suspended in 3% (v/v) acetonitrile, 0.1% (v/v) TFA for analysis by mass spectrometry.

2.2.7.2 In-solution digestion

For in-solution digestion, acetone precipitated PM material was solubilised with 0.1% (v/v) Rapigest (Waters Corp.) in 25 mM ammonium bicarbonate. The sample was heated at 80ºC for 10 min, reduced with 3 mM DTT (Sigma) at 60ºC for 10 min, and then alkylated with 9 mM iodoacetamide (Sigma) at room temperature for 30 min in the dark. Proteomic grade trypsin (Sigma) was added at a protein:trypsin ratio of 50:1, and samples were incubated at 37ºC overnight. Rapigest was removed by adding TFA to a final concentration

of 1% (v/v) with incubation at 37ºC for 2 hours. The peptide samples were then centrifuged at 12,000x*g* for 60 min at 4ºC to remove precipitated Rapigest. Peptides were desalted using C18 Stage tips (Thermo scientific), then reduced to dryness centrifugal evaporator (Jouan RC10-22), and re-suspended in 3% (v/v) acetonitrile, 0.1% (v/v) TFA for analysis by mass spectrometry.

2.2.8 Liquid chromatography-mass spectrometry (LC-MS) analysis of tryptic peptides

Peptide mixtures, generated by in-gel proteolysis of excised protein bands from polyacrylamide gels, were analysed by reverse-phase liquid chromatography (RPLC) using an UltiMate™ 3000 LC system (DIONEX) coupled to an LTQ (Thermo Fisher Scientific) mass spectrometer. Peptides (10 µl) were injected onto a C18 column (2 µm particle size (100), 75 µm diameter x 150 mm long) at nanoflow rate (300 nl/min) and separated over a 50 minutes linear chromatographic gradient. The gradient consisted of the following phases: 0-30 min, 0-50% buffer B (linear); 30-30.1 min, 50-100% buffer B (linear); 30.1-35 min, 100% buffer B; 35.1-50 min, 0% buffer B. Full scan MS spectra (*m/z* range, 400-2000) were acquired by the LTQ operating in triple-play acquisition mode. The top three most intense ions were selected for zoom scan and tandem MS by collision-induced dissociation (CID).

Peptide mixtures, generated by in-solution proteolysis, were analysed by on-line LC using the nanoACQUITY-nLC system (Waters Corp.) coupled to an LTQ-Orbitrap Velos (Thermo Fisher Scientific) mass spectrometer. Peptides (~500 ng) were injected onto the analytical column (nanoACQUITY UPLC™ BEH130 C18. 15 cm x 75 µm, 1.7 µm capillary column) at nanoflow rate (300 nl/min). The linear gradient consisted of 3-40% acetonitrile in 0.1% formic acid (v/v) over 120 min, followed by a ramp of 40-85% acetonitrile in 0.1% formic acid for 3 min. Full scan MS spectra (*m/z* range. 300-2000) were acquired by the Orbitrap at a resolution of 30,000. A data-dependent CID data acquisition method was used. The top 20 most intense ions from the MS1 scan (full MS) were selected for CID in the LTQ ion trap.

2.2.9 Protein identification

2.2.9.1 *Glossina* Proteins

Preliminary MS data were searched against the Arthropoda non-redundant database in NCBI. Subsequently the Glossina genome was publicly released and data resulting from the MS analyses from urea/SDS extracted PMs were searched against Glossina specific peptides. Tandem MS data were searched against the *Glossina morsitans morsitans* database Glossina-morsitans-Yale_PEPTIDES_GmorY1.1.fa.gz downloaded from

VectorBase (https://www.vectorbase.org/proteomes) using the Mascot (version 2.3.02, Matrix Science, Liverpool) search engine. Search parameters were a precursor mass tolerance of 10 ppm for the in-solution digest using the LTQ-Orbitrap Velos and 0.6 ppm for the lower resolution LTQ instrument. Fragment mass tolerance was 0.6 Da for both instruments. One missed cleavage was permitted, carbamidomethylation was set as a fixed modification and oxidation (M) was included as a variable modification. For in-solution data, the false discovery rate was filtered at 1%, and individual ion scores ≥30 were considered to indicate identity or extensive homology (p <0.05). Individual MS/MS spectra for single peptide hits with an ion score of 30 or above were inspected manually and only included if a series of at least four continuous fragment ions were observed (Ch2_SF1 and Ch2_SF2).

2.2.9.2 *Sodalis* Proteins

Tandem MS data were also searched against the *Sodalis glossinidius* peptide database generated from the latest re-annotated coding sequences [96] using the same search engine and parameters as described above.

2.3 Results

2.3.1 PM protein extraction using 3% SDS-DTT

Initial analyses of female and male tsetse were carried out on teneral, 24 and 48 hour old flies and on blood-fed flies at 15 days post-eclosion (not all data not shown); with blood-fed flies being fed every 2 days on an artificial membrane system. These comparisons were made as structural changes within the PM protein composition as the fly ages may contribute to flies becoming more resistant to trypanosome infection. The potential for infection of the tsetse fly is thought to be highest when the fly first takes its blood-meal, and then becomes increasingly refractory to trypanosome infection at each subsequent blood meal 38. 1D SDS-PAGE using 2% SDS as a detergent, showed slight dissociation of proteins as most of the resolved proteins were no smaller than 17 kDa. SDS is able to denature secondary and non-disulfide-linked tertiary structures, but as SDS is only a mild detergent, not all the proteins linearise sufficiently. Preliminary mass spectrometry (MS) analysis resulted in the identification of only a few *G. m. morsitans* proteins.

2.3.2 PM protein extraction using TFA

Due to the low number of proteins that were solubilised and separated using SDS in Laemmli buffer, a new batch of PM material was then treated with 40 mM (0.3%) of triflouroacetic acid (TFA) for 30 minutes in order to separate those proteins that were attached to chitin. TFA is a mild acid that preferentially cleaves Asp-Pro bonds and, to a lesser extent Asp-Gly bonds (Fig. 2.1). Bioinformatics studies have shown that proteins from the tsetse midgut appear to contain extensive Asp-Pro peptide bonds [97]. Therefore fragments of those proteins that are extensively cross-linked with chitin and possibly other sugars, which were previously unable to be detected, may be cleaved, visualised by SDS-PAGE analysis and identified by MS.

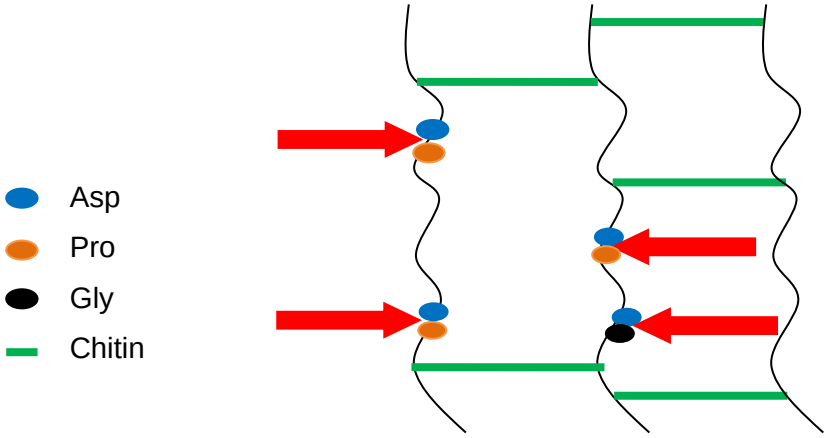


Figure 2.1. Schematic showing the potential cleavage sites (red arrows) that result from PMs treated with TFA.

Acid hydrolysis using TFA, followed by extraction with Laemmli buffer, was shown to be more efficient than using SDS alone as there was a higher amount of proteins as visualised by SDS-PAGE, including proteins smaller than 17 kDa (Fig. 2.2).

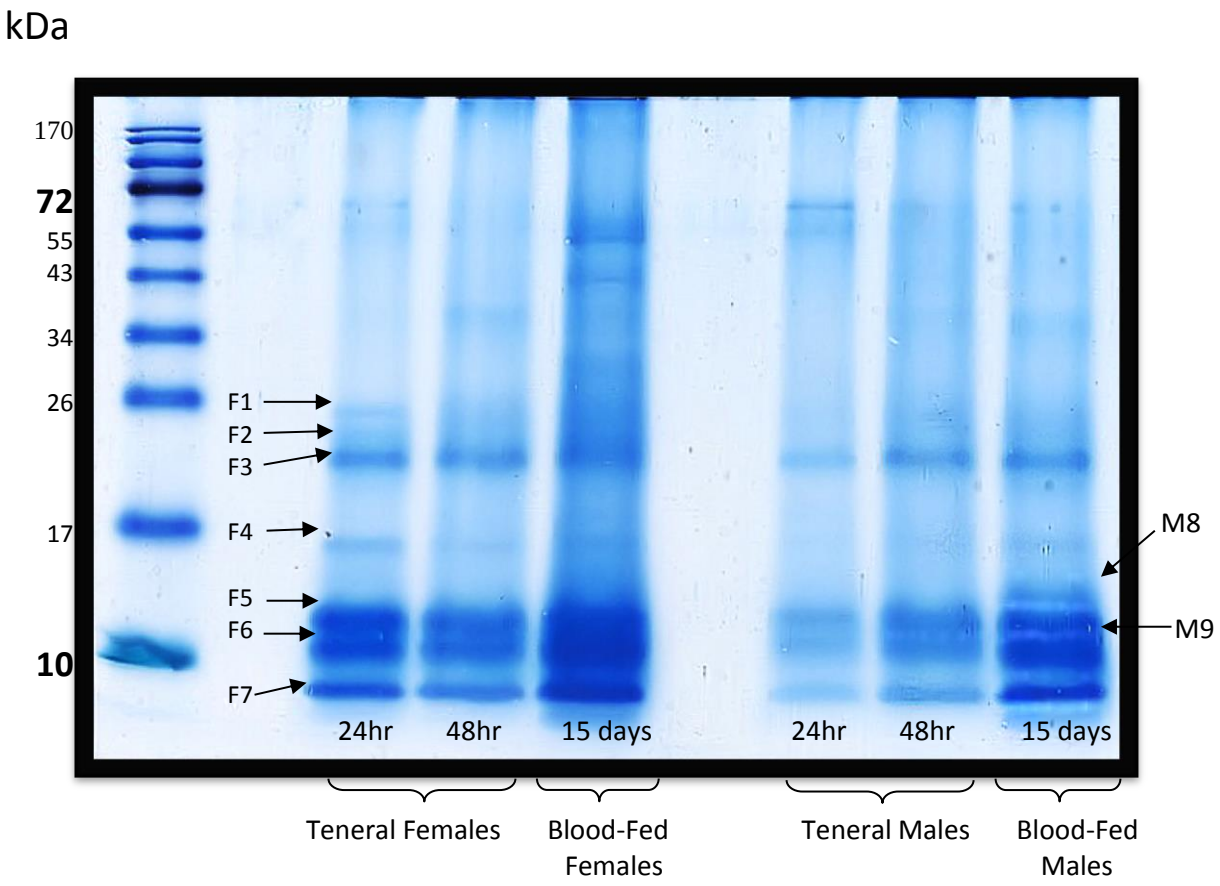


Figure 2.2. Proteins from the PMs of* G. m. morsitans *of different sexes and ages. *Proteins were extracted as described in the text above, separated by 1-D SDS-PAGE and visualised by Colloidal Coomassie Blue staining.*

By looking at the pattern of proteins on the gel it was clear that PM proteins between males and females and at different ages are the same, although there were minor differences in the intensity of bands. There was, however, a single band that appeared only in the males that were 15 days old (band M8; Fig. 2.2), which did not appear at any other age and did not seem to be there in the females at any age. Subsequent mass spectrometry analysis on band M8 after trypsinisation and a BLAST search against the NCBI database did not deliver any results. Other relevant bands (F1-F7 & M9; Fig. 2.2) were cored from the gel and sent off for mass spectrometry resulting in the identification of approximately 42 *Glossina* proteins, 12 of which were hypothetical, and 21 *Sodalis* proteins (Ch2_ST1).

2.3.3 PM protein extraction using TFMS

It has been shown that removal of glycans from the PM of several different insects is relatively easy using trifluoromethanesulfonic acid (TFMS). This acid is one of the most efficient chemical reagents used to selectively cleave O-linked and N-linked sugars from glycoproteins without disturbing the integrity of the polypeptide (Fig. 2.3). This type of chemical deglycosylation produces a high yield of proteins that are suitable for mass spectrometry analysis and has been shown to be very successful in the deglycosylation of proteins from the PM of many insects

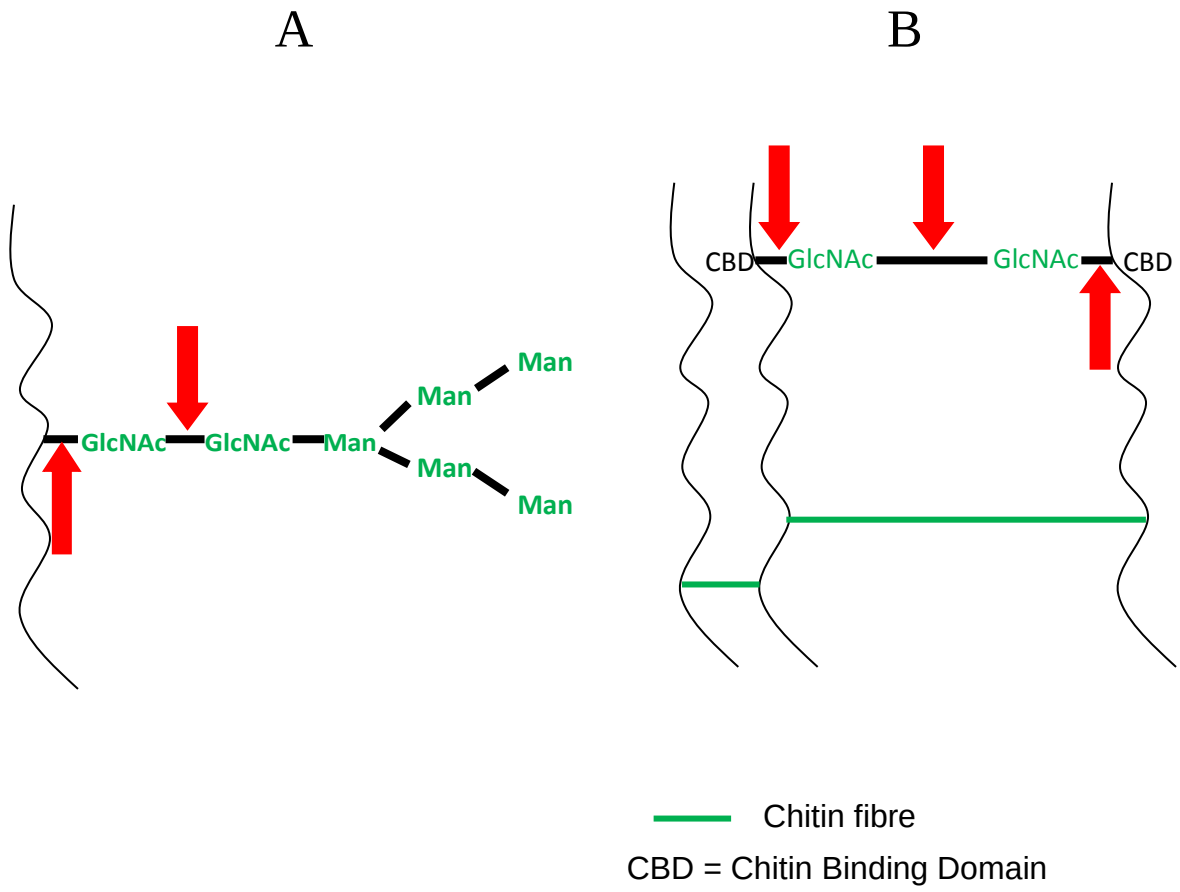


Figure 2.3. Schematic of cleavage sites of TFMS. *A; Schematic of high mannose; an N-linked glycoprotein that is found attached to the Glossina peritrophic matrix. B; Schematic of cross-linked peritrophins that are predicted to be in the peritrophic matrix of G. m. morsitans. TFMS is able to cleave at sites indicated by the red arrows.*

This methodology was first set up using a known glycosylated protein, Ribonuclease B, which contains one *N*-glycan chain. Subsequent 1D SDS-PAGE of TFMS treated Ribonuclease B showed a significant shift in the band pattern of proteins compared to the

untreated samples of RNase B (Fig. 2.4). The reduction in the apparent molecular mass (~2000 Da) is consistent with the complete removal of one N-linked oligosaccharide. The cleavage was efficient; as short as 5 minutes, and there was very little protein degradation (compare 5 vs. 60 minutes, Fig. 2.4).

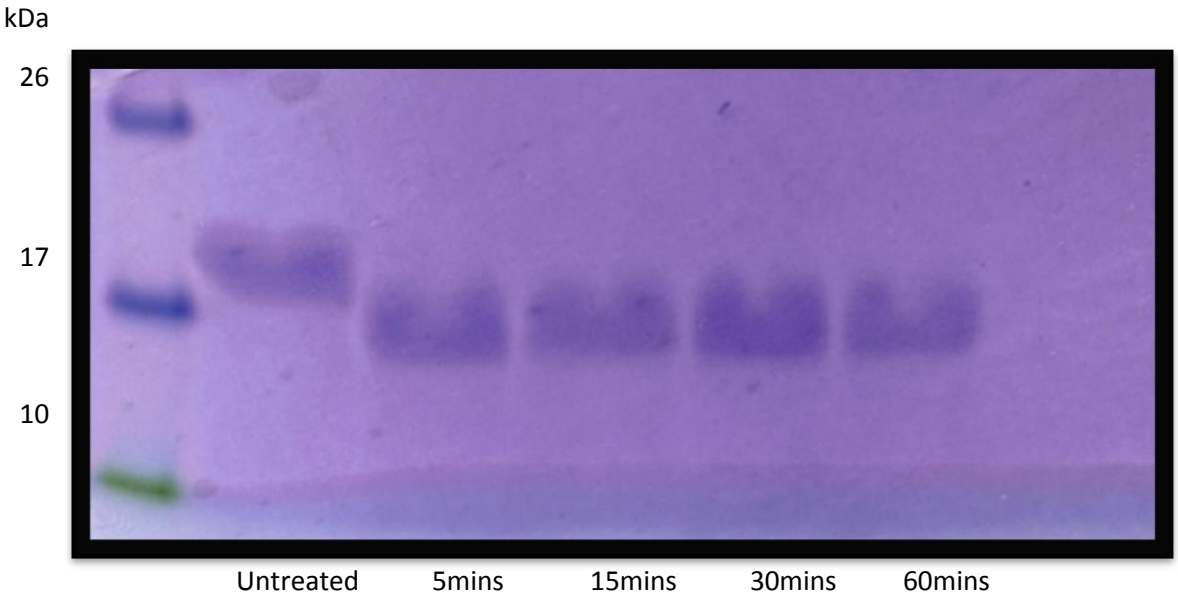


Figure 2.4. Time course of TFMS de-glycosylation of Ribonuclease B. *Batches of 20 µg of proteins were treated with TFMS as indicated in the methods section and analysed by SDS-PAGE. Proteins were visualised by Colloidal Coomassie Blue.*

Unfortunately, due to type 2 PMs being rich in *O*-glycosylated proteins, the TFMS needed a longer incubation time for a complete removal of this type of glycan which completely degraded the protein backbone. In addition, presumably the innermost Asn-linked GlcNAc residues of N-linked glycans would remain attached to the protein as they are not cleaved by TFMS. A shorter incubation time did not affect the protein but it failed to deglycosylate any glycoproteins.

2.3.4 PM protein extraction using 8M urea and 3% SDS

Approximately two batches of 150 male teneral (young, unfed) flies were used to determine the proteome of *G. m. morsitans* PM (one batch of 150 were used for in-gel analyses, the other 150 were used for in-solution analyses). The PM samples were extensively washed before homogenisation in Tris-HCl buffer containing 8M urea and 3% SDS, and cold-precipitation with acetone/TCA (Fig. 2.5a). The urea/SDS homogenisation produced a higher yield of proteins (as judged by analysis of Coomassie blue-stained SDS-PAGE gels) compared to the extraction with 3% SDS-DTT alone followed by either mild TFA hydrolysis [97] or anhydrous trifluoromethanesulfonic acid (TFMS) [98]. In addition, there were no significant differences in the pattern of bands on SDS-PAGE between PM samples extracted from either teneral or 15 day old, bloodfed flies, or in samples from either sex. Nevertheless, to avoid contamination from horse blood proteins in the mass spectrometry analyses and considering their high susceptibility to a trypanosome infection, only analyses of PMs extracted from teneral male flies were carried out.

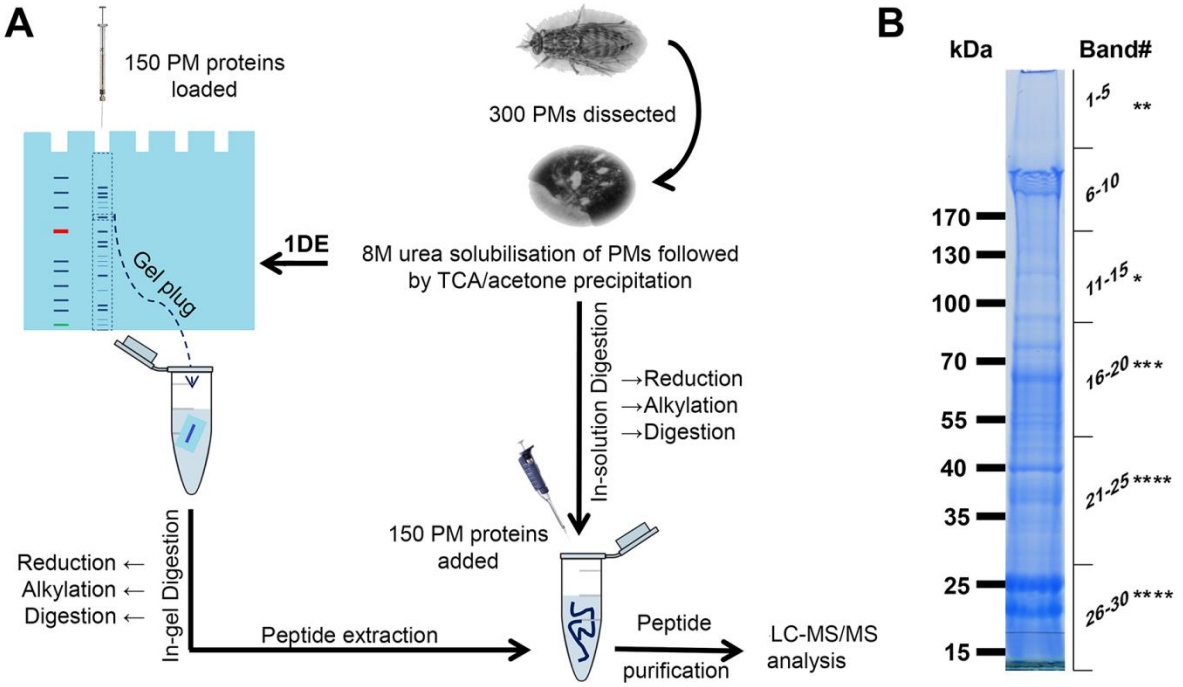


Figure 2.5. Experimental flow diagram. A total of 300 PMs were dissected and split equally into 2 tubes for urea/SDS solublisation and acetone precipitation. (A) One batch of PMs were sent directly for in-solution trypsin digestion whilst the other were fractionated on a 4-20% gradient gel. (B) 30 bands were cored from the gel from top to bottom and each gel plug was processed for in-gel trypsination mass spectrometry analysis. The position and number of asterisks indicate the regions of the gel where GmmPer66 was identified.

In order to increase the number of proteins identified, two approaches were used. One sample (containing ~150 PMs) recovered after acetone/TCA precipitation was fractionated on a 4-20% NuPAGE precast gel (Figure 2.5b) and the excised bands were digested with trypsin and processed for LC-MS/MS analysis (below), whilst another batch of ~150 PMs were directly trypsinised in-solution after urea solubilisation and precipitation as above, before mass spectrometry analysis. Furthermore, peptide sequences from both analyses were BLAST searched against the genomes of *G. m. morsitans* and the midgut symbiont *S. glossinidus*.

2.3.5 In-gel analysis

After colloidal Coomassie blue staining, many proteins with apparent molecular masses from ~21 kDa up to >200 kDa were visualised, although a slight smeariness in a number of bands indicated the presence of highly modified proteins. Since many proteins do not stain with Coomassie blue (e.g. mucins and peritrophins due to their high negative charge and acidity [99]), coverage was increased by slicing the stained gel lane in 30 pieces from top to bottom (Figure 2.5b). The individual bands were then excised, the proteins in-gel trypsinised and analysed by LC-MS/MS (Figure 2.5 and Table 2.1). This approach provided useful information regarding the relative abundance and masses of the different proteins (Table 2.1), whilst the in-solution analysis (below) increased the number of proteins identified.

The most visually abundant proteins on the gel were a doublet migrating with relative molecular masses around 26 and 21 kDa (bands 27 and 28, respectively), which were identified as midgut trypsins. However, the most abundant and frequent hit in many of the bands analysed was a new type of peritrophin herein referred to as GmmPer66 (discussed below in the peritrophins section and additional information in Chapters 3 and 4). In addition, GmmPro2, another known peritrophin-like protein that is produced in the PV [85] and the immunomodulatory TsetseEP protein [100, 101], were also detected in several bands (Table 2.1). The possible significance of the high occurrence of these proteins is discussed below. Furthermore, other peptidases, including GmmPro3 [85], one serine peptidase and one putative metalloprotease, one chitinase, and several uncharacterised/conserved/hypothetical proteins were also found. Not surprisingly, abundant hits were also found for metabolic proteins, transporters and extracellular matrix proteins. The significance of the presence of these proteins is discussed below.

Table 2.1. List of the most abundant proteins detected by mass spectrometry from in-gel analyses of the peritrophic matrix of teneral *G. m. morsitans*

VectorBase ID #[1]	Protein name	Occurrence	Main feature(s)
GMOY002708	GmmPer66/Peritrophin-like	14[2]	3 CBDs, 2 *N*-glycs.
GMOY009892	Dynein AAA+ ATPase	13	AAA+ ATPase domain
GMOY011773	Perlecan	11[3]	Ig set I-domain, ConA-like lectin
GMOY001776	Actin	9	Actin-related domain
GMOY005703	Myosin heavy chain	9	Myosin motor domain
GMOY007063	Midgut trypsin	8[4]	Trypsin-like serine peptidase domain, SP
GMOY009248	Lamin	5	Lamin tail domain
GMOY004611	Vesicular transport factor dp115	4	Armadillo repeats
GMOY006294	Glutamate semialdehyde dehydrogenase	4	Uridylate kinase, aldehyde dehydrogenase
GMOY009756	Trypsin/Proventriculin3 (Pro3)	4[5]	Trypsin-like serine peptidase domain, SP
GMOY000153	Chitinase Chit1 precursor	3	1 CBD (PAD), glycosidase
GMOY003579	Na/K transport. ATPase	3	P-type ATPase, 4 TMD
GMOY005442	Lipophorin	3	Vitellogenin lipid transport domain, SP
GMOY009587	Proventriculin2 (Pro2)	2[6]	Partial CBD (PCD)
GMOY003306	TsetseEP	2[7]	Tsetse-specific, immunity, SP
GMOY000672	Serine protease 6	2	Trypsin-like Ser peptidase domain, SP
GMOY009757	Serine type endopeptidase	2	Trypsin-like serine peptidase domain
GMOY011520	Alanyl aminopeptidase	2	Peptidase M1, DUF domain, *N*-glyc.
GMOY002421	Chaperonin-60 kDa	2	Cpn60/TCP-1, GroEL-like apical domain
GMOY011805	Choline O-acyltransf.	2	1 TMD, *O*-glyc.
GMOY007524	Hypothetical	2	*O*-glyc.
GMOY007847	Hypothetical	2	unknown
GMOY008627	Hypothetical	2	Unknown, SP
GMOY008757	Hypothetical	2	Unknown
GMOY001198	Hypothetical	2	Ig I-set domain

[1]VectorBase *G. m morsitans* database version GmoY1.1, 2013. Glossina-morsitans-Yale_PEPTIDES_GmorY1.1.fa.gz.
[2]GmmPer66 was found in bands 1, 3, 12, 16, 18, 20 to 24, 26 and 28 to30. CBD (PAD type).
[3]Identified from bands 4, 7 to 15 and 24.
[4]Found in bands 19 and 24 to 30.
[5]Identifed from bands 24 to 26 and 28.
[6]Identified from bands 29 and 30.
[7]Found in bands 23 and 24, which is consistent with the protein's predicted *Mr* of 37.5 kDa.
SP: Signal Peptide.
TMD: Transmembrane Domain.
doi:10.1371/journal.pntd.0002691.t001

2.3.6 In-solution analysis

In order to increase detection of PM proteins, a urea/SDS extract was also trypsinised in-solution and directly analysed by LC-MS/MS. A minimum of 195 *G. m. morsitans* proteins were identified. Only those with an ion score cut off of 30 or above were considered, with the majority of them having 2 or more identifying peptides and annotated on the VectorBase database (version GmorY1.1, 2013) and *S. glossinidius* genome. Proteins were classified and grouped by functional classifications (Fig. 2.6), according to their GO terms and domain

features as predicted by ExPASy Prosite, VectorBase and EMBL-EBI InterProScan. Hypothetical proteins were classified based on the presence of family domains.

The majority of tsetse proteins (92%) fit into 13 of the categories, whereas 15 proteins (8%) could not be assigned to any category. However, all of these unknown proteins had orthologues in several insects and insect vectors, most of which had either no description or were described as conserved hypothetical proteins, suggestive of being ubiquitous among insects. Of the 195 proteins, 28 contained a predicted signal peptide (SP), 26 were found to contain a transmembrane domain (TM) only, 16 had both a predicted signal peptide and at least one TM and the remainder (125) were predicted to be soluble (i.e. neither SP nor TM domain). Interestingly, one of the most abundant hits corresponded to GmmPer66. Two other novel peritrophins were also discovered: GmmPer12 (GMOY011810) and GmmPer108 (GMOY007191) (See Chapter 3 for more details).

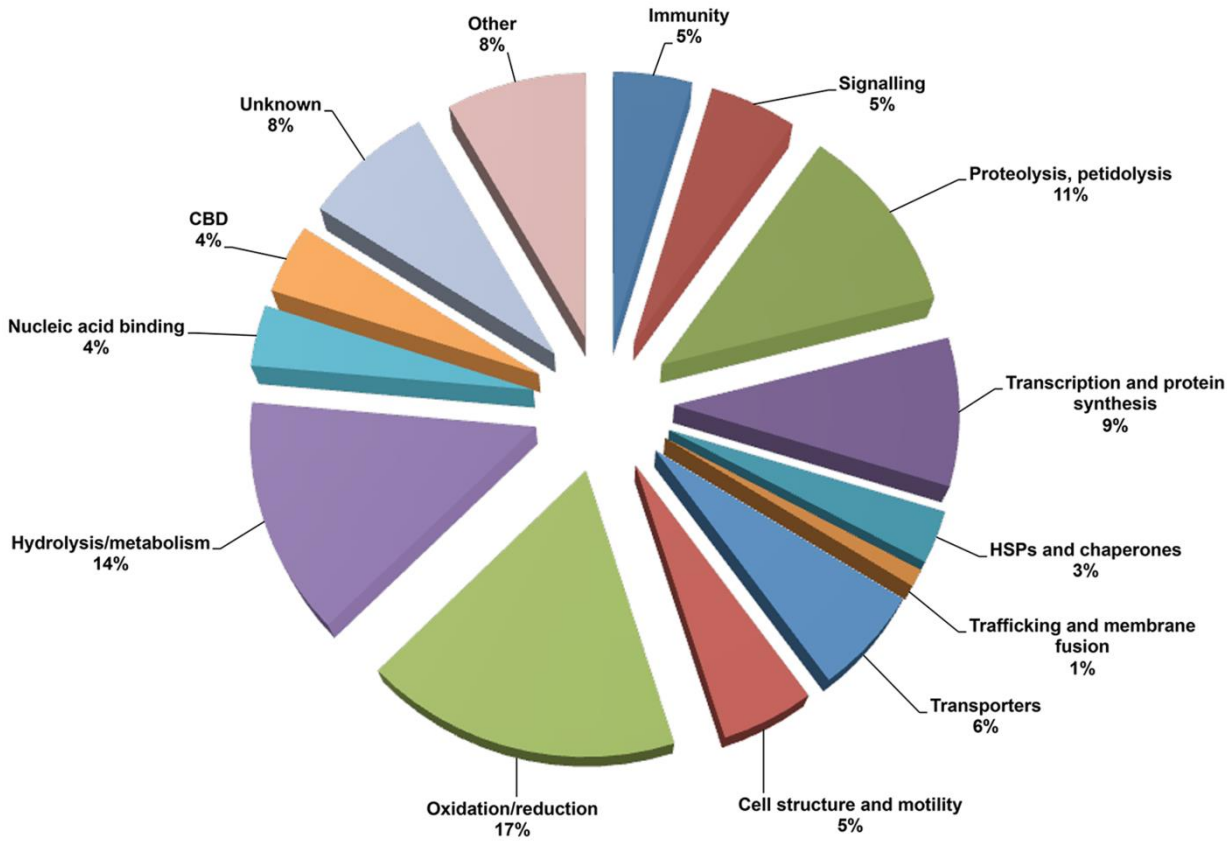


Figure 2.6. **Categorisation of the *G. m. morsitans* peritrophic matrix proteins as identified through LC-MS/MS according to their putative functions.**

2.3.7 Validation of some of the proteins identified by western blotting

Western blot analysis was performed in order to validate some of the protein hits identified in both the in-gel and in-solution digested samples. Tsetse PMs were dissected, washed and solubilised with urea/SDS, processed for Western blotting and ~10 PMs per lane probed separately with several anti-tsetse and two anti-*Sodalis* antibodies.

As shown in Fig. 2.7, western blotting confirmed the presence of one C-type lectin (lane 1), TsetseEP protein (lane 2) and Pro2 (lane 3). In addition, the presence of symbiont proteins were confirmed using an anti-GroEL monoclonal antibody (lane 4), which cross-reacts with the GroEL of *Wigglesworthia glossinidia* and *Sodalis glossinidius*. To confirm that *S. glossinidius*, and not *W. glossinidia,* was isolated with the PM, an anti-*Sodalis* polyclonal antiserum was used (lane 5). This antiserum recognizes a suite of *S. glossinidius* proteins that produces a characteristic banding profile, including GroEL (*Mr* ~60 kDa) (Haines, L., unpublished) (Fig. 2.7).

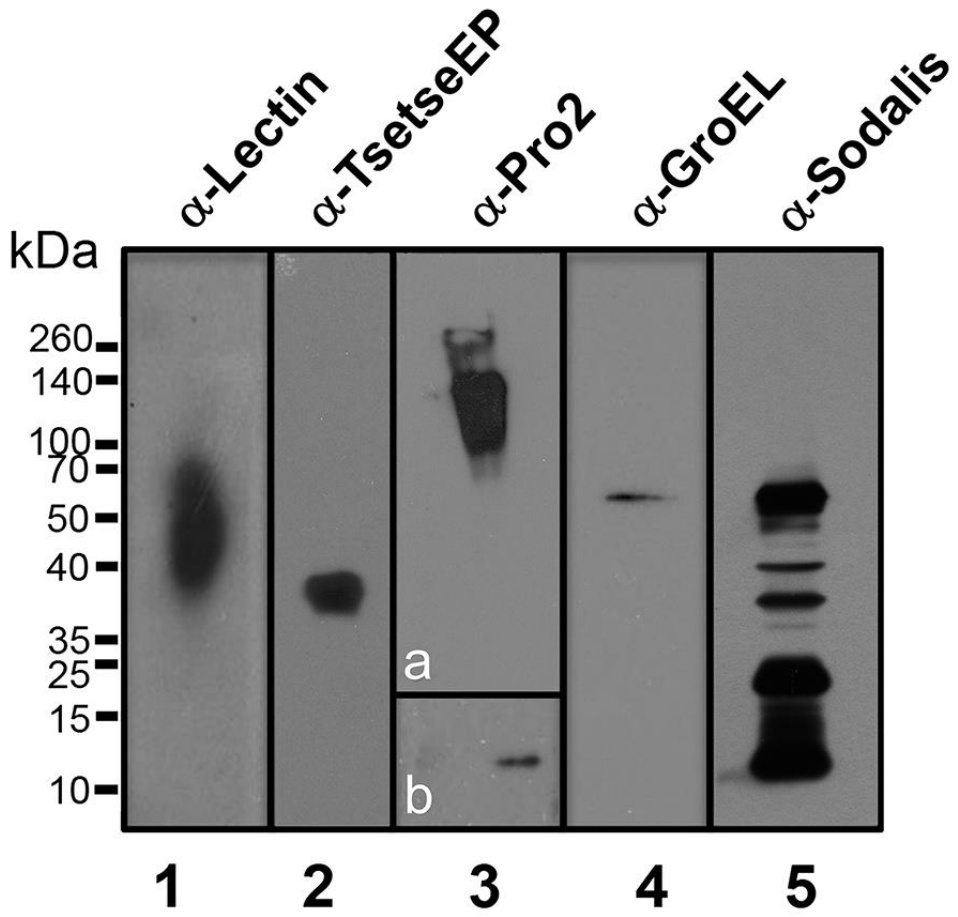


Figure 2.7. Western blotting analysis of tsetse PM proteins. Homogenates from ~10 PM equivalents were loaded per lane and after transferred to PVDF separately probed against an anti-tsetse lectin antibody (lane 1), anti-procyclin mAb 247 (lane 2), anti-Pro2 mAb 4A2

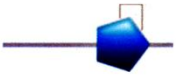
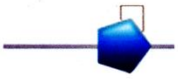
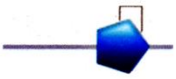
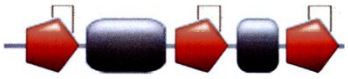
(lane 3), polyclonal anti-GroEL mAb 1H1 (lane 4) and polyclonal anti-*Sodalis* (lane 5) and developed by chemiluminiscence. Developing continued for 30 seconds (lane 1, 2, 3a, 4 and 5) or up to 3 hours (lane 3b).

2.4 Discussion

Many types of arthropods possess a peritrophic matrix. They differ in their production and protein composition and in many cases PM protein extraction can be difficult. This may explain the huge variation in many PM protein studies. For example, there were only two major proteins reported from the PM of adult dipterans *Similium vitatum* and *lutzomyia longipalpis* whereas a study on the PM of *Aedes aegypti* showed there were 20-40 major PM proteins. In addition, the PM proteome of *Glossina morsitans* has previously been shown to comprise of 40 proteins as seen by 2D SDS-PAGE. To date, only a handful of studies has utilised an LC-MS/MS proteomics approach to investigate the protein content of insect PMs, including *anopheles gambiae* where up to 209 proteins where found to be constitutes of the PM. This current study represents the first tsetse proteome PM analysis using a LC-MS/MS approach and may provide insights as to how trypanosomes interact with the tsetse PM during their development in the fly. It has been hypothesised that African trypanosomes may bind to specific receptors before crossing the PM [102]. Binding could facilitate PM crossing by concentrating degrading enzymes at the point of parasite entry. Given their high abundance, it is tempting to suggest that the PM peritrophins may serve as receptors for specific parasite flagella surface ligands.

2.4.1 Analysis of the new tsetse peritrophins

In total, five peritrophins were identified by mass spectrometry analysis from both in-solution and in-gel digestion and as such, has more than doubled the number of previously reported peritrophins from the *Glossina* PM (Fig. 2.8).

Protein ID #[1]	Protein Length	Predicted Signal Peptide[2]	# CBD	Type CBD[3,4]	Peritrophin classification[4]	Structural Organisation[4,5,6,7,8]
GMOY011809 (PRO1)[9]	92	1-20//21	1	Partial PCD	Simple peritrophin	
GMOY009587 (PRO2)[10]	116	None predicted	1	Partial PCD	Simple peritrophin	
GMOY011810 (GmmPer12)	100	1-19//20	1	Partial PCD	Simple peritrophin	
GMOY007191 (GmmPer108)	1040	1-19//20	2	PAD	Binary Peritrophin	
GMOY002708 (GmmPer66)	603	None predicted	3	PAD	Complex Peritrophin	

[1]Protein ID numbers as found in VectorBase
[2]As predicted by SignalP
[3]Predicted using ExPASy Prosite, VectorBase and EMBL-EBI InterProScan. Domains represented in relation to the whole protein and are not to scale. Disulphide bridges demonstrate the potential formation of intradomain disulphide bonds
[4]CBD represented in relation to the whole protein. Types of CBD are shown in the figure and are defined according to Toprak *et al* (2010)
[5]Images created using ExPASy Prosite
[6]The typical arrangement of the 6 cysteine residues for a dipteran PAD is $CX_{11-21}CX_5CX_{9-19}CX_{10-14}CX_{4-16}C$ where X is any amino acid other than cysteine
[7]The typical arrangement of the 8 cysteine residues for a dipteran PBD is $CX_{12-14}CX_{18-21}CX_{10-18}CX_{12}CX_2CX_8CX_{7-12}C$ where X is any amino acid other than cysteine
[8]The typical arrangement of the 10 cysteine residues for a dipteran PCD is $CX_{17}CX_{9-10}CX_{14}CX_9CX_{8-9}CX_{19}CX_{9-11}CX_{14}CX_{11}C$ where X is any amino acid other than cysteine
[9]Protein has been identified and partially characterised as Proventriculin 1 (PRO1) Hao *et al* (2002). NCBI accession # **AAN52276.1,** UniProt accession # **Q8ITJ7**, GeneDB ID ***Gmm-0757*** (now redundant)
[10]Protein has been identified and partially characterised as Proventriculin 2 (PRO2) Hao *et al* (2002). NCBI accession # **ANN52277.1,** UniProt accession # **Q8ITJ6**, GeneDB ID ***Gmm-2445*** (now redundant)

Figure 2.8. Classification and partial characterization of *G. m. morsitans* peritrophic matrix (PM) peritrophin and peritrophin-like proteins, containing 1 or more chitin binding domains (CBD), as identified by LC-MS/MS.

2.4.2 Non-mucin peritrophins

Both GmmPro1 (GMOY011809) and GmmPro2 (GMOY009587) are known to be synthesised in the tsetse PV and secreted during the formation of the PM [85]. The remaining three are novel, and this study is the first to positively identify them as being PM constituents. GmmPer12 (GMOY011810) is a small peritrophin of 100 aa with a predicted molecular mass of ~12 kDa and has a partial Peritrophin C Domain (PCD). Originally, the PCD was thought to consist of 6 conserved cysteine residues [49] with the domain spanning 68-70 residues. Only recently has the PCD been shown to be composed of 120-

121 residues and have a motif of 10 conserved cysteines [103] consisting of $CX_{17}CX_{9-10}CX_{14}CX_{9}CX_{8-9}CX_{19}CX_{9-11}CX_{14}CX_{11}C$ and those peritrophins thought to have a full PCD are now categorized as having partial domains. Partial domains may have come about through multiple duplication events or proteolytic degradation of full length proteins whilst retaining the ability to bind chitin. This proteolysis may occur before or after such CBD proteins have been incorporated into the matrix. Some partial CBDs have been shown to have trypsin and chymotrypsin cleavage sites embedded within the CBDs [104] suggesting these proteins are highly resistant to proteolysis owing to the folded nature of their structure through disulphide bond formation. GmmPer12 has a PCD of 4 conserved cysteine residues similar to that of GmmPro1 and GmmPro2 and is analogous to peritrophins found in other insects such as LcPer15, a peritrophin found in the PM of the sheep blowfly *Lucilia cuprina* [105] (Fig. 2.9). A predicted signal peptide between residues 19/20 suggests that GmmPer12 is secreted into the PM after synthesis. GmmPro1, GmmPro2 and GmmPer12 are related to the peritrophin-15 family of proteins, integral proteins from the PMs of many insects [106]. This protein family is suggested to associate with the PM by binding to the ends of chitin fibrils giving structural support and preventing exochitinase action. The lack of *N*- and *O*-glycosylation on these 3 peritrophins supports this assumption. However, their intact forms appear to be absent in the PM, suggesting these three peritrophins are degraded and incorporated into the PM as partial fragments that have retained their ability to function as a chitin-binding domain. The updated *Glossina* VectorBase genome annotation has revealed that GmmPro2 is not 93 amino acids as previously reported [85] (AAN52277.1), but instead has an extension at its N-terminus making the protein 116 amino acids long. This is perhaps evidence that at least GmmPro2 (and probably also GmmPro1 and GmmPer12) have evolved from a larger protein containing many CBDs (Chapter 3).

2.4.3 Insect intestinal mucins

GmmPer108 (GMOY007191) is a 1040 amino acid long binary peritrophin, which has a conserved insect intestinal mucin (IIM) flanked by 2 Peritrophin A Domains (PAD) and a predicted signal peptide between residues 19/20. The PADs are typically 48-57 residues in length and have a consensus consisting $CX_{11-21}CX_{5}CX_{9-19}CX_{10-14}CX_{4-16}C$ where X is any amino acid other than cysteine. PADs are ubiquitous among all insects and are the most common CBD in dipteran larvae and, presumably, dipteran adults. This type of domain is abundant in invertebrate cuticular proteins and can also be found in other PM associated proteins such as mucins, serine proteases and chitinases [107] since these proteins are also associated with chitin. The mucin domain is rich in serine, proline and threonine residues resulting in multiple *O*-glycosylation sites, which likely contributes to the

peritrophins unusually large mass of 108 kDa. Binary peritrophins have been found in at least 3 lepidopteran families, *Bombyx mori*, *Epiphyas po stvittana* and *Helicoverpa armigera* [103], but until now have only been observed in one dipteran family: 1Chit from *Anopheles gambiae* [108]. This 40.6 kDa protein is expressed in the mosquito midgut and has been shown to be an immunoresponsive gene in adult mosquitoes when bacterially challenged. The large mucin domain in GmmPer108 may help to provide tensile strength to the tsetse PM during blood feeding in addition to providing protection against enzymatic attack due to its potential extensive glycosylation.

GmmPer66 is a complex peritrophin containing 3 PADs interspersed with 2 mucin domains. It is 603 aa in length and has a predicted mass of ~66 kDa. It is the most abundant protein found both in-solution and in-gel analyses. Out of 30 gel bands excised, GmmPer66 was found in 12 bands throughout the gel at molecular masses between 10->170 kDa (Fig. 2.5b). This is suggestive of either protein degradation, (unlikely due to its possible high degree of crosslinking), protein modification or expression of different gene products. Transcriptional analysis suggests that GmmPer66 is expressed exclusively in the PV (Lehane, SM and Lehane MJ, unpublished). Multi-gene families may be required to synthesise sufficient amounts of protein to support rapid PM synthesis, which can be up to 5-10 mm/hr in certain insects with a type II PM [109]. Unusually for peritrophins, GmmPer66 does not contain a predicted signal peptide or transmembrane domain indicating that this protein possibly enters a non-classical secretory pathway before association with the PM. Further characterization of these three novel peritrophins should provide additional insights into the molecular function of the tsetse PM.

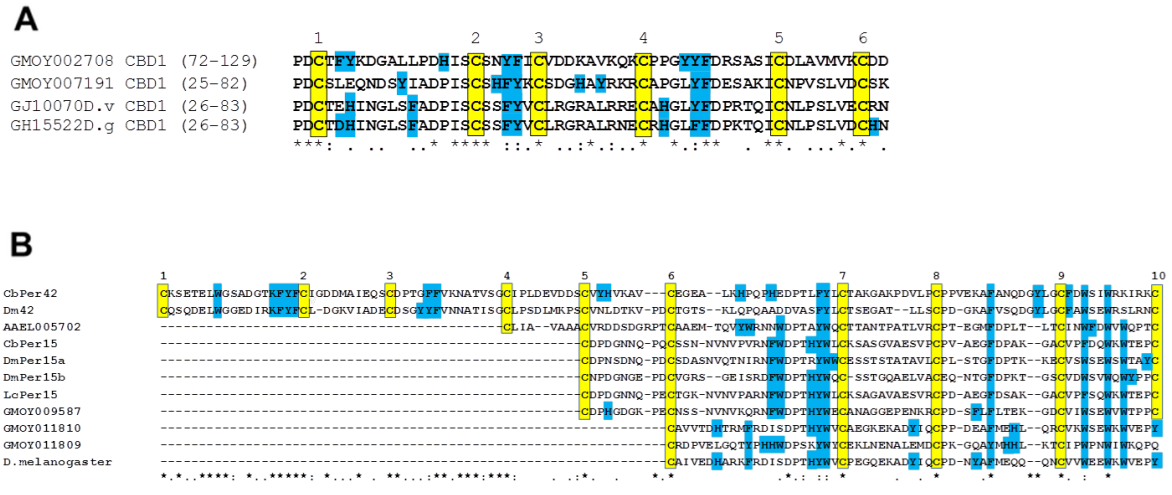


Figure 2.9. **Chitin Binding Domains of tsetse peritrophins as identified by LC-MS/MS aligned against other representative domains from putative Dipteran peritrophins.** The positions of the domains within the sequence are shown next to their respective protein

IDs for panel A only. The 6 and 10 conserved cysteine residues of a PAD (A) and PCD (B) respectively, which are potentially involved in disulphide bridge formation, are indicated by yellow boxes and asterisks (*). Note the partial PCD of many peritrophins including those in *Glossina.* The numbers above the cysteine residues depict the order of cysteines in the CBD. The conserved aromatic residues, characteristic of chitin binding domains, are denoted by blue boxes and may be involved in carbohydrate binding. Identity of amino acid residues is depicted according to ClustalW.

2.4.4 Non-structural PM proteins

The majority of the proteins identified were hydrolytic enzymes including chitinases, amylases, exopeptidases and digestive enzymes such as trypsin. Although these may be midgut secreted proteins and only transiently associated with the PM, studies have shown these enzymes remain in the PM even after repeated washes and extraction with strong denaturants [82]. A tsetse Chitinase (Cht1) was identified from both in-gel and in-solution analyses. Chitinases have been found associated in the PM of lepidopteran larvae where they are involved in the larvae moulting process [110] and are also found in the PM of adult mosquitoes where their role is less understood [111]. It has been suggested that during insect growth and development, chitin containing structures require the capacity to undergo remodelling and modification in order to allow for growth, maturation and repair [112]. This is especially true under certain conditions such as periods of moult or starvation where PM production can stop. In order for this to happen, tissue specific chitinolytic enzymes and chitin synthases are produced periodically. Chitinases are important in both the shedding of the cuticle during moults and growth and for the degradation and turnover of both the PM and trachea [112]. The fact that chitinases have now been identified in the tsetse PM suggests that PM chitinases in adult tsetse may be involved in degradation of the chitin fibrils thereby modifying the thickness, porosity and tensile strength of the PM during its extension along the length of the midgut.

As expected, a large percentage of proteins (11% from the in-solution digestion) were digestive enzymes such as trypsins, chymotrypsins, peptidases and serine proteases. Their identification may simply reflect transit across the PM to the endoperitrophic space in response to a blood meal, but given the fact that the flies used in this study did not receive a blood meal, it is possible that these enzymes are directly interacting with the PM in anticipation of feeding. These findings add to the reputable evidence that the PM improves digestion by concentrating the food bolus and filtering out indigestible components [113]. One serine protease, Proventriculin 3 or GmmPro3, previously shown to be expressed in

the tsetse PV, was also identified in this study suggesting that it might be physically associated to the PM. GmmPro3 is homologous to proteins of the serine protease S3 family and shares similarities with serine proteases from other haematophagus insects such as *Stomoxys calcitrans* and *A. gambiae* [85]. One serine protease inhibitor (Serpin), GmmSpn4, was also identified from the PM suggesting that serine proteases and serpins have a co-relationship involving blood meal digestion and may also modulate the PM structure until it is fully formed. Finally, proteases may also protect the passage of pathogens through the PM. In fact, the surface of procyclic trypanosomes gets "re-shaped" due to extensive proteolysis of the main surface glycoproteins, procyclins [114], which partially may occur during PM crossing.

2.4.5 Proteins involved in immunity

From the in-solution analysis, 9 proteins (~5%) were identified as being involved in host-parasite interactions. These proteins were mainly C-type lectins (CTLs), whose presence in the PM was corroborated by Western blotting (Fig. 2.7). CTLs are Ca^{2+}-dependent glycan binding proteins and play important roles in insect defence [115]. Carbohydrate binding events mediate a range of processes including cell/cell interactions, cell adhesion and are involved in cell apoptosis. They are also capable of recognizing pathogen-associated molecular patterns in a variety of microbes and in tsetse it has been suggested to be involved in the initial elimination of trypanosome burden by agglutinating parasites [116], although so far no experimental evidence has proved this.

Interestingly, from the in-gel analyses there were many hits for basement membrane-specific heparin sulfate proteoglycan core protein (perlecan). Perlecan is a large proteoglycan with a multitude of diverse domains [117]. These domains bind to and cross-link numerous extracellular components and cell surface molecules. The N-terminal domain consists of ~195 aa and contains three Ser-Gly-Asp attachment sites for large heparin sulfate chains or, occasionally, chondroitin sulfate. There is microscopy evidence showing that the *G. m. morsitans* PM contains glycosaminoglycan's (GAGs) in the layer facing the ectoperitrophic space (epithelium side) [39], suggesting that this may be the location where perlecan may accumulate after secretion. Other domains include immunoglobulin, laminin and low-density lipoprotein (LDL) receptors that contain multiple cysteine residues able to form disulphide bridges. Perlecan also has an epidermal growth factor (EGF) domain, which is involved in ligand-recognition and protein-protein interactions. It is possible that identification of this protein is due to basement membrane contamination, however, if

perlecan is a true PM protein, this may explain why proteins such as collagen, actin, lamin, laminin and fibronectin are found in a number of PM proteome studies [118, 119]. It would be interesting to determine the exact place of perlecan synthesis.

TsetseEP protein was also identified in both the in-gel and in-solution analyses. This is a unique tsetse protein of *Mr* ~36 kDa, which contains a characteristic extended glutamic acid-proline (EP) repeat domain at the C-terminus. Interestingly, its structure resembles that of the *T. brucei* EP-procyclins [14, 97, 100]. Studies have shown that TsetseEP probably acts as an antagonist to trypanosome infection [100]. TsetseEP is also highly upregulated in flies that have been challenged with gram-negative bacteria, which would suggest this protein may have an immunoprotective role [102]. The finding of TsetseEP in our analyses is intriguing. Although secretion of this molecule is enhanced by the presence of pathogenic microorganisms and it contains a lectin domain that may directly interact with pathogen's surface glycans, its elevated production during a midgut infection may also contribute to PM thickening, thus creating a stronger protective barrier. In *Drosophila*, there is genetic evidence showing that the PM structure changes in the presence of pathogenic bacteria [70, 79]. In addition, an interesting recent work has shown that the *Glossina* PM becomes thinner in aposymbiotic flies (i.e. lacking a midgut microbiome), which in turn increases PM permeability and allows an "easier" passage of trypanosomes through the PM [83]. Thus, although it remains to be determined how the structure of the tsetse PM changes in response to either pathogenic or non-pathogenic organisms, it may be possible that TsetseEP has a role in PM remodelling.

Tsetse antigen 5 (Tag5) was identified from the in-solution analysis. This protein of 259 amino acids is related to the large Crisp-Antigen 5 Plant pathogenesis protein families that are found in a huge diversity of organisms [120]. Mostly found in saliva of many insects, these proteins share a core sequence of approximately 200 amino acids that are responsible for their multiple functions. Antigen 5 has been proven as a potent venom allergen in hornets, wasps and fire ants and causes allergic reactions in humans [121, 122]. Although primarily found in the salivary gland tissue of tsetse, it is reported to be expressed in the PV and midgut tissues [123]. A related protein in *Drosophila*, Antigen 5 related (Agr), is also expressed in the PV of both larvae and adult flies [124, 125]. Tag5 has also shown to be upregulated in a susceptible strain of tsetse (salmon flies) [86]. Tag5 may be a true constituent of the tsetse PM and as such may have a bearing on the digestion of the bloodmeal as studies have shown Tag5 prevents homeostasis [126]. As tsetse take up to 3 days to digest a bloodmeal, it is possible that the presence of Tag5 in the PM prevents the ingested bloodmeal from clotting quickly, thus aiding and facilitating digestion.

Another protein identified and involved in immunity was glycoprotein CD36, whose family members are conserved within mammals and have many representative orthologues in

insects. They have a variety of functions including lipid transport, immune regulation, homoeostasis and adhesion. One function of CD36 is as a scavenger receptor, which recognizes molecular patterns presented by bacteria, pathogens and viruses and also pathogen infected cells [127, 128]. An orthologue of CD36 in *C. elegans*, CO3F11.3, is responsible for mediating host defences against fungal infection by stimulating the production of cytokines [129]. As a PM constituent, CD36 may have multiple roles from anti-homoeostasis to immune system mediation possibly involved in initial clearance of pathogens. In addition, this protein is highly resistant to proteolysis, which would be favourable given its putative location.

Hemomucin, a 61.7 kDa protein containing extensive *O*-glycosylation at its C-terminus was also identified. It contains a domain showing strictosidine synthase, which is a key enzyme in alkaloid biosynthesis. Alkaloids are important in the immunity of plants and have been shown to be secreted in the venom of the fire ant where they act as potent inhibitors of bacteria [130]. Hemomucin from *Drosophila* proved likely to be involved in induction of antibacterial effector molecules after showing affinity for the snail lectin (*Helix pomatia* hemagglutinin A). This protein was found to be expressed in the PV, suggesting that it may be incorporated into the PM after synthesis [131].

2.4.6 Other proteins

Proteins involved in stress response (oxidation and reduction) and protein folding (heat shock and chaperones) comprises a total 20% of the detected proteins. Some of these proteins may originate from the layer of epithelial cells that is in close proximity with the PM of teneral (unfed) flies. However, they may have a role in detoxification. Bloodmeal digestion leads to the rapid production of reactive oxygen species (ROS), due to the breakdown of red blood cells, which causes the release of haem and iron. Accumulation of free haem leads to oxidative stress and these oxidation/reduction proteins are needed to detoxify the midgut environment [132]. It has been demonstrated in female *Aedes aegypti* that the PM of these insects are capable of binding haem during bloodmeal digestion as shown by histochemical studies [68] and a subsequent study has shown that at least one PM protein, the peritrophin AeIMUCI, is responsible for this interaction [69]. Haem-regulatory motifs (HRM) have also been found in peritrophins from two species of sandfly, *Phlebotomus papatasi* and *Lutzomyia longipalpis* [79].

2.4.7 *Sodalis* proteins

A total of 27 *S. glossinidius* proteins were identified (Table 2.2). Given that *Sodalis* proteins have been identified within the PM and their presence verified by Western blotting (Fig. 2.7) suggests that secondary symbionts are intimately associated with the tsetse PM. Alternatively, these proteins may be secreted and incorporated into the PM, and thus they may have a functional role. The majority of these proteins were found to relate to metabolic activities within the bacteria. It has been well documented that *Sodalis* are important for many aspects of tsetse metabolism for example, cofactor and vitamin synthesis to compensate for the restricted diet of blood-meals [133]. Genes encoding biotin, lipoic acid, molybdenum cofactor, riboflavin and folic acid have all been found to be present in the genome of *Sodalis glossinidius* [134].

Table 2.2. List of Sodalis glossinidus proteins associated with the Glossina morsitans morsitans peritrophic matrix

Protein ID[1]	Protein Name	In-gel[2] (# peptides)	In-solution (# peptides)
SG0306	GroEL 60 kDa chaperonin[3]	(1)	(3)
SG0058	FliC flagellin	N.D.[4]	(2)
SG1007	OmpF outer membrane pore protein[5]	N.D.	(2)
SG1030	OmpA outer membrane protein A[5]	N.D.	(2)
SG1474	Putative chitinase	N.D.	(2)
SG2412	AtpA ATP synthase subunit alpha	(2)	(1)
SG0131	RplA 50S ribosomal protein L1	N.D.	(1)
SG0127	Tuf Elongation factor Tu (EF-Tu)	(1)	(1)
SG1759	HisS Histidyl-tRNA synthetase	(1)	(1)
Ps_SGL0763	Putative phage integrase	(1)	(1)
SG0134	RpoB DNA-directed RNA Pol subunit β	N.D.	(1)
SG0135	RpoC1 DNA directed RNA Pol subunit γ	N.D.	(1)
SG2230	GlnA glutamine synthetase	N.D.	(1)
SG1265	MsbB lipid A biosynthesis acyltransferase	N.D.	(1)
SG0887	peptidoglycan-associated lipoprotein precursor	(1)	(1)
Ps_SGL0342c	Putative exported protein	(1)	(1)
SG1015	AsnS asparaginyl-tRNA synthetase		(1)
SG0430	ATP dependent helicase HepA	(2)	

Ps_SGL0112c	Periplasmic trehalase precursor	(2)
SG0482	Hypoxanthine phosphoribosyltransferase	(1)
Ps_SGL1307c	CheA chemotaxis protein	(1)
SG0955	LysP lysine specific permease/transporter	(2)
SG0087	RpoH RNA polymerase factor sigma 23	(1)
SG2405	tRNA uridine 5-carboxymethylaminomethyl modification protein GidA	(1)
SG0284	N-acylhomoserine lactone synthase Ypel1	(1)
SG1864	Multidrug efflux transport protein	(1)
Ps_SGL1035c	Putative N-ethylmaleimide reductase	(1)

[1]Genome downloaded from Eugeni, E. et al., (2010). doi:10.1186/1471-2164-11-449

[2]As shown in Fig 2.5b

[3]Its presence in the PM was validated by Western blotting (Figure 2.7)

[4]Not detected

[5]The presence of both OMP isoforms has been also detected by Western blotting using an anti *Sodalis* OMP polyclonal serum (data not shown).

One interesting protein identified by mass spectrometry analysis was the *Sodalis* putative chitinase (Accession No SG1474). Studies have shown that when flies harbour a high density of *Sodalis,* they are more susceptible to trypanosome infection, thus it is entirely feasible to assume that these endosymbionts confer susceptibility to tsetse [135-137]. One possible explanation for this is that *Sodalis* may degrade chitin fibrils that comprise the tsetse PM, effectively remodelling it and providing an opportunity for trypanosomes to penetrate [135, 138, 139]. The primary carbon source during the growth of *Sodalis* is *N*-acetyl-β-D-glucosamine (a monomer of chitin), which it produces from the breakdown of chitin using a secreted chitinase. Given that trypanosomes have no chitinase activity, it is reasonable to speculate that *Sodalis* breaks down PM chitin, leaving the PM vulnerable and unknowingly facilitating trypanosome crossing (reviewed in [140]). In addition, the prevalence of trypanosome infection is highest when the fly is young and the PM is not yet fully formed. Proteins containing CBDs such as the peritrophins may have not yet been fully incorporated into the PM leaving the ends of PM chitin fibrils exposed. This may be the critical time point of chitinase activity, thereby degrading the PM and allowing trypanosomes to break through. Other parasites like *Brugia malayi*, *Leishmania spp* and *Plasmodium spp* secrete chitinases and proteases to degrade the proteins within the chitin meshwork and allow penetration of the PM [46, 141, 142]. Although the quantity of chitin has yet to be measured in the *G. m. morsitans* PM, the lack of chitinase expression in trypanosomes

suggests that the chitin content of the tsetse PM may be probably low as reported in *Lucilia cuprina* larvae (which also express a type II PM). Therefore, the tsetse PM chitin may not a real barrier to trypanosome infection [143].

2.4.8 What does the *G. m. morsitans* PM proteome tell us about its possible architecture?

Contrary to type I PMs, there is no molecular model representing the architecture of type II PMs. In the case of the *Glossina* PM, it is challenging to predict a model considering that it is composed of three layers [39] (each one of different thickness and also in composition) and because of the lack of EM localization of major peritrophins. However, based on its high abundance, number of CBDs and mucin domains, we hypothesise that GmmPer66 may play an essential role in interconnecting chitin fibres with other GmmPer66 monomers and/or other PM peritrophins, like GmmPer108 (with 2 CBDs and 1 mucin domain) or Pro2 (with one CBD and several *O*-glycosylation sites). As suggested for other highly glycosylated molecules, the *O*-glycans from these peritrophins may serve to protect the PM from protease attack and retain water thus allowing the selective trafficking of molecules between the lumen and the ectoperitrophic space. It is conceivable that other peritrophins are also part of the tsetse PM, but their identification by MS was missed due to their resistance to trypsin. In fact, the *G. m. morsitans* genome contains a minimum of 36 putative peritrophins (Chapter 3) [144].

2.5 Conclusion

The study presented here has given a comprehensive overview of the main proteins that make up the tsetse PM identified using mass spectrometric techniques. Identification of at least 209 proteins from in-solution analysis and many more from in-gel analysis has provided a foundation of knowledge for which there is potential to develop. The identification of 3 novel peritrophins has expanded the list of known tsetse PM peritrophins from 2 to 5. In addition, the unique banding pattern of one of these peritrophins, GmmPer66, has provided us with useful insights into how their putative degree of crosslinking and how they are potentially incorporated into the PM. Although the quantity of chitin in the PM of *G. m. morsitans* has yet to be confirmed, the lack of chitinase activity in procyclic trypanosomes would suggest that the chitin component of the tsetse PM is extremely low and that chitin is not a real barrier to infection, proposing that the PM is

composed mainly of glycoproteins rather than chitin. Therefore, a direct degradation of integral proteins may provide a pathway for trypanosome invasion through the PM. We are currently investigating candidate trypanosome proteases that may be participating in PM degradation. However, for proteases to act, glycosidases must first remove glycans (i.e. chitin and GAGs). Therefore, it is intriguing that procyclic trypanosomes do not express the glycosidases to degrade any of these complex sugars. Alternatively, some of the PM-degrading glycosidases may be supplied by bacterial symbionts present in the tsetse midgut. With the completion of the *Glossina* genome project, a collaborative effort involving the VectorBase community and the Sanger Centre, there is great potential to reveal novel concepts about type II PMs. Insects with a type II PM are often more refractory to infection that those with a type I PM such as mosquitoes and sand flies. Whilst huge efforts have gone into researching larval type II PMs, this is the first study to concentrate on the protein composition of the adult type II PMs from an insect vector.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [145] with the dataset identifier PXD000594 and DOI 10.6019/PXD000594

Chapter 3. Comparative Peritrophin Analysis and Annotation of orthologous genes across major *Glossina* vectors

An African adventure……..

[2]3.1 Introduction

Tsetse flies (Family, Glossinidae) are the sole vectors of African trypanosomes and are responsible for the transmission of the parasites which cause devastating illnesses in humans (sleeping sickness) and livestock (nagana). The impracticability of rearing livestock in endemic areas causes economic losses of several billion US$ per annum and affects over 50,000 people in 36 countries of Sub-Saharan Africa [146]. Although trypanosomes are one of the best studied organisms and are indeed a model organism for kinetoplastids, advances in drug and vaccine development against the disease have been stalled mainly due to the huge repertoire of antigenic variation displayed by trypanosomes [147]. In addition, treatment for HAT is risky due to the dangerous side effects from the toxic drugs used as well as an increase in the number of reports of trypanosome resistance to treatment [148]. Thus, with chemical warfare against the parasites deemed unfeasible, control of the disease is mainly down to vector control which has proven to be the most efficient when bringing outbreaks of the disease under control [72]. Indeed, the African Union has envisaged that vector control will alleviate the burden of HAT so effectively that the disease will be eradicated by 2031 with only 7 countries in Africa still plagued by the disease by 2010 [149]. Although eliminating Animal African Trypanosomiasis (AAT) will prove much harder than eradication of Human African Trypanosomiasis (HAT), vector control nonetheless will be paramount in the success of relieving the burden of the disease and ensuring that the reclaimed land areas are sustainably and economically exploited. This will allow many of the most marginalised people in Africa a chance to overcome poverty constraints and help alleviate food poverty. Tsetse's unique life-cycle and biology set them aside from other Dipterans providing a weak-link that can be exploited and may give rise to alternative control strategies. In addition, identifying differences between numerous Glossina species and related Dipterans such as *Drosophila melanogaster* and *Musca domestica* (a non blood-feeding mechanical vector of various communicable diseases of humans), may provide invaluable insights into genes that are important for multiple aspects of tsetse biology. Even within the same genus, different species occupy different niches, have specific host seeking and feeding behaviours and have slightly different physiological conditions, all of which are a contributing factor to the vectoral competence of each species (Ch3_SI1). The recently sequenced genome of *Glossina morsitans morsitans* has provided multiple genomic and functional biology findings that reflect the unique physiology of this disease vector, some of which are summarised below;

[2] Parts of this chapter have been published in Science; DOI: 10.1126/science.1249656

3.1.1 Olfaction and chemosensory

In comparison to *Drosophila* and several mosquito species, there was found to be an overall reduced number of odorant, gustatory and ionotrophic receptor genes [150]. Unlike Glossina, who are obligate blood feeders, the absence of sweet-tasting gustatory receptors that can be found in all other Diptera may reflect their feeding habits of taking both blood and nectar meals and males of other species that do not feed on blood.

3.1.2 Reproductive biology

There was found to be a reduced number of yolk protein genes in Glossina, indeed only one was found to be from the yolk protein family which reflects the tsetse flies unique vivaparity. In comparison, 12 tsetse-specific milk proteins comprising transferrin, lipocalin, acid sphingomyelinase and a novel 9-gene family were identified (the highest number of milk proteins of any insect to date) and are thought to be vital in the intrauterine larval survival [140, 151, 152]. .

3.1.3 Salivary and digestion

Approximately 250 salivary proteins including many anticoagulants and proteins for host immune suppression such as nucleic acid binding proteins and adenosine deaminase-related growth factors were found to comprise the tsetse saliva. In addition, there was shown to be a global reduction of proteins in infected glands [153].
A total of ten aquaporin genes were identified which is the most of any characterised insect [154]. These genes are required for water transport and are involved in diuresis which may be implicated in the vectoral capacity of the fly (Chapter 5).
Approximately 300 PM proteins were found to be constitutes of the tsetse PM (Chapter 2) through mass spectrometry [59]. Among these, five peritrophins were identified of which, 3 have never been characterised. By mining the genome, a further 31 peritrophin-like proteins were identified (a total of 36).

3.1.4 Tsetse Peritrophins

As the tsetse peritrophic matrix (PM) imposes a physical barrier to trypanosomes (Chapter 1) and peritrophins are found to be important PM components (Chapter 2 and 4), comparative analyses of peritrophins across 5 tsetse species a non-blood feeder vector, *Musca domestica*, and a non- blood feeder non-vector *Drosophila melanogaster* may provide insights into the broad range of vectorial competence and differences in vectorial

capacity. In addition, the chitin binding domains (CBDs) of insect peritrophins differ and these differences may lead to a better understanding of the evolution of peritrophins. The Peritrophin-A-Domains (PADs) are the most common among all insect species, whereas the Peritrophin-B-Domains (PBDs) and the Peritrophin-C-Domains (PCDs) have been so far only been found in a few Dipteran insects. PADs are usually indicative of peritrophins, i.e. proteins of the peritrophic matrix but can also be found in various other proteins such as chitinases, chitin deacetylases, cuticular proteins and cuticular proteins analogous to peritrophins (CPAPs). However, these proteins also contain additional signature domains whereas peritrophins contain only CBDs with some also containing mucin domains. In addition, peritrophins containing PADs can also be found in insects that lack a PM, such as the cat flea and triatomine bugs, where they are expressed in certain tissues. Although there have been functional studies on peritrophins that are localised to the PM, for many insect species there is a distinct lack of data regarding the numbers and types of peritrophins which could provide insights into the insects physiology. There is also no known 3D model of any PADs, although it has been inferred to compare to the CBD found in tachycitin, an antimicrobial peptide from the horseshoe crab, *Tachypleus tridentatus.*

This chapter aimed to identify Glossina peritrophins through their characteristic chitin binding domains and compare them with orthologues of comparative species. Structural organisation of each putative peritrophin was determined for each species and placed into one of four peritrophin groups; simple, binary, complex or repetitive. In addition, phylogenetic analysis and a protein modelling analysis was carried out on the CBDs from known *Glossina morsitans* peritrophins that are constitutes of the peritrophic matrix (Chapter 2).

3.2 Methods and materials

3.2.1 *Glossina morsitans* genomic sequencing, sequence assembly and gene annotation

An international collaboration between 146 research scientists (collectively known as the International *Glossina* genome Initiative (IGGI)) contributed to the sequencing, assembly and annotation of the *Glossina morsitans* genome. In depth details can be found at http://www.sciencemag.org/content/344/6182/380/suppl/DC1. Briefly, numerous sequencing methods were adopted on gDNA extracted from larvae and male and female adults and adult female tsetse devoid of symbiotic bacteria. This produced a high yield of sequences that were subsequently assembled into 13,807 scaffolds. Based on automated and manual annotations, the 366 Mb genome of *G. morsitans* was predicted to contain 12,308 protein-encoding genes. Clusters of orthologous proteins were compared against complete, known genomes from other Dipteran species, namely *Drosophila melanogaster*, *Aedes aegypti*, *Anopheles gambiae*, *Culex quinquefasciatus,* and *Phlebotomus papatasi*. Gene model corrections and renaming were performed via the Artemis sequence viewing and annotation software and through Web Apollo. Automatic and manual annotations were merged into a single gene set. Community based manual annotations were favoured over automatically generated ones and the final gene set was released in VectorBase (https://www.vectorbase.org/) as gene set GmorY1.2. The most recent gene set was updated in 2014 as GmorY1.4 (https://www.vectorbase.org/organisms/glossina-morsitans/yale/GmorY1.4).

3.2.2 Manual annotation and identification of proteins containing CBDs in *Glossina morsitans*

Glossina sequence and annotation data were obtained from the Community Annotation Portal at VectorBase. A search was carried out on proteins that contained chitin binding domains (CBDs) based on domain consensus and homology based searches by BLAST against characterised dipteran and insect gene sequences. Three different types of CBD were searched for; Peritrophin A Domains (PAD) which have 6 conserved cysteine (C) residues in a consensus of $CX_{13\text{-}20}CX_5CX_{9\text{-}19}CX_{10\text{-}14}CX_{4\text{-}14}C$, Peritrophin B Domains (PBD) which have 8 conserved cysteine residues in the consensus of $CX_{12\text{-}13}CX_{20\text{-}21}CX_{10}CX_{12}CX_2CX_8CX_{7\text{-}12}C$ and Peritrophin C Domains (PCD) which have 10 conserved cysteine residues with the consensus $CX_{17}CX_{9–10}CX_{14}CX_9CX_{8–9}CX_{19}CX_{9–11}CX_{14}CX_{11}C$, where X is any amino acid other than cysteine. Proteins containing partial PCDs were also searched for as many characterised peritrophins contain these domains. In order to be determined as putative peritrophins, conserved aromatic residues namely tryptophan (W), phenylalanine (F) and tyrosine (Y) but also histidine (H), had to be present between the 2nd

and 3rd cysteine residues and/or between the 4th and 5th cysteine residues. Additionally, a GO-term search for chitin binding proteins (GO: 0008061) was carried out using the VectorBase Ontology browser to ensure no proteins containing CBDs were missed during the initial search. Paralogues and orthologues to *G. m. morsitans* proteins containing a CBD but were not peritrophins, i.e. chitinases, chitin deacetylases, cuticular proteins or cuticular proteins analogous to peritrophins (CPAPs) were removed from the list. Moreover, those proteins that had no associated metadata but contained other domains other than CBDs were also dismissed as putative peritrophins.

3.2.3 Identification of proteins containing CBDs in other major *Glossina* species and comparative non blood-feeders *Musca domestica and Drosophila melanogaster*

Glossina and *Musca* orthologues to *G. m. morsitans* putative peritrophins were searched for through VectorBase using the BLAST tool against the following proteomes; *Glossina austeni* (gene set GausT1.2), *Glossina brevipalpis* (gene set GbreI1.2), *Glossina fuscipes* (gene set GfusI1.2), *Glossina pallidipes* (gene set GpalI1.2) and *Musca domestica* (gene set MdomA1.1). Additional putative peritrophins for each gene set that were not in the *G. m. morsitans* genome but contained signature CBDs were included and were applied to the same stringencies as shown above. Protein sequences for each *G. m. morsitans* putative peritrophin were used to conduct a BLAST search in FlyBase against the *Drosophila melanogaster* proteome. Again, after omitting orthologues already found, additional peritrophins were searched for using the methods described above and also by searching against a list of putative structural constitutes of the peritrophic matrix (GO:0016490) in *Drosophila* as inferred by prediction or experimental evidence.

3.2.4 Computational analysis

3.2.4.1 Phylogenetic analysis of known peritrophins from *G. m. morsitans* against orthologues in other *Glossina, Musca* and *Drosophila* species

In order to investigate the evolutionary relationship of the known *G. m. morsitans* peritrophin genes to other Dipteran species, a phylogenetic analysis of *G. m. morsitans* peritrophins GmmPer66 (GMOY002708), GmmPer108 (GMOY007191), GmmPer12 (GMOY011810), Pro1 (GMOY011809) and Pro2 (GMOY009587) was carried out using their corresponding genes from other species obtained from VectorBase and FlyBase. Analysis was carried out using the web-based service Phylogeny.fr, using MUSCLE for alignments and optimised with Gblocks. Dendograms were constructed using the maximum likelihood tree implemented in the PhyML v3.0 programme [155, 156]. Confidence values

for each internal branch was determined using the bootstrapping method (500 replicates) and graphical representation was performed with TreeDyn [157].

3.2.4.2 Structural analysis of CBDs from known *G. morsitans* PM peritrophins

Homology modelling was used to predict the three dimensional (3D) structures of the CBDs from 5 peritrophins known to be components of the tsetse PM (Chapter 2); GMOY002708 (GmmPer66), GMOY007191 (GmmPer108), GMOY009587 (Pro2), GMOY011809 (Pro1), and GMOY011810 (GmmPer12). The PADs associated with each protein were uploaded into Phyre2 with tachycitin (Protein Data Bank (PDB) code 1dqc) used as a template structure to model PADs. Homology prediction was also carried out by I-Tasser for increased confidence of structure before refinement of the predicted model was carried out through ModRefiner. Final structure predictions were modified in PyMol

3.2.4.3 Disulphide bond prediction for cysteine residues of PADs

PADs from orthologous known PM peritrophins across *Glossina* and *Musca* species were uploaded into the DISULFIND S-S bond prediction software v1.1. Only those predicted S-S bonds were considered if they had a confidence value of 6 or more (where 0=lowest confidence and 9=highest confidence). The prediction of the disulphide bonds in tachycitin were used as a guide and to ensure a more accurate prediction, PAD sequences were also uploaded into the following S-S bond prediction software packages; EDBCP (Ensemble-based Disulfide Bonding Connectivity Pattern), CYSPRED and CysCon. Only those bonds that were predicted in agreement between each of the software packages were taken into consideration for the homology modelling.

3.3 Results

3.3.1 Characterisation of putative peritrophins containing CBDs in *Glossina morsitans*

Proteins containing CBDs are found throughout the animal kingdom but are more commonly found in insects as structural components of certain tissues or as functional domains of enzymes. CBDs can be arranged into groups depending on the number of conserved cysteine residues present in the domain. PADs are by far the most common both within Lepidopterans and Diptera, with only a handful of Dipteran species identified to have proteins containing either a PBD or a PCD [85, 158, 159]. Those peritrophins that also contain mucin domains are sometimes referred to as insect intestinal mucins (IIMs). Through a combination of proteomics (Chapter 2) and mining of the recently released *Glossina morsitans morsitans* genome, it was possible to identify 36 proteins that contained signature CBDs with some also containing mucin domains.

Table 3.1 shows a list of putative peritrophins identified from the *G. m. morsitans* genome (VectorBase ID numbers) and additional information regarding the potential signal peptides, types and structural organisations of CBDs and any putative post-translational modifications have been included. Four types of peritrophins can be determined by the structural organisation of CBDs within the protein. Simple peritrophins are those that consist of one or two CBDs which may contain small mucin domains. Binary peritrophins consist of a single large mucin domain flanked by one to multiple CBDs. Complex peritrophins are those that have multiple mucin domains (large and small) interspersed with CBDs and repetitive peritrophins are those that consist of multiple CBDs (3 or more) but do not contain any mucin domains. *G. m. morsitans* peritrophins consist of 25 simple, 1 binary, 4 complex and 6 repetitive peritrophins (Fig. 3.1a). Over half (19) of the peritrophins contained a predicted signal peptide and all but three identified peritrophins contained PADs. Those peritrophins without a PAD contained partial PCDs and have been found through mass spectrometry to be constitutes of the tsetse PM (Chapter 2). Two of them, (GMOY009587 or Pro2 and GMOY011810 or Pro1) have also been described and partially characterised elsewhere. Of all the peritrophins identified, 11 contained predicted mucin domains and most of the peritrophins contained fewer than 4 CBDs. The exceptions include a repetitive peritrophin (GMOY003840) containing 6 PADs and a complex peritrophin (GMOY005278) which contains 13 PADs.

Table 3.1. List of peritrophins and peritrophin-like proteins, containing 1 or more chitin binding domains (CBD), from Glossina morsitans morsitans. [1]Protein has been identified and partially characterised as Proventriculin 1 (PRO1). NCBI: **AAN52276.1,** UniProt: **Q8ITJ7**, GeneDB ID: ***Gmm-0757*** (now redundant). [2] Protein has been identified and partially characterised as Proventriculin 2 (PRO2). NCBI: **ANN52277.1,** UniProt: **Q8ITJ6**, GeneDB ID: ***Gmm-2445*** (now redundan
t). NP = non predicted. Highlighted IDs depict those that were found through mass spectrometry (Chapter 2).

Protein ID #	Protein length (aa)	Signal Peptide	# CBD	Type CBD	Peritrophin classification	*O*-glycosylation	*N*-glycosylation
GMOY000880	173	1-17//18	1	PAD	Simple	NP	NP
GMOY000968	254	1-26//27	2	PAD	Simple	NP	NP
GMOY001312	507	1-21//22	2	PAD	Simple	NP	NP
GMOY001522	1030	NP	1	PAD	Simple	NP	NP
GMOY001662	345	NP	1	PAD	Simple	Multiple	NP
GMOY002084	187	NP	1	PAD	Simple	NP	NP
GMOY002086	1246	NP	1	PAD	Simple	NP	Multiple
GMOY002141	398	1-22//23	1	PAD	Simple	Multiple	NP
GMOY002142	1209	NP	1	PAD	Simple	Multiple	Multiple
GMOY002339	446	NP	1	PAD	Simple	Multiple	NP
GMOY002708	603	NP	3	PAD	Complex	Multiple	NP
GMOY002812	545	1-21//22	1	PAD	Simple	NP	NP
GMOY003840	469	1-21//22	6	PAD	Repetitive	NP	NP
GMOY004823	89	1-24//25	1	PAD	Simple	NP	NP
GMOY004893	295	1-18//19	3	PAD	Repetitive	NP	NP
GMOY005251	939	1-20//21	3	PAD	Complex	Multiple	Multiple
GMOY005235	659	NP	1	PAD	Simple	Multiple	NP
GMOY005278	2247	NP	13	PAD	Complex	Multiple	Multiple
GMOY006485	607	NP	3	PAD	Repetitive	NP	NP
GMOY006713	384	NP	1	PAD	Simple	NP	NP
GMOY007191	1040	1-19//20	2	PAD	Binary	Multiple	Multiple
GMOY007895	263	NP	1	PAD	Simple	NP	NP
GMOY008030	366	NP	1	PAD	Simple	NP	NP
GMOY008032	973	1-23//24	1	PAD	Simple	Multiple	NP
GMOY009587[2]	116	NP	1	Partial PCD	Simple	NP	NP
GMOY009788	173	1-19//20	2	PAD	Simple	NP	NP
GMOY009805	565	1-22//23	3	PAD	Complex	Multiple	NP
GMOY009806	282	NP	2	PAD	Simple	NP	NP
GMOY009807	216	1-17//18	2	PAD	Simple	NP	NP
GMOY009571	216	NP	3	PAD	Repetitive	NP	NP
GMOY009572	236	1-20//21	3	PAD	Repetitive	NP	NP
GMOY009826	236	NP	3	PAD	Repetitive	NP	NP
GMOY011054	112	1-27//28	1	PAD	Simple	NP	NP
GMOY011777	108	1-19//20	1	PAD	Simple	NP	NP
GMOY011809[1]	92	1-20//21	1	Partial PCD	Simple	NP	NP
GMOY011810	100	1-19//20	1	Partial PCD	Simple	NP	NP

Table 3.2. List of all *Glossina morsitans* peritrophin orthologues from haematophagous vectors of African trypanosomiasis, a non blood-feeding infectious disease vector and a non blood-feeding Dipteran. Those boxes coloured black with a white text indicate orthologues to entire protein sequences but lack any CDBs. Those boxes highlighted yellow indicate proteins that have been shown to be a structural constitute of the peritrophic membrane (GO:0016490) in *Drosophila*.

Morsitans	Austeni	Brevipalpis	Fuscipes	Pallidipes	Musca	Drosophila
GMOY000880	GAUT003677	GBRI031179	GFUI027404	GPAI017590	MDOA007958	Fbgn0052036
GMOY000968	GAUT010978		GFUI013195	GPAI009989	MDOA006197 MDOA012735	Fbgn0052024
GMOY001312	GAUT020830	GBRI007737	GFUI048226	GPAI009178	MDOA05592	Fbgn0030999
GMOY001522	GAUT026255	GBRI023956	GFUI039257	GPAI027229	MDOA007383	Fbgn0035430
GMOY001662		GBRI000180	GFUI007094		MDOA009492	
GMOY002084	GAUT027066	GBRI004278	GFUI036017	GPAI012018	MDOA000150 MDOA013821	Fbgn0038632
GMOY002086	GAUT027081	GBRI004277	GFUI036027	GPAI011997	MDOA011153	Fbgn0038629
GMOY002141	GAUT044450	GBRI023660	GFUI027472 GFUI048634	GPAI005401	MDOA002563	Fbgn0037488
GMOY002142					MDOA006218	Fbgn0037487
GMOY002339	GAUT014249	GBRI000180	GFUI016406	GPAI001697 GPAI021942	MDOA009492	
GMOY002708	GAUT001531 GAUT001532	GBRI011740 GBRI011741	GFUI006263	GPAI021775 GPAI021777	MDOA000494 MDOA000046 MDOA000998 MDOA013786	Fbgn0263748
GMOY002812	GAUT028770	GBRI019411	GFUI040609	GPAI030232	MDOA001441	Fbgn0260386
GMOY003533	GAUT001431	GBRI034366	GFUI034672 GFUI034673		MDOA011477	Fbgn0052656

GMOY003840	GAUT040598	GBRI039371	GFUI049678	GPAI033642	MDOA004664 MDOA009969 MDOA010543 MDOA000318 MDOA006800	Fbgn0031737
GMOY004823	GAUT015903 GAUT029663	GBRI027291	GFUI012030 GFUI017945	GPAI009564 GPAI016225	MDOA002969	Fbgn0260653
GMOY004893	GAUT011300	GBRI022913	GFUI015986	GPAI048004	MDOA009132	Fbgn0027600
GMOY005251		GBRI026889	GFUI045474 GFUI050311	GPAI039542	MDOA004810	Fbgn0025390
GMOY005235	GAUT010029	GBRI008201	GFUI023663	GPAI033432	MDOA000619	Fbgn0038422
GMOY005278	GAUT029524 GAUT029525	GBRI026281	GFUI015251	GPAI036992	MDOA003858	Fbgn0038492
GMOY006485	GAUT032284	GBRI024198	GFUI022163	GPAI033201	MDOA008538	
GMOY006713	GAUT023971		GFUI042924	GPAI039717	MDOA001947	Fbgn0034030
GMOY007191	GAUT013406	GBRI016187	GFUI029181	GPAI035996	MDOA002379 MDOA002027	Fbgn0036203
GMOY007476	GAUT039321	GBRI029781	GFUI006871	GPAI013818	MDOA009131	Fbgn0040601
GMOY007895	GAUT042124	GBRI006105		GPAI032440	MDOA001232	Fbgn0039172
GMOY008030	GAUT042200	GBRI039692	GFUI011847	GPAI019290	MDOA002964	Fbgn0036845
GMOY008032	GAUT042214	GBRI039674	GFUI011826	GPAI019295	MDOA011524	Fbgn0035844
GMOY009587	GAUT024681 GAUT008830	GBRI017781	GFUI019092	GPAI002755	MDOA015471	Fbgn0040959
GMOY009788	GAUT012373		GFUI027349	GPAI002186	MDOA004255	Fbgn0036230 Fbgn0262986
GMOY009805	GAUT012335		GFUI026256	GPAI002122	MDOA004255	Fbgn0036230 Fbgn0262986
GMOY009806	GAUT012312	GBRI029501	GFUI026249	GPAI002108	MDOA010787 MDOA006949	Fbgn0036226

GMOY009807	GAUT012316	GBRI029501	GFUI026252 GFUI026249	GPAI002112 GPAI002186 GPAI002122 GPAI002124	MDOA010787 MDOA006949	Fbgn0036226
GMOY009571	GAUT032039		GFUI025256	GPAI004770	MDOA001685	Fbgn0022770
GMOY009572	GAUT032040	GBRI003555	GFUI025254	GPAI004766	MDOA008182	Fbgn0031097
GMOY009826	GAUT041130	GBRI031719	GFUI037199 GFUI049166	GPAI031701	MDOA007176	Fbgn0026077
GMOY011054	GAUT026340	GBRI013108	GFUI045125	GPAI027171		
GMOY011777	GAUT036980 GAUT039327	GBRI013425 GBRI034385	GFUI006839 GFUI006847	GPAI013834 GPAI013840	MDOA003890 MDOA004772 MDOA012281 MDOA008142	Fbgn0039452
GMOY011809		GBRI007305			MDOA003564 MDOA003371	Fbgn0038643
GMOY011810	GAUT038910	GBRI007300 GBRI007301	GFUI000259		MDOA013567 MDOA008499	Fbgn0085311
	GAUT012314	GBRI029506	GFUI026251	GPAI002117		
	GAUT044451	GBRI023665	GFUI048633	GPAI005399		
	GAUT013406	GBRI016187	GFUI029181			Fbgn0036203
		GBRI011741	GFUI006267			
			GFUI037693			
			GFUI022165			
	GAUT036465	GBRI019611		GPAI033924	MDOA011173	
	GAUT051817	GBRI017597	GFUI036185	GPAI002755	MDOA002764 MDOA003298 MDOA005779 MDOA006933 MDOA007638 MDOA009096	FBgn0040958

					MDOA009405 MDOA011156 MDOA013209 MDOA014615 MDOA015066	
		GBRI023004	GFUI007992	GPAI006518	MDOA011511	
					MDOA006539	FBgn0035607
					MDOA007551	FBgn0053983
					MDOA008488	FBgn0053265
					MDOA006939	FBgn0035427
					MDOA009303 MDOA008963	FBgn0051439
					MDOA000944 MDOA015194	FBgn0053985
					MDOA001072 MDOA002800 MDOA004458 MDOA008137	FBgn0036950
					MDOA002131	FBgn0034662
					MDOA002136	FBgn0036361 FBgn0036362 FBgn0036363
					MDOA002322	FBgn0053263
					MDOA002412	FBgn0053986
					MDOA004034	FBgn0036940
					MDOA004171 MDOA008346	FBgn0038646
					MDOA005227 MDOA009442	FBgn0053263
					MDOA005641	FBgn0039042
					MDOA006479	

					MDOA006915	
					MDOA007356	FBgn0052302
					MDOA007741	FBgn0085456
					MDOA008065	FBgn0264488
					MDOA010052	FBgn0039453
					MDOA010250	FBgn0035325
					MDOA002314	FBgn0036949
					MDOA003800	
					MDOA004573	FBgn0013000
					MDOA005090	
					MDOA005986	FBgn0085455
					MDOA011640	
					MDOA006702	FBgn0036953
					MDOA007686	FBgn0036951
					MDOA009891	FBgn0036952 FBgn0036220
					MDOA012188	FBgn0036225
					MDOA002178	FBgn0036948
					MDOA004365	
						FBgn0040609
						FBgn0034301
						FBgn0036229
						FBgn0036563
						FBgn0035932
						FBgn0035933
						FBgn00.35931
						FBgn0035845
						FBgn0036232

						FBgn0040607
						FBgn0039453
						FBgn0037488
						FBgn0040687
						FBgn0035412
						FBgn0260393
						FBgn0051077
						FBgn0052284
						FBgn0053258
						FBgn0085249
						FBgn0085353
						FBgn0259192
						FBgn0259748
						FBgn0260026
						FBgn0260430
						FBgn0261681
						FBgn0261682
						FBgn0262854
						FBgn0040950
						FBgn0036947
						FBgn0036228
						FBgn0052304

3.3.2 Characterisation of putative peritrophins containing CBDs in other major *Glossina* species and comparative non blood-feeders, *Musca domestica and Drosophila melanogaster*

3.3.2.1 Peritrophins of *Glossina austeni*

All orthologues to *Glossina morsitans morsitans* peritrophins can be seen in table 3.2 and the structural organisation of the CBDs can be seen in Fig. 3.1b. *Glossina austeni* has orthologues to all *G. m. morsitans* peritrophins except for 3; GMOY001662, GMOY002142 and GMOY011809 (Pro1). Five of the orthologues in *G. austeni* have protein conservation other than the CBD, which suggests the acquisition of the CBD in *G. morsitans* or the loss of the CBD in *G. austeni*. However, *G. austeni* has additional paralogues to 3 of the orthologues, one being an extra orthologue to GMOY009587 (Pro2), and an additional 6 peritrophins that have no orthologues in *G. morsitans*. Moreover, although consisting of a similar number of peritrophins (40), the structural organisations of the CBDs differ from those in *G. morsitans*. The peritrophins of *G. austeni* consist of 29 simple, 3 binary, 2 complex and 6 repetitive peritrophins. There also more of the 3rd type of CBD. Five simple peritrophins contain either a complete or partial PCD, and one of these (GAUT010978) comprises of one PAD and one PCD flanking a small mucin-like domain. Thirteen of the peritrophins contained mucin domains and only 2 contained more than 3 CBDs; GAUT040598 with 6 CBDs and GAUT029525 with 11 CBDs.

3.3.2.2 Peritrophins of *Glossina brevipalpis*

All orthologues to *Glossina morsitans morsitans* peritrophins can be seen in table 3.2 and the structural organisation of the CBDs can be seen in Fig. 3.1c. *Glossina brevipalpis* has one fewer peritrophins than *G. morsitans* (35), with 7 missing paralogues. Instead this species has 3 additional paralogues including an additional orthologue to GMOY011810 (Per12) and another additional orthologue to GMOY002708 (Per66). *G. brevipalpis* also has a further 4 additional peritrophins than *G. morsitans*. Again, there are structural differences between the peritrophins of the two species. The peritrophins of G. brevipalpis comprise of 25 simple, 3 binary, 2 complex and 6 repetitive peritrophins. Moreover there are 8 simple peritrophins that contain a complete or partial PCDs and one that contains a complete PBD. Nine of the total number of peritrophins contains a mucin domain and only 2 peritrophins contained more than 3 CBDs; GBRI026281 and GBRI039371 which are both comprised of 6 CBDs. In addition, only 3 orthologues are missing a CBD with the rest of the protein being conserved.

3.3.2.3 Peritrophins of *Glossina fuscipes*

All orthologues to *Glossina morsitans morsitans* peritrophins can be seen in table 3.2 and the structural organisation of the CBDs can be seen in Fig. 3.1d. *Glossina fuscipes* has the most peritrophins of all *Glossina* species analysed with 44. Four of the orthologues have protein conservation other than the CBD, which they lack. This species has orthologues to all *G. m. morsitans* peritrophins except for 3; GMOY007895, GMOY002142 and GMOY011809 (Pro1), and an additional 9 peritrophins with no orthologues in *G. morsitans*. *G. fuscipes* has additional paralogues to 8 orthologues of *G. morsitans*. The structural organisation of the peritrophins is different in this species compared to *G. morsitans*, consisting of 31 simple, 2 binary, 3 complex and 8 repetitive peritrophins. Of the simple peritrophins, 3 have full PCDs and out of all the peritrophins 12 have mucin domains. Most of the peritrophins of *G. fuscipes* contain 3 or less CBDs, the exceptions being GFUI015251 with 13 CBDs, GFUI026256 with 4 CBDs and GFUI049678 which contains 6 CBDs.

3.3.2.4 Peritrophins of *Glossina pallidipes*

All orthologues to *Glossina morsitans morsitans* peritrophins can be seen in table 3.2 and the structural organisation of the CBDs can be seen in Fig. 3.1e. *Glossina pallidipes* has the same number of peritrophins to *G. brevipalpis* and one less than *G. morsitans* with 35. However, eight of the orthologues do not contain a CBD; instead the rest of the protein is conserved. *G. pallidipes* has 6 missing orthologues to *G. morsitans* including GMOY011809 (Pro1) and GMOY011810 (Per12). *G. pallidipes* has additional paralogues to 6 orthologues of *G. morsitans* including 4 to GMOY009807, the highest number of orthologues to any *G. morsitans* peritrophins. There is also an additional 6 peritrophins with no orthologues to any *G. morsitans* proteins. The organisation of the CBDs places the peritrophins into 21 simple, 5 binary, 1 complex and 8 repetitive peritrophins. Only 1 simple peritrophin consists of a full PCD, the rest all have PADs and a total of 11 peritrophins contain a mucin domain. Most of the peritrophins have 3 or less CBDs with only 2 containing more; GPAI036992 with 13 CBDs interspersed with mucin domains and GPAI033642 which contains 6 CBDs.

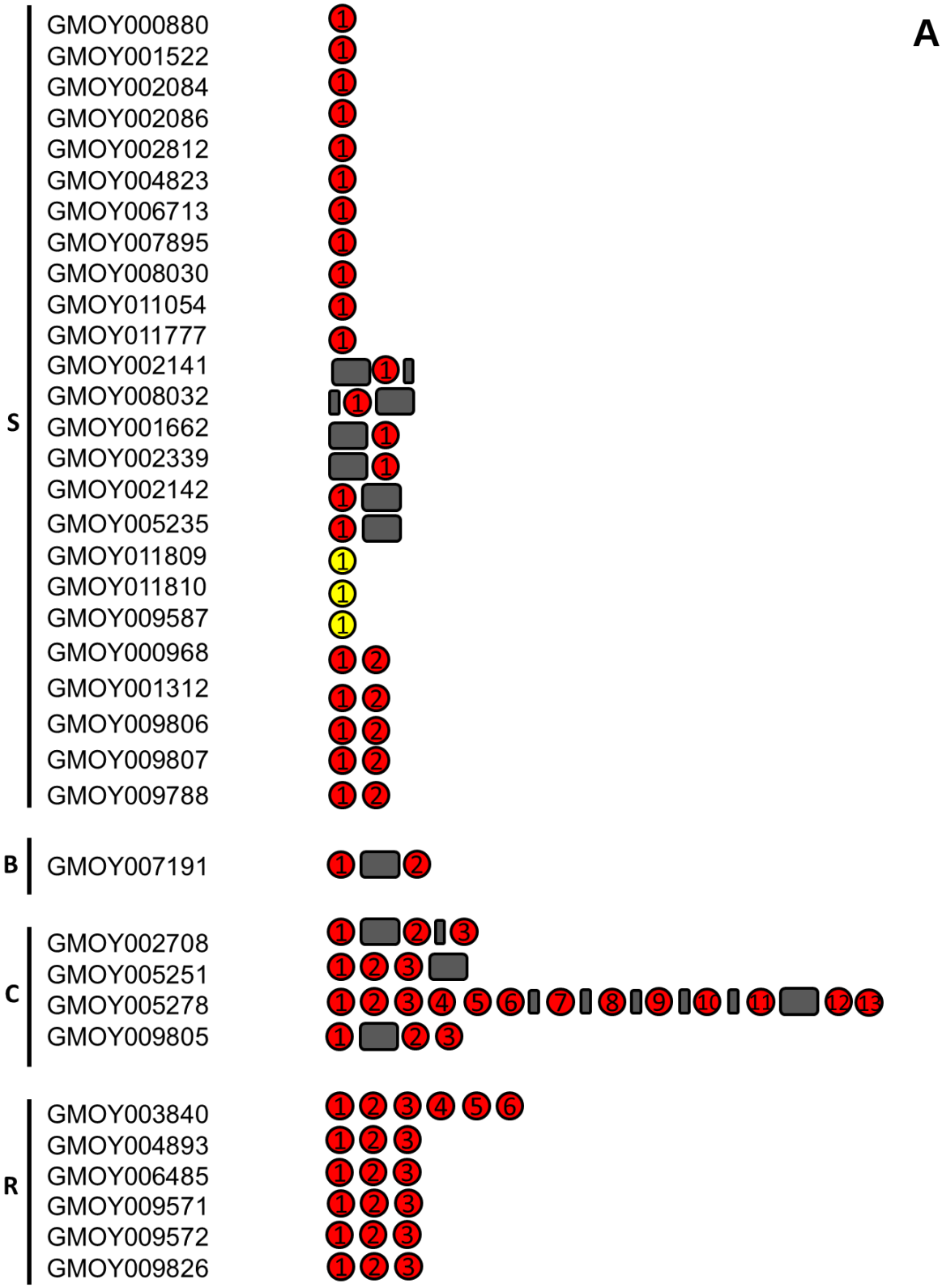
3.3.2.5 Peritrophins of *Musca domestica*

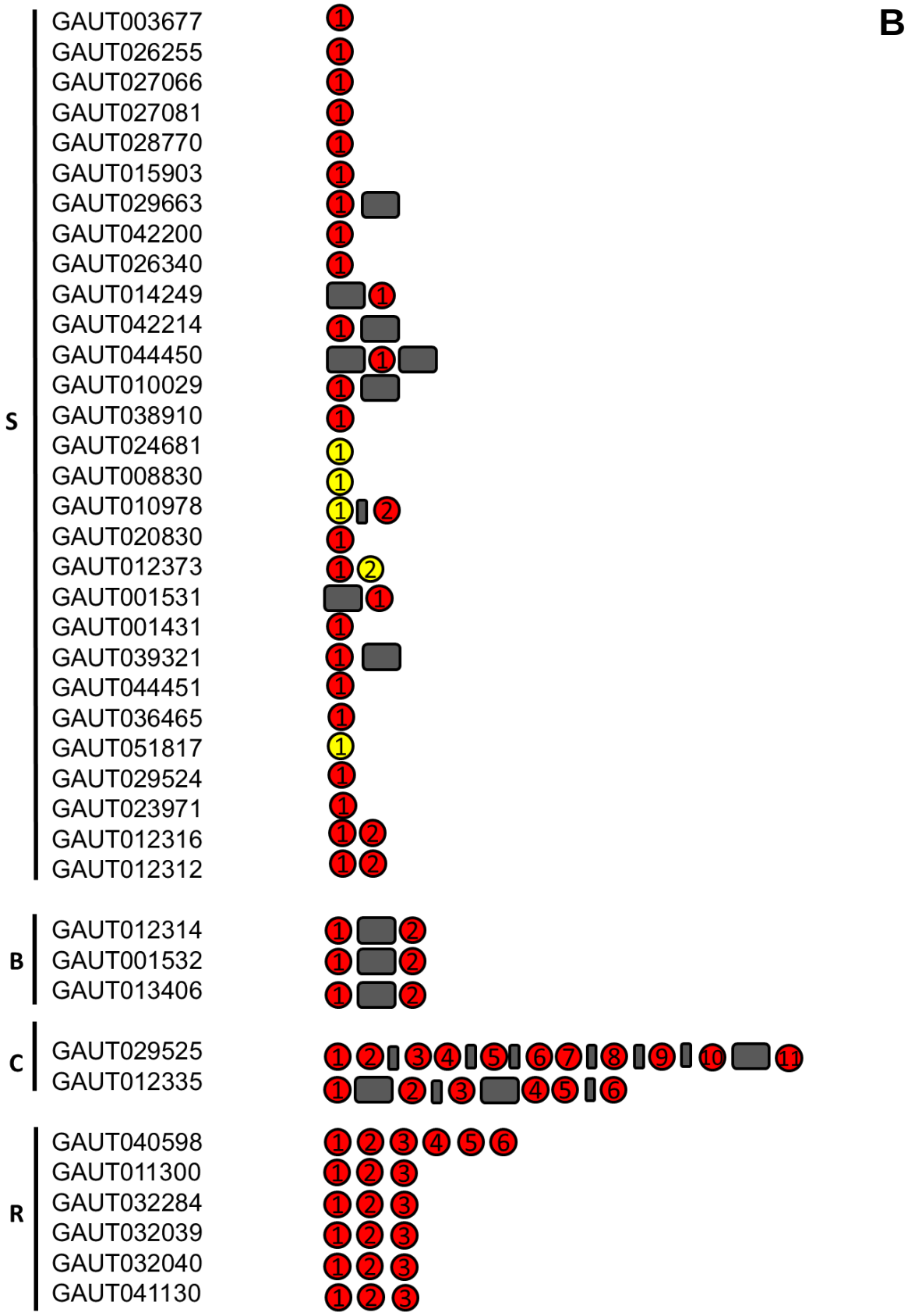
All orthologues to *Glossina morsitans morsitans* peritrophins can be seen in table 3.2 and the structural organisation of the CBDs can be seen in Fig. 3.1f. *Musca domestica* has more than double the number of peritrophins of all *Glossina* species with 102. There are 62 extra peritrophins that are not found in *Glossina morsitans,* but only 1 that *G. morsitans* has that *M.*

domestica does not; GMOY011054. Only 2 orthologues lack a CBD and have conservation in the rest of the protein. *M. domestica* has additional paralogues to 16 *G. morsitans* orthologues including 2 paralogues to GMOY011809 (Pro1), 2 paralogues to GMOY011810 (Per12) and 2 paralogues to GMOY007191 (Per108). All but 2 additional *M. domestica* peritrophins with no orthologues to *G. morsitans* had orthologues to *Drosophila melanogaster*. As expected there are major differences in the structural organisation of the peritrophins in *M. domestica* compared to *Glossina* species. The peritrophins comprise 57 simple, 3 binary, 17 complex and 25 repetitive. Out of all the peritrophins, 14 contain full or partial PCDs and 8 contain full or partial PBDs. All peritrophins containing a CBD other than a PAD are found within the simple peritrophins and a total of 32 peritrophins have a mucin domain. Unlike most peritrophins of *Glossina* species, many of the peritrophins in *Musca* are larger than 3, the largest one, MDOA008065, containing 39 PADs which is the highest number of CBDs on a single Dipteran peritrophin so far described.

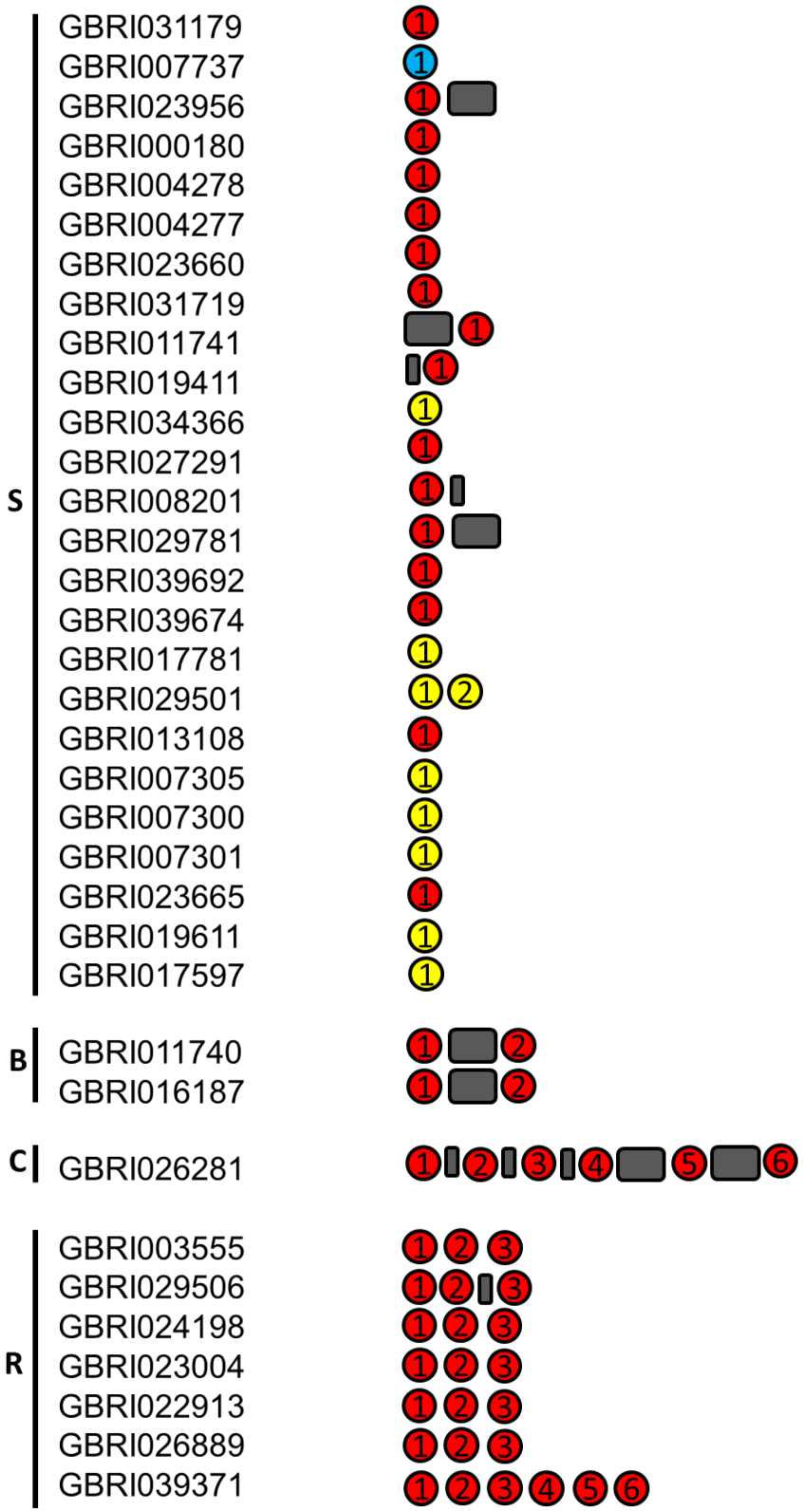
3.3.2.6 Peritrophins of *Drosophila melanogaster*

All orthologues to *Glossina morsitans morsitans* peritrophins can be seen in table 3.2 and the structural organisation of the CBDs can be seen in Fig. 3.1g. *Drosophila melanogaster* has a higher number of peritrophins than those in *Glossina* species, but fewer than those in *M. domestica*, with a total of 90. *D. melanogaster* has an additional 64 peritrophins than *G. morsitans* with 34 out of the 36 *G. morsitans* peritrophins orthologous. Like *M. domestica*, *Drosophila* lacks an orthologue to GMOY011054, suggesting that this peritrophin is unique to *Glossinidae,* however does have a single orthologue to each of the *G. morsitans* peritrophins found by mass spectrometry. Of the 24 *Drosophila* peritrophins determined to be a structural constitute of the peritrophic membrane (GO:0016490) as inferred electronically or experimentally, 8 had orthologues in *Glossina* including GMOY007191 which has previously been found as PM component (Chapter 2), an additional 12 had orthologues in *M. domestica* and 4 which have no orthologues in any *Glossina* or *Musca* species. The structural organisation of *Drosophila* peritrophins is more similar to *Musca* than *Glossina*, comprising of 49 simple, 4 binary, 9 complex and 28 repetitive peritrophins. However, unlike *Musca,* the peritrophins of *D. melanogaster* do not contain any PBDs; instead 10 consist of a full or partial PCD, 3 consist of a mixture of PADs and PCDs and the remainder (77) consist of PADs. *Drosophila* has the most repetitive peritrophins with 28 including one, FBgn0264488, which contains 31 PADs and interestingly is one of the peritrophins found to be a constitute of the peritrophic membrane.

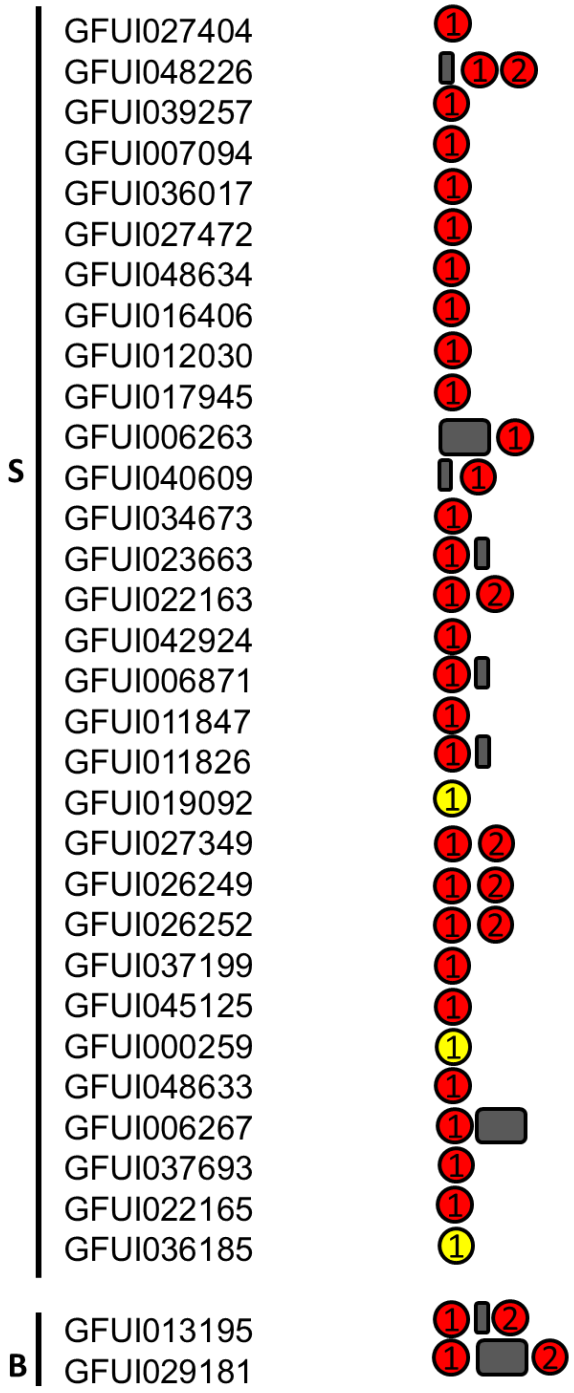

A
S
GMOY000880
GMOY001522
GMOY002084
GMOY002086
GMOY002812
GMOY004823
GMOY006713
GMOY007895
GMOY008030
GMOY011054
GMOY011777
GMOY002141
GMOY008032
GMOY001662
GMOY002339
GMOY002142
GMOY005235
GMOY011809
GMOY011810
GMOY009587
GMOY000968
GMOY001312
GMOY009806
GMOY009807
GMOY009788
B
GMOY007191
C
GMOY002708
GMOY005251
GMOY005278
GMOY009805
R
GMOY003840
GMOY004893
GMOY006485
GMOY009571
GMOY009572
GMOY009826

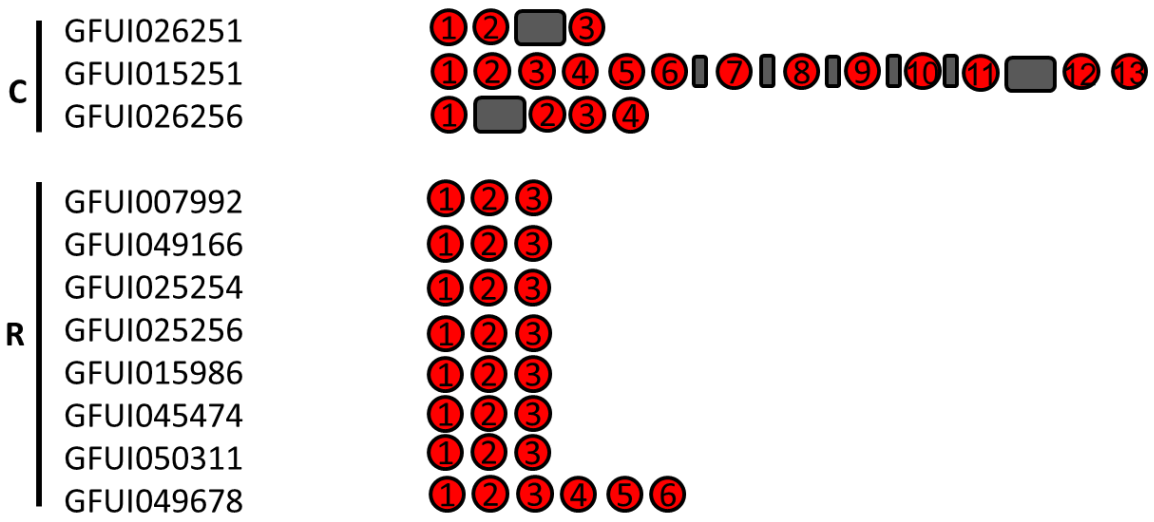

B
S
GAUT003677
GAUT026255
GAUT027066
GAUT027081
GAUT028770
GAUT015903
GAUT029663
GAUT042200
GAUT026340
GAUT014249
GAUT042214
GAUT044450
GAUT010029
GAUT038910
GAUT024681
GAUT008830
GAUT010978
GAUT020830
GAUT012373
GAUT001531
GAUT001431
GAUT039321
GAUT044451
GAUT036465
GAUT051817
GAUT029524
GAUT023971
GAUT012316
GAUT012312
B
GAUT012314
GAUT001532
GAUT013406
C
GAUT029525
GAUT012335
R
GAUT040598
GAUT011300
GAUT032284
GAUT032039
GAUT032040
GAUT041130

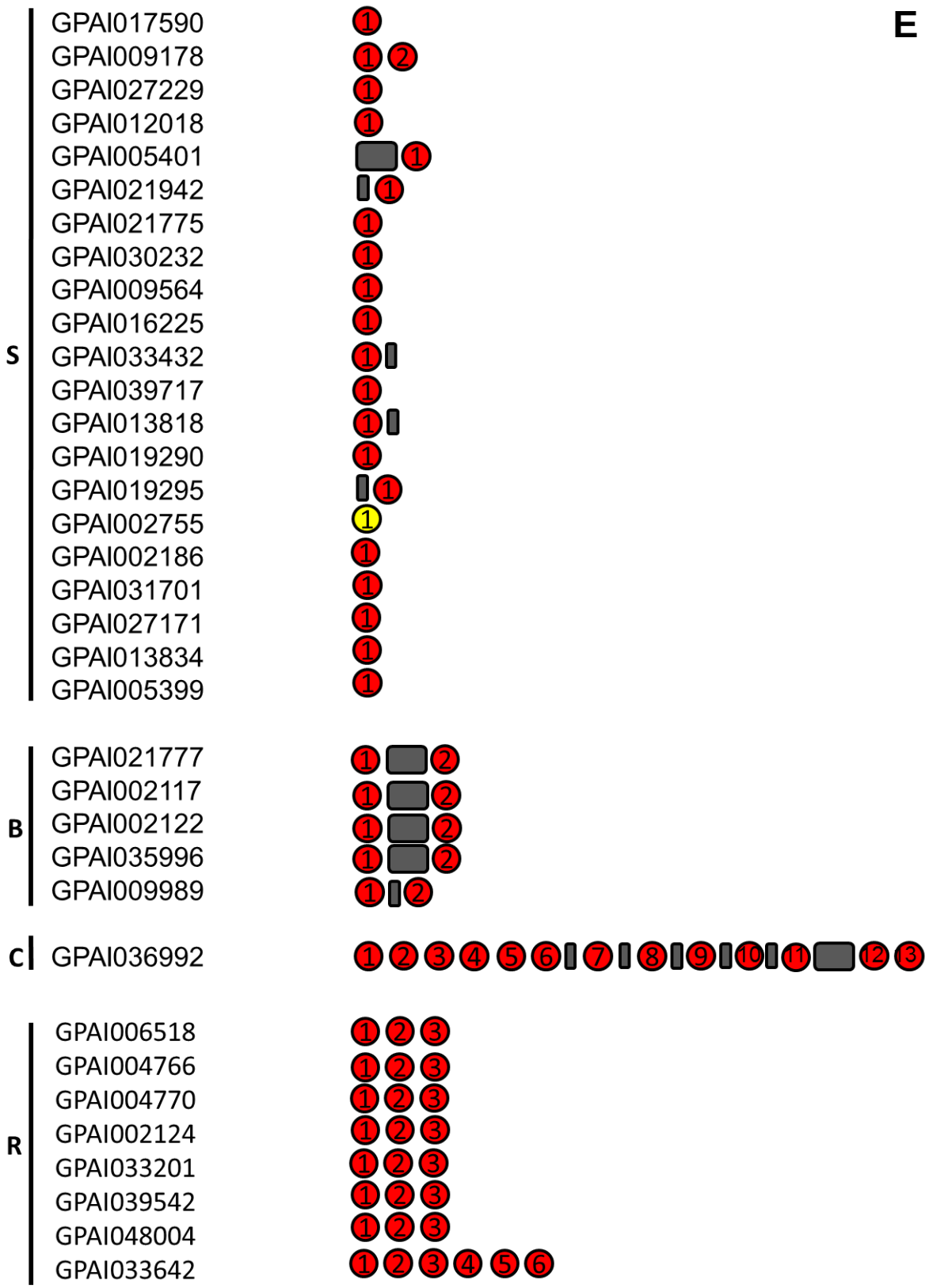
C


S
GBRI031179
GBRI007737
GBRI023956
GBRI000180
GBRI004278
GBRI004277
GBRI023660
GBRI031719
GBRI011741
GBRI019411
GBRI034366
GBRI027291
GBRI008201
GBRI029781
GBRI039692
GBRI039674
GBRI017781
GBRI029501
GBRI013108
GBRI007305
GBRI007300
GBRI007301
GBRI023665
GBRI019611
GBRI017597
B
GBRI011740
GBRI016187
C
GBRI026281
R
GBRI003555
GBRI029506
GBRI024198
GBRI023004
GBRI022913
GBRI026889
GBRI039371

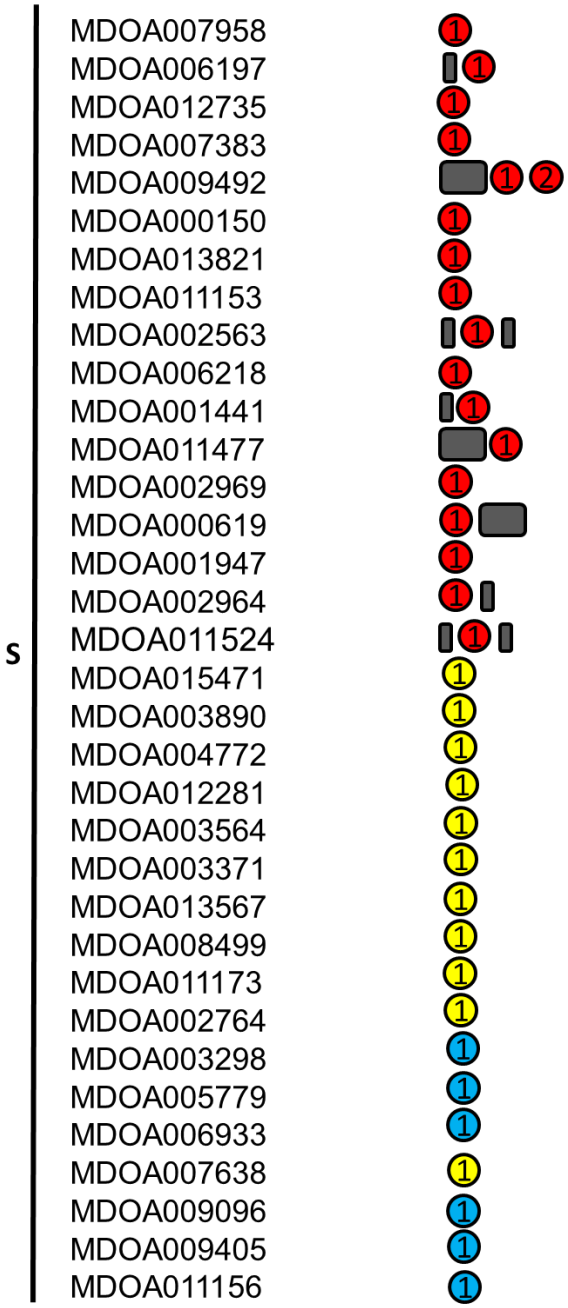
D


S
GFUI027404
GFUI048226
GFUI039257
GFUI007094
GFUI036017
GFUI027472
GFUI048634
GFUI016406
GFUI012030
GFUI017945
GFUI006263
GFUI040609
GFUI034673
GFUI023663
GFUI022163
GFUI042924
GFUI006871
GFUI011847
GFUI011826
GFUI019092
GFUI027349
GFUI026249
GFUI026252
GFUI037199
GFUI045125
GFUI000259
GFUI048633
GFUI006267
GFUI037693
GFUI022165
GFUI036185
B
GFUI013195
GFUI029181


C
GFUI026251
GFUI015251
GFUI026256
R
GFUI007992
GFUI049166
GFUI025254
GFUI025256
GFUI015986
GFUI045474
GFUI050311
GFUI049678


E
S
GPAI017590
GPAI009178
GPAI027229
GPAI012018
GPAI005401
GPAI021942
GPAI021775
GPAI030232
GPAI009564
GPAI016225
GPAI033432
GPAI039717
GPAI013818
GPAI019290
GPAI019295
GPAI002755
GPAI002186
GPAI031701
GPAI027171
GPAI013834
GPAI005399
B
GPAI021777
GPAI002117
GPAI002122
GPAI035996
GPAI009989
C
GPAI036992
R
GPAI006518
GPAI004766
GPAI004770
GPAI002124
GPAI033201
GPAI039542
GPAI048004
GPAI033642

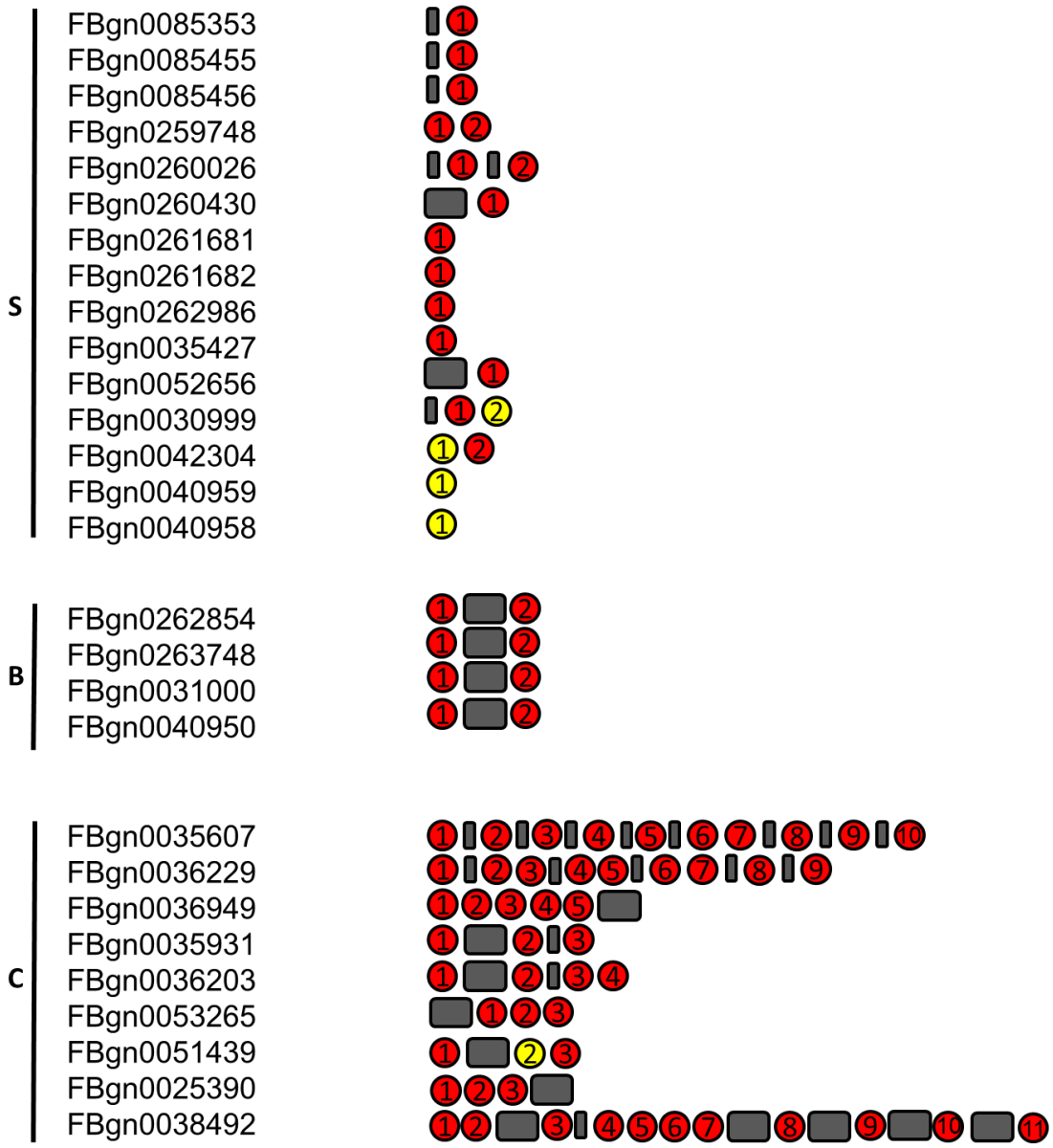
F


S
MDOA007958
MDOA006197
MDOA012735
MDOA007383
MDOA009492
MDOA000150
MDOA013821
MDOA011153
MDOA002563
MDOA006218
MDOA001441
MDOA011477
MDOA002969
MDOA000619
MDOA001947
MDOA002964
MDOA011524
MDOA015471
MDOA003890
MDOA004772
MDOA012281
MDOA003564
MDOA003371
MDOA013567
MDOA008499
MDOA011173
MDOA002764
MDOA003298
MDOA005779
MDOA006933
MDOA007638
MDOA009096
MDOA009405
MDOA011156


S
MDOA013209
MDOA014615
MDOA015066
MDOA000318
MDOA006939
MDOA009303
MDOA000046
MDOA000944
MDOA000998
MDOA002322
MDOA004171
MDOA006915
MDOA007356
MDOA007741
MDOA008142
MDOA009442
MDOA010052
MDOA010250
MDOA005986
MDOA013786
MDOA006800
MDOA008137
MDOA008346
B
MDOA005227
MDOA002379
MDOA000494
C
MDOA002800
MDOA011640
MDOA006539
MDOA003858
MDOA002027
MDOA010787
MDOA007551
MDOA008488
MDOA006949
MDOA003800
MDOA004573
MDOA005090
MDOA005592
MDOA004810
MDOA002314
MDOA004458
MDOA002178


R
MDOA015194
MDOA005641
MDOA002412
MDOA002136
MDOA002131
MDOA011511
MDOA007176
MDOA008182
MDOA001685
MDOA004255
MDOA008538
MDOA009132
MDOA010543
MDOA009969
MDOA004664
MDOA001072
MDOA004034
MDOA006479
MDOA008065
MDOA008963
MDOA006702
MDOA007686
MDOA009891
MDOA012188
MDOA004365

G


S
FBgn0085353
FBgn0085455
FBgn0085456
FBgn0259748
FBgn0260026
FBgn0260430
FBgn0261681
FBgn0261682
FBgn0262986
FBgn0035427
FBgn0052656
FBgn0030999
FBgn0042304
FBgn0040959
FBgn0040958
B
FBgn0262854
FBgn0263748
FBgn0031000
FBgn0040950
C
FBgn0035607
FBgn0036229
FBgn0036949
FBgn0035931
FBgn0036203
FBgn0053265
FBgn0051439
FBgn0025390
FBgn0038492

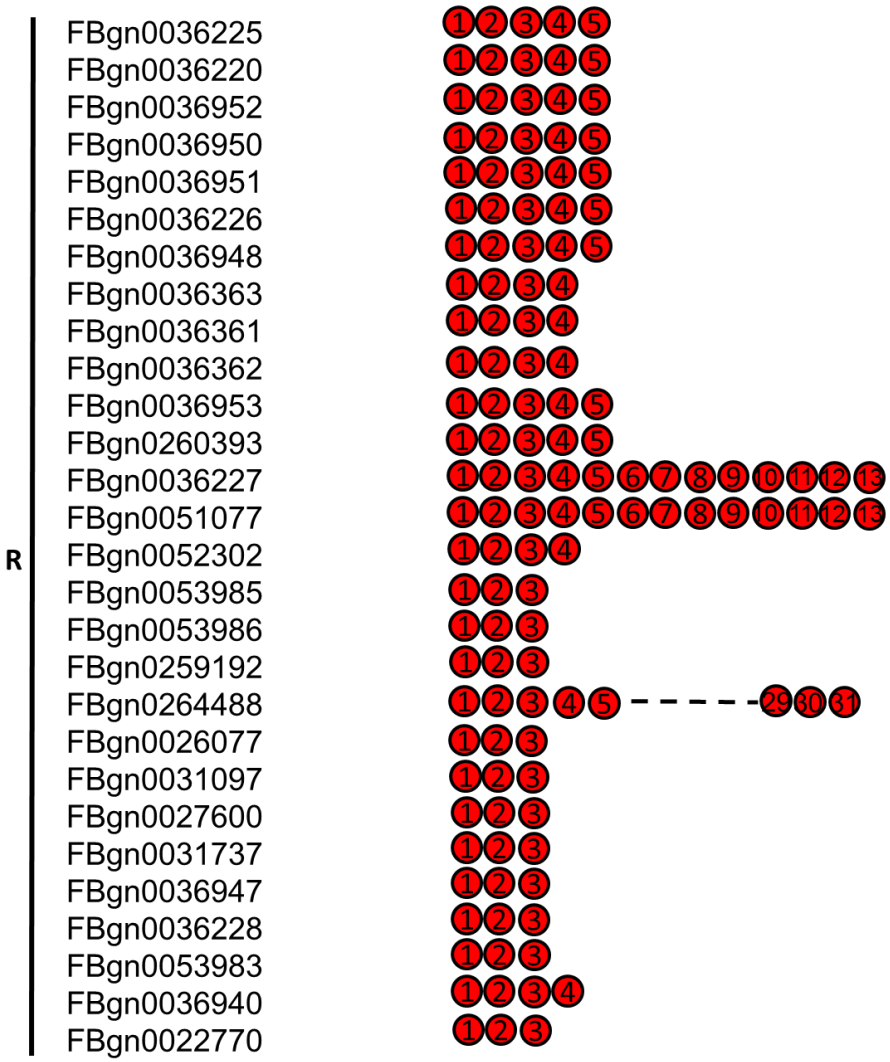
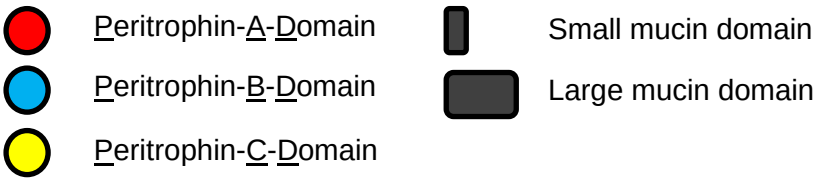


Figure 3.1. The structural organisation of putative peritrophins from several Dipteran species. *A-G; G. morsitans, G. austeni, G. brevipalpis, G. fuscipes, G. pallidipes, M. domestica and D. melanogaster respectively. S; simple peritrophins, B; binary peritrophins, C; complex peritrophins and R; repetitive peritrophins. CBDs are numbered sequentially from N to C terminus.*

Key;



3.3.3 Phylogenetic analysis of known *G. morsitans* PM proteins and their orthologues

3.3.3.1 GMOY002708 (GmmPer66)

All species of Glossina so far annotated possess an orthologue to GmmPer66, the major PM protein of *G. m. morsitans*, with *G. Brevipalpis* possessing two. There are also orthologues in *D. melanogaster* and *M. domestica* possess 3 but none were identified in *Stomoxys calcitrans*, a comparative bloodfeeder that can mechanically transmit many diseases to both humans and animals but is not a biological vector [160]. The phylogenetic branching pattern of *Musca* and *Drosophila* GmmPer66 genes (Fig. 3.2) suggests there could be an ancestral as well as a modern origin of this peritrophin in *Musca* in relation to *Drosophila*. In addition, GmmPer66 in all *Glossina* species diverged later in evolution; with the gene in *G. brevipalpis* diverging most recently which is surprising considering this species is thought to be the most ancient species of tsetse.

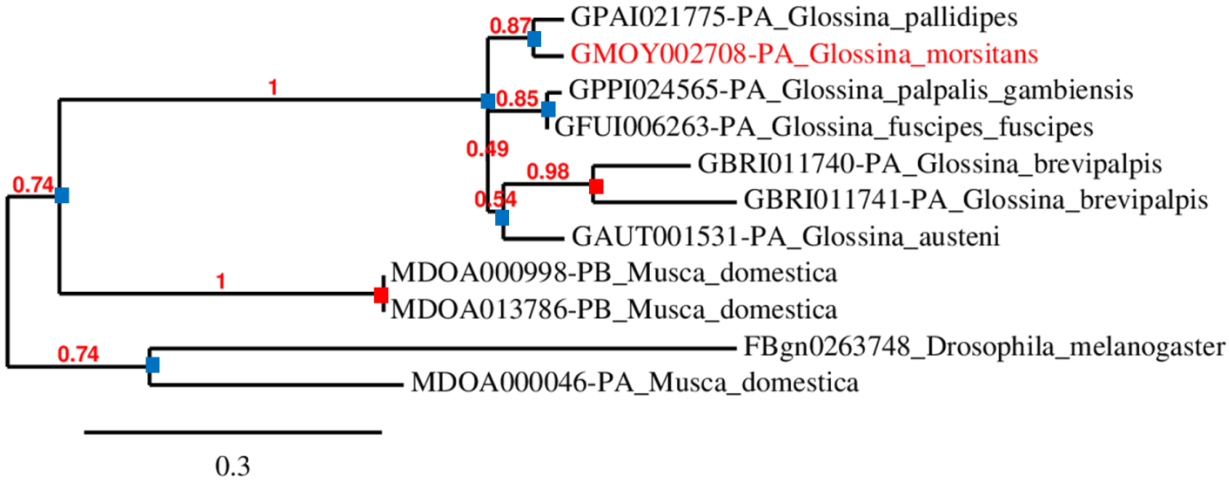


Figure 3.2 Phylogenetic analyses of GmmPer66 from several Dipteran species. *Phylogenetic tree was constructed by maximum likelihood criteria with bootstrap replicates set at 500. Taxa complete names and the accession numbers of the sequences used in the analysis were obtained from VectorBase and FlyBase. Branch values indicate bootstrap support. Red boxes at nodes depict duplication events. Blue boxes at nodes depict speciation events*

3.3.3.2 GMOY007191 (GmmPer108)

All *Glossina* species have an orthologue to GMOY007191, *Musca* has two and *Drosophila* and *Stomoxys* possess one each. The dendogram shown in Fig. 3.3 suggests the genes in *Stomoxys* and *Musca* are closely related and both have a common origin to the peritrophin in *Drosophila.* A second orthologue in Musca appeared after a duplication event before speciation events gave rise to the gene that can be found in all *Glossina* species.

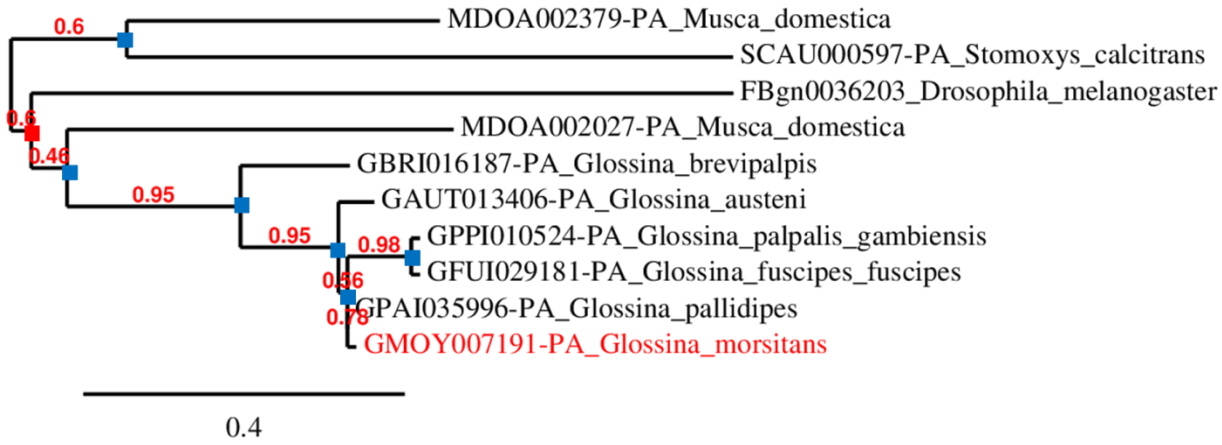


Figure 3.3 Phylogenetic analyses of GmmPer108 from several Dipteran species. *Phylogenetic tree was constructed by maximum likelihood criteria with bootstrap replicates set at 500. Taxa complete names and the accession numbers of the sequences used in the analysis were obtained from VectorBase and FlyBase. Branch values indicate bootstrap support. Red boxes at nodes depict duplication events. Blue boxes at nodes depict speciation events*

3.3.3.3 GMOY009587 (Pro2)

Again, all species of *Glossina* have orthologues to Pro2. Whilst *Drosophila* has one, *Musca* and *Stomoxys* have multiple orthologues. The branching pattern of the dendogram (Fig. 3.4) shows two distinct clusters of the gene in *Musca* which suggests two seperate expansions of Pro2 in this species which could indicate the importance of this gene in *Musca*. Two further genes in *Musca* appear to have diverged separately from these distinct groups; one closely related to a *Stomoxys* gene and the other closely related to the only *Drosophila* orthologue. Interestingly, Pro2 of *G. brevipalipis* appears to be more closely related to the genes in one of the clusters of *Musca* rather than the Pro2 found in all other *Glossina* species. There are also several orthologues in *Stomoxys* which appear to have evolved independently following speciation events. It remains to be seen if these are functionally homologous.

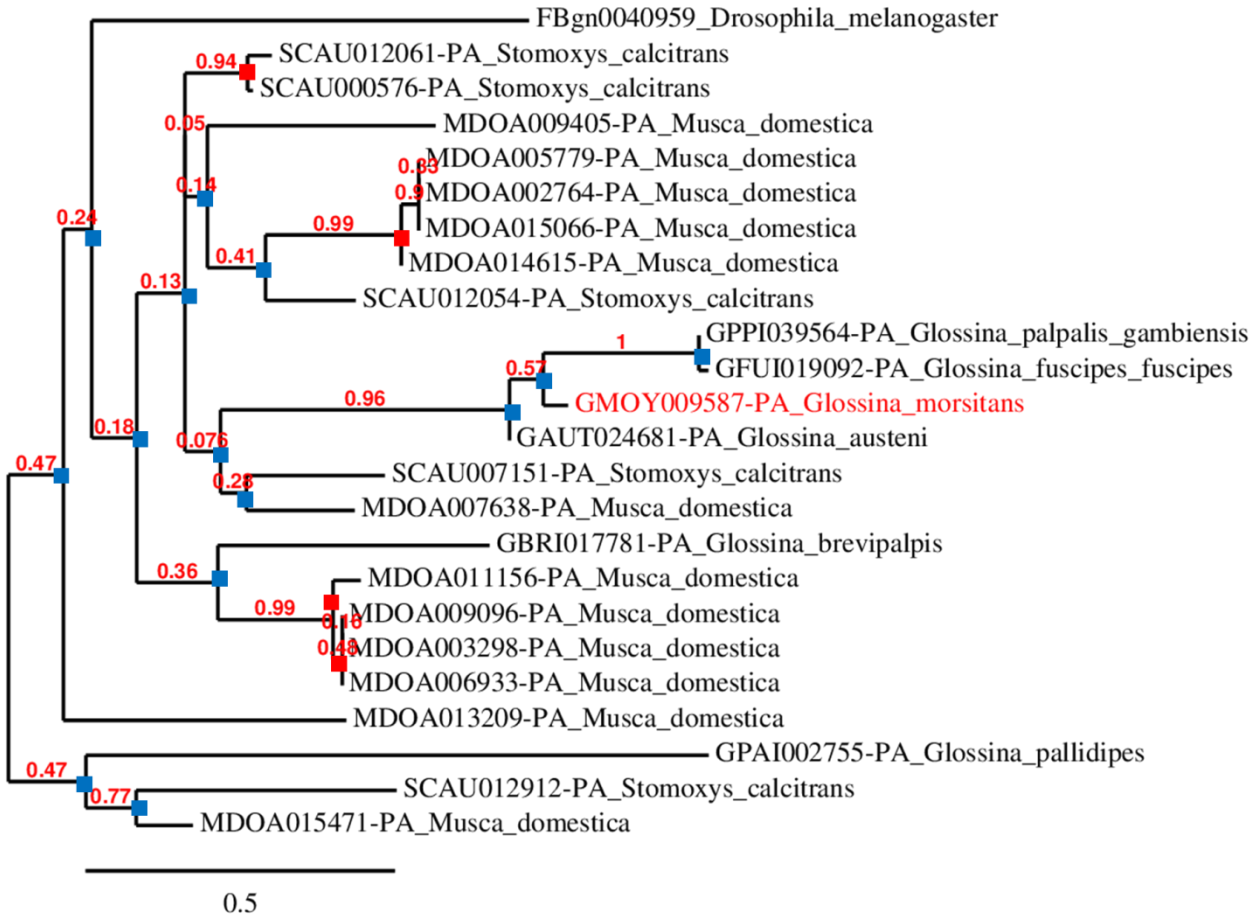


Figure 3.4 Phylogenetic analyses of Pro2 from several Dipteran species. *Phylogenetic tree was constructed by maximum likelihood criteria with bootstrap replicates set at 500. Taxa complete names and the accession numbers of the sequences used in the analysis were obtained from VectorBase and FlyBase. Branch values indicate bootstrap support. Red boxes at nodes depict duplication events. Blue boxes at nodes depict speciation events*

3.3.3.4 GMOY011809 (Pro1)

Unlike the three previously described peritrophins, Pro1 is only found in 3 species of *Glossina*; *G. morsitans*, *G. brevipalpis* and *G. palpalis* suggesting that this peritrophin may not be criitcal in all *Glossina* but may have an important structural or functional role in those tsetse that do have the gene. The dendogram in Fig. 3.5 shows an expansion of Pro1 in Stomoxys with 4 ortholgues as well as an expansion in Musca with two orthologues. The close phylogenetic relationship suggests Pro1 is important in these two species.

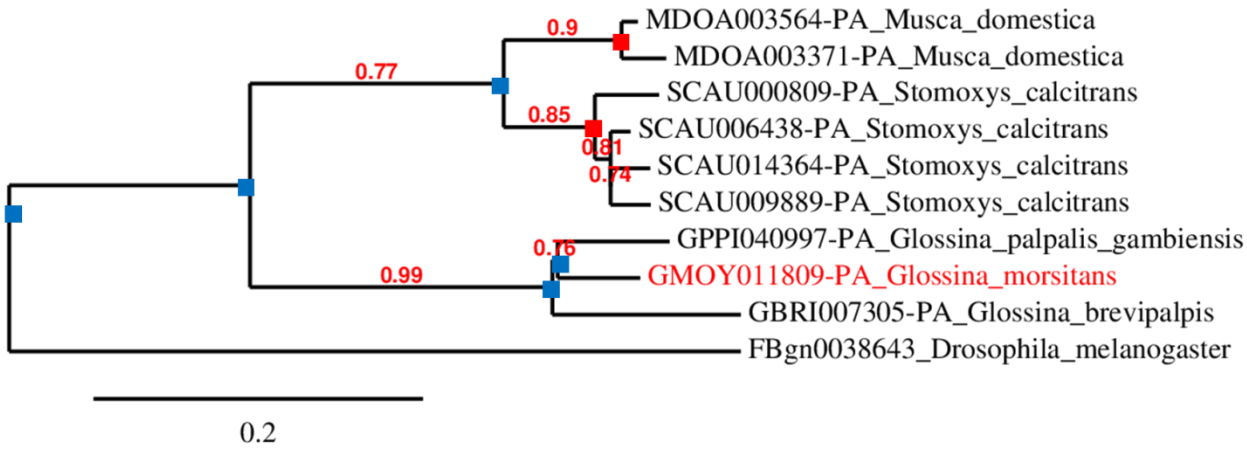


Figure 3.5 Phylogenetic analyses of Pro1 from several Dipteran species. *Phylogenetic tree was constructed by maximum likelihood criteria with bootstrap replicates set at 500. Taxa complete names and the accession numbers of the sequences used in the analysis were obtained from VectorBase and FlyBase. Branch values indicate bootstrap support. Red boxes at nodes depict duplication events. Blue boxes at nodes depict speciation events*

3.3.3.5 GMOY011810 (GmmPer12)

All species of *Glossina* with the exception of *G. pallidipes* posess an othologue of GmmPer12. The phylogenetic branching pattern of the Pro1 genes (Fig. 3.6) suggests they appeared first in *Glossina* and later evolved in *Drosophila* possibly from recombination of other, similar, peritrophins of the same nature. Similarly to Pro1, there is an expansion of GmmPer12 in *Stomoxys* with 3 genes and in *Musca* with 2 genes and all are closely related.

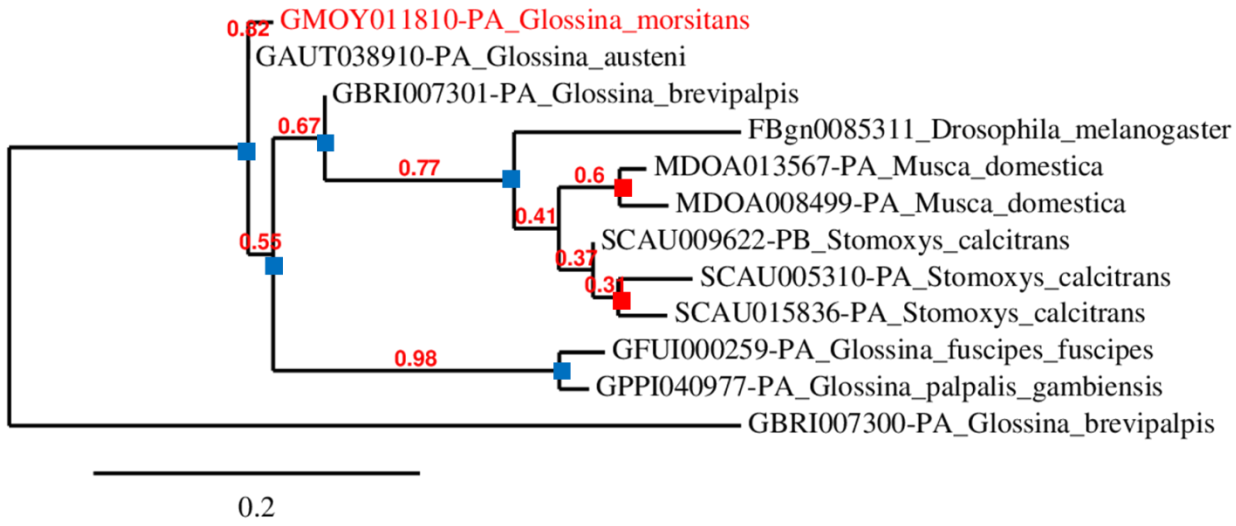


Figure 3.6 Phylogenetic analyses of GmmPer12 from several Dipteran species. *Phylogenetic tree was constructed by maximum likelihood criteria with bootstrap replicates set at 500. Taxa complete names and the accession numbers of the sequences used in the analysis were obtained from VectorBase and FlyBase. Branch values indicate bootstrap support. Red boxes at nodes depict duplication events. Blue boxes at nodes depict speciation events*

3.3.4 Alignments of PADs from known *G. morsitans* PM peritrophins and their orthologues

Tachycitin, a small antimicrobial protein with chitin-binding activity from the horseshoe crab, provides an excellent guide structure to model PADs against. It contains 10 cysteine residues (Fig. 3.7a), of which, 6 residues can be perfectly aligned with the six cysteine residues of PADs. Cysteine numbers 1, 3, 5, 7, 8 and 9 from tachycitin correspond and align with cysteine numbers 1, 2, 3, 4, 5 and 6 respectively in the PADs of Dipteran peritrophins.

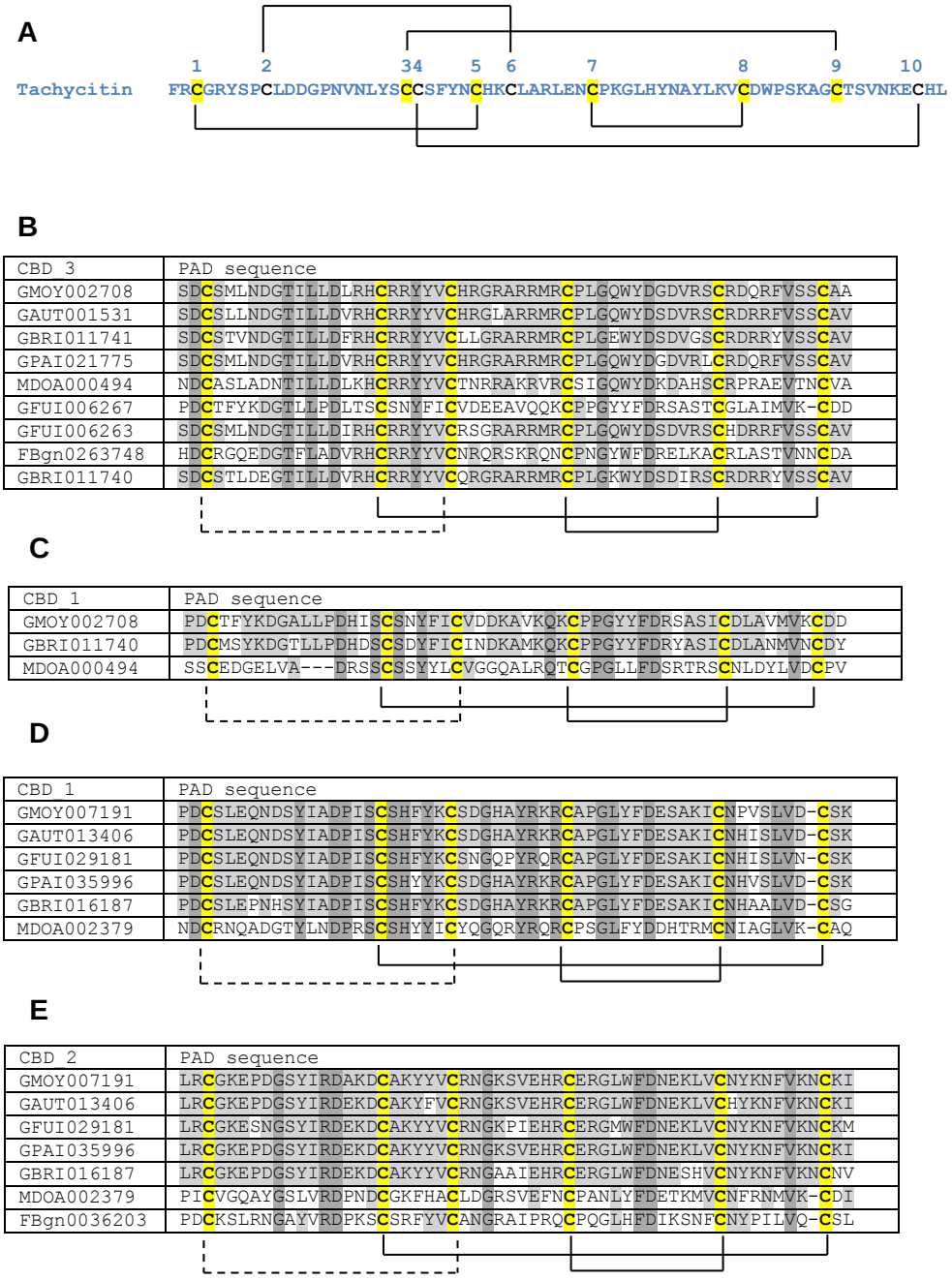


Figure 3.7. Alignments of PADs from several Dipteran species against their respective orthologues using tachycitin as a guide. *A; the 10 cysteine residues of the CBD in tachycitin, B-C; The alignments of the carboxy and amino terminal PADs from GMOY002708 and their respective orthologues. D-E; Alignments of the amino and carboxy terminal PADs from GMOY007191 against their respective orthologues. Conserved cysteine residues are in bold and highlighted yellow and the highlighted cysteines in the CBD of tachycitin correspond to those in the Dipteran PADs. One hundred percent conserved residues are highlighted in dark grey whilst ≥60% of conserved residues are highlighted in light grey. Putative S-S bonds as predicted by multiple software programmes in agreement are indicated by solid lines and putative S-S bonds predicted by only one software programme are indicated by dashed lines*

3.3.4.1 GMOY002708

GMOY002708 (GmmPer66 – Chapter 2), is a protein of 603 amino acids in length and has three PADs. The PAD at the N-terminus end is followed by a large mucin domain which is predicted to be heavily *0*-glycosylated, a second PAD followed by a smaller mucin domain and finally the third PAD at the C-terminus. Most of the orthologues to GMOY002708 from all species analysed were comprised of only the small mucin domain and the 3rd PAD at the C-terminus (Fig. 5b-c), none contained the 2nd PAD and only *G. brevipalpis* and *M. domestica* had the additional N-terminus PAD.

3.3.4.2 GMOY007191

GMOY007191 (GmmPer108 – Chapter 2), is a protein of 1040 amino acids in length and has two PADs. As previously stated, it is a binary peritrophin and has a huge mucin domain of 865 amino acids flanked by the two PADs. All *Glossina* species analysed contained both PADs as did *M. domestica* (Fig. 5d-e).

3.3.5 Alignments of PCDs from known *G. morsitans* PM peritrophins and their orthologues

3.3.5.1 GMOY011809

GMOY011809 (Pro1 – Chapter 2) is a small protein of 92 amino acids and is entirely comprised of a partial PCD. Four of its 5 cysteine residues (cysteine numbers 6-9) align with the cysteines at the carboxy terminus end of peritrophins with full PCDs which contain 10 cysteines. Of all the *Glossina* species analysed, only *G. brevipalpis* has an orthologue of Pro1, whereas there are 2 orthologues in *M. domestica.* Within the latter species, the protein and therefore the CBD is longer than in *Glossina* species. Although still only a partial PCD, this orthologue has an additional 2 cysteines towards the amino terminus which are both predicted to form disulphide bridges (Fig. 3.8a).

3.3.5.2 GMOY009578

GMOY009578 (Pro2 – Chapter 2) is a small protein of 112 amino acids and is also comprised of an entire partial PCD. It contains nine cysteine residues but only the last five (carboxy end) cysteine residues (cysteine numbers 6-10) align with full PCDs of peritrophins from other insects (Fig. 3.8b). All other *Glossina* species and *M. domestica* have Pro2 orthologues,

although *G. pallidipes* has only 6 cysteines and *G. brevipalpis* has 9 but in a slightly different consensus.

3.3.5.3 GMOY011810

GMOY011810 (GmmPer12 – Chapter 2) is a small protein of 100 amino acids with a predicted molecular mass of ~12 kDa and is also comprised of a partial PCD. The CBD consists of 5 cysteine residues but only the last 4 align with the carboxy terminus end (cysteine numbers 6-9) cysteines from peritrophins with full PCDs (Fig. 3.8c). All other *Glossina* species have an orthologue to GmmPer12 except for *G. pallidipes.* In addition, *M. domestica* has 2 orthologues as does *G. brevipalpis.* However, in the latter species one orthologue contains only the 4 cysteines which align to full PCDs.

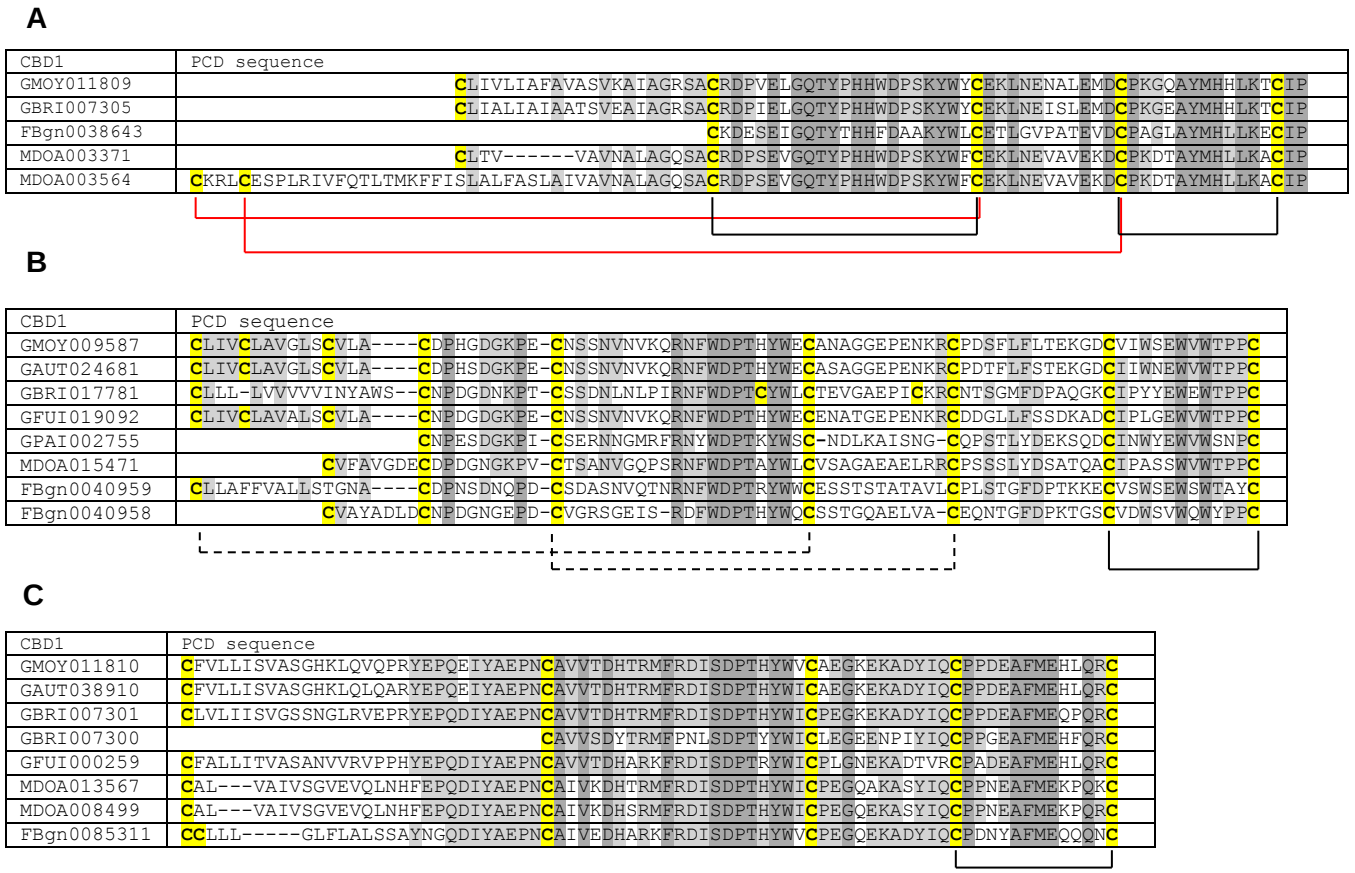


Figure 3.8. Alignments of PCDs from several Dipteran species against their respective orthologues. A; *The alignments of the CBD of GMOY01809 and its respective orthologues. B; Alignments of the CBD of GMOY009758 against its respective orthologues. C; Alignments of the CBD of GMOY011810 against its respective orthologues. Conserved cysteine residues are in bold and highlighted yellow. One hundred percent conserved residues are highlighted in dark grey whilst ≥60% of conserved residues are highlighted in light grey. Putative S-S bonds as predicted by multiple software programmes in agreement are indicated by solid lines and putative S-S bonds predicted by only one software programme are indicated by dashed lines. Red lines are predicted S-S bonds that are found in the larger orthologue of M. domestica.*

3.3.6 Homology modelling of CBDs of five known *G. morsitans* PM peritrophins

Using tachycitin as a guide, it was possible to accurately predict the structures of the chitin binding domains of peritrophins that have been identified as constitutes of the tsetse PM. CBD 1 and 3 from GMOY002708 (Fig. 3.9c and 3.9b respectively) and CBD 1 and 3 from GMOY009171 (Fig. 3.9d and 3.9e respectively), showed remarkable secondary structure similarity to the structure of tachycitin. All structures consisted of 3 anti-parallel beta sheets at the amino terminus, followed by a hevein-like fold consisting of 2 further anti-parallel beta sheets separated by a hairpin loop and followed by an alpha twist at the carboxy terminus. In the case of tachycitin, five S-S bond formations were predicted to occur between C1-C6, C2-C5, C3-C9, C4-C10 and C7-C8, and had a single tryptophan residue in the alpha twist near the C-terminus. The modelling software predicted only 2 disulphide bridge formations and that the cysteines involved in both CBDs of GMOY002708 and GMOY007191 were C2-C6 and C4-C5. Cysteine residues 1 and 3 were not predicted to form disulphide bridges. The presence of tryptophan residues were observed only in CBD3 of GMOY002708 between C4 and C5 and in CBD2 of GMOY007191 between C5 and C6.

The tertiary structure predictions for the peritrophins containing PCDs, i.e. GMOY009578, GMOY011809 and GMOY0011810 (Fig. 3.9f, g and h respectively), showed similarity to the structures of the C-terminus ends of the PADs. They consisted of two anti-parallel beta sheets separated by a hairpin loop although there was no alpha twist following at the C-terminus. However, the structures are similar to the hevein-like fold found in tachycitin and the other PADs. GMOY009578 is predicted to have a single disulphide bridge occurring between cysteine residues 9 and 10, whilst C8 does not appear to be involved in S-S bond formation. Two tryptophan residues are found at the N-terminus before C8. The structure of GMOY011809 is similar although it is predicted to have two disulphide bridge formations; one occurring between C6 and C7 and the other occurring between C8 and C9. It contains two tryptophan residues, both of which are between cysteines 6 and 7. The structure of GMOY011810 is more similar to the structure of Pro2 in that it has only one predicted disulphide bridge. The S-S bond formation is thought to occur between C8 and C9 near to the C-terminal end of the protein. It also contains only one tryptophan residue which is located on the first beta sheet before C7 at the N-terminus.

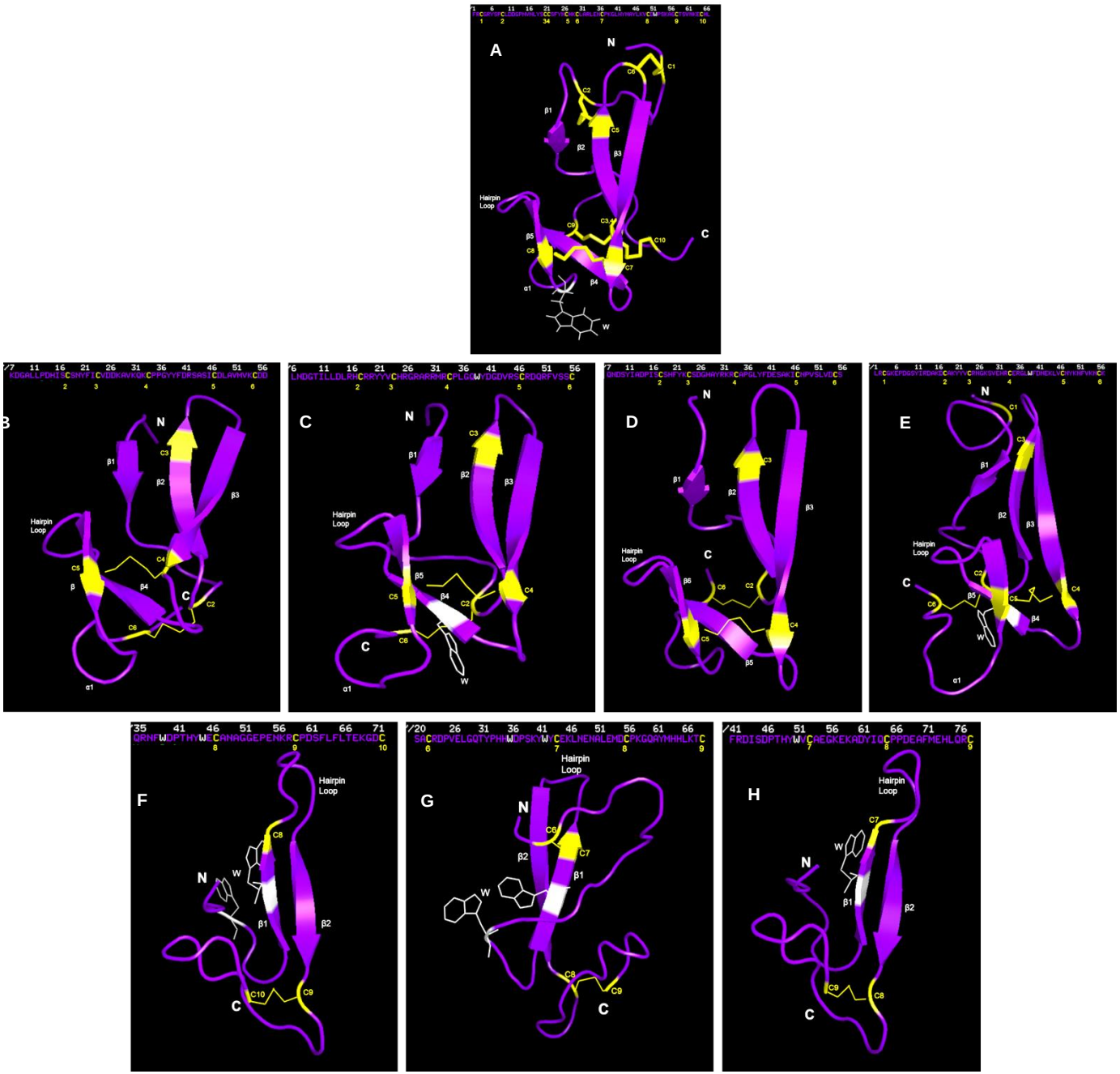

A
N
C
Hairpin Loop
B
C
D
E
F
G
H
W

Figure 3.9. Homology modelling of Glossina morsitans CBDs from known PM peritrophins using the chitin binding structure of tachycitin as a model. *A-H; Tertiary structure models of tachycitin (Protein Data Bank code 1dqc), CBD1 and CBD3 of GMOY002708, CBD1 and CBD3 of GMOY007191 and CBDs of GMOY009578, GMOY011809 and GMOY011810 respectively. Each shows the amino acid sequence of which the model was predicted from and the cysteine numbers predicted to be involved in disulphide bridge formation. Yellow residues in the structure indicate the respective cysteine residue as shown in the amino acid sequence. White amino acid projections are tryptophan residues.*

3.4 Discussion

The annotation of the tsetse genome started with whole genome shotgun sequencing of *G. m. morsitans* using *Drosophila melanogaster*, *Aedes aegypti*, *Anopheles gambiae*, *Culex quinquefasciatus,* and *Phlebotomus papatasi* as reference genomes [161]. Since being released, work has started on the annotation of several other major species of *Glossina* and a comparative species; *Stomoxys calcitrans*. The results presented in this study provide the first insights into the identification of peritrophins as a result of this annotation. By using reference genomes, errors in assembly can be minimised and the existing genome assembly is continually being improved for accuracy. Bearing this in mind, some of the putative peritrophins identified here may be incorrect particularly those that have been identified in *G. m. morsitans* but not in other *Glossina* species. This may be due to the other *Glossina* species having a more accurate and closely related reference genome (i.e. *G. m. morsitans*) on which to assemble than *G. m. morsitans* which, although used Diptera as reference genomes, they were of different families. Nevertheless the list of peritrophins described here is up to date and many peritrophin genes that have been identified in *Drosophila* as part of the PM structure can be found in *Glossina* species.

3.4.1 Characterisation of CBDs from Dipteran species

3.4.1.1 PADs

The consensus sequence from identified PADs from peritrophins of *Glossina* species, *M. domestica* and *D. melanogaster* falls into the previously characterised consensus of $CX_{13\text{-}20}CX_5CX_{9\text{-}19}CX_{10\text{-}14}CX_{4\text{-}14}C$, with a few exceptions having a further two cysteine residues either near to the N-terminus or the C-terminus. No additional cysteine residues were found in the middle portions of the PADs, giving 100% identity between all PAD orthologues at C2, C3 and C4 (Fig 2). Peritrophins containing PADs were the most common in all species analysed and indeed have been shown to be the most common type in other insects such as lepidopterans and coleopterans. Regions between each cysteine residue and between each CBD within the same peritrophin contains a large number of charged residues such as arginine (R), histidine (H), lysine (K), aspartate (D) and glutamate (E) interspersed with proline residues. Regions such as these have previously been shown to form tightly coiled structures which would bring each CBD into close proximity along the length of chitin bundles of the PM, giving a stability and strength to the PM. In addition, multiple trypsin and chymotrypsin cleavage sites have been predicted to be embedded within the tightly coiled PADs which would allow these proteins to be highly resistant to proteolysis by digestive enzymes. This, in combination with mucin domains that some peritrophins possess would further add to the robustness of the PM

3.4.1.2 PBDs

The consensus sequence of PBDs consists of $CX_{12\text{-}13}CX_{20\text{-}21}CX_{10}CX_{12}CX_{2}CX_{8}CX_{7\text{-}12}C$, and only 2 species contained this type of CBD; *G. brevipalpis* which had 1 and *M. domestica* which had 8. PBDs have so far been found only in one other dipteran species *Lucilia cuprina*. This species have been shown to possess 2 proteins which contain a PBD, although only one is shown to be from the PM. Interestingly, analysis of *Drosophila* peritrophins revealed no PBDs, suggesting this type of CBD evolved in *Musca* after the divergence of *Drosophila* and *Musca* species. Fig. 3.10 shows the predicted monophyly of the Glossinidae family, supported by molecular taxonomy, and shows it as a sister group to all Pupipara. *Glossina brevipalpis* of the *Fusca* species group emerges as a sister group to all remaining Glossinidae and is suggested to be the most ancient of all tsetse. The Muscoidea and blowfly species are predicted to be older than the Glossinidae, with *Musca* species being the oldest evolutionary. It may be that the appearance of PBDs in *Musca* occurred as they confer an advantage to this species but the disappearance of PBDs in all other *Glossina* species came about as they were no longer needed. It would be interesting to see if PBDs occur in all other species evolutionary older than *Glossina* in order to see the point at which they disappeared.

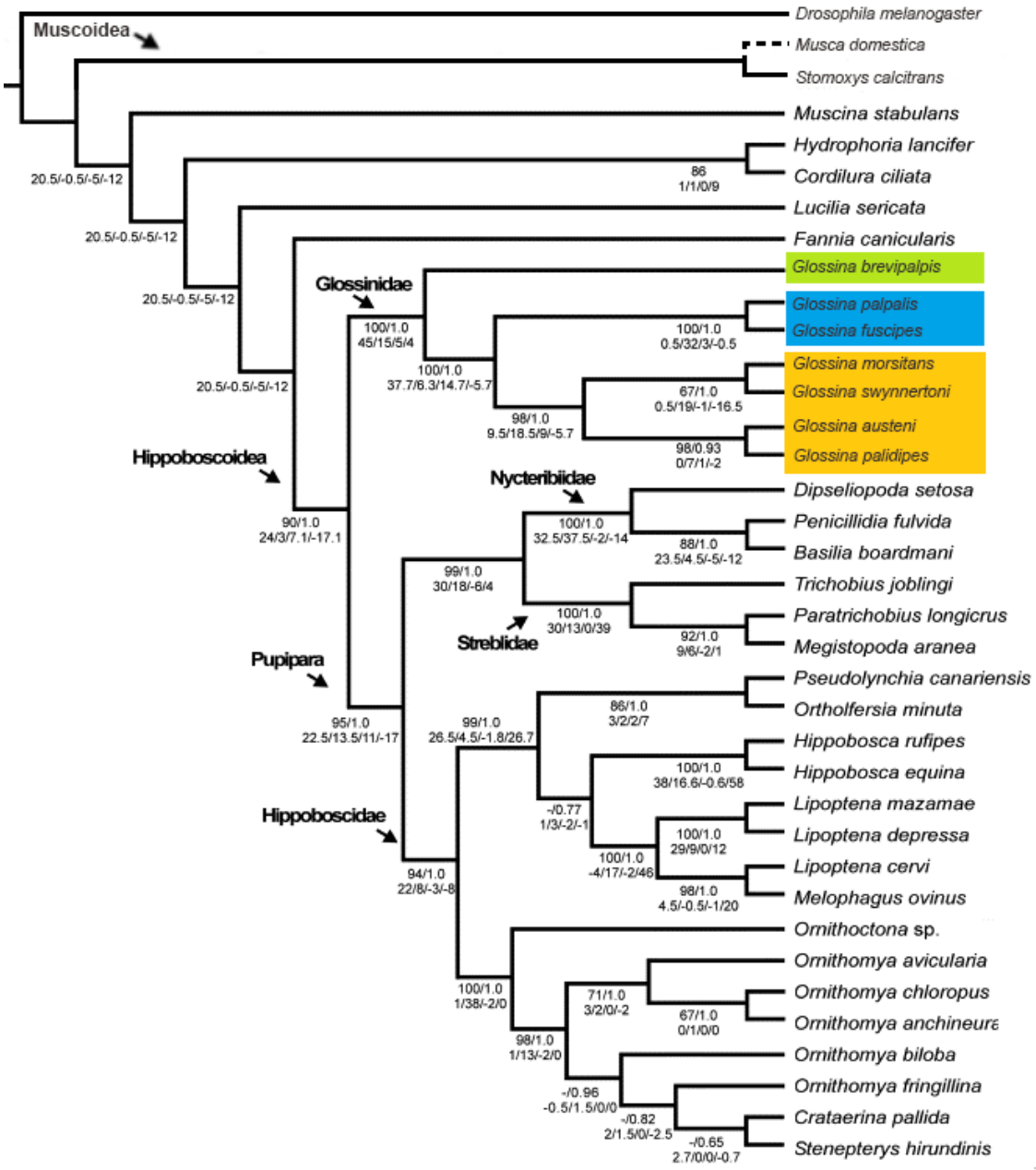


Figure 3.10. Phylogeny relationships between selected Dipteran species

Maximum parsimony tree based on combined sequence data from CAD, COI, 16s, and 28s (numbers in first line are bootstrap support values and posterior probabilities; − = bootstrap support <50; numbers in second line are PBS values for CAD/COI/16s/28s. The position of Glossina subgroups are depicted by colour; Fusca group – green, Palpalis group – blue and Morsitans/Austeni groups – orange. The predicted position of Stomoxys calcitrans and Musca domestica are depicted in relation to the Muscoidea. Adapted from [162].

3.4.1.3 PCDs

The consensus sequence of PCDs consists of $CX_{17}CX_{9}–_{10}CX_{14}CX_{9}CX_{8}–_{9}CX_{19}CX_{9}–_{11}CX_{14}CX_{11}C$, and all species analysed contained at least one peritrophin that contained this type of CBD. Although not as common as the PADs, PCDs have been shown to be important PM proteins. In tsetse, two of its 3 peritrophins containing PCDs are expressed in the proventriculus where they are thought to be involved in immunity in the fly. The other has been identified to be a component of the *G. morsitans* peritrophic matrix. *G. brevipalpis* contains the most peritrophins (8) with this type of CBD, whilst *M. domestica* has 14. Again, this suggests the possibility that the reduction in this type of CBD may be due to a lack of needed functionality. However, as they are not entirely obsolete suggests that the remaining ones play a vital role in the synthesis and/or structure of the PM. Indeed, PCDs are related to the peritrophin-15 family, which are known to cap the ends of chitin fibrils within insect PMs, giving added strength to the matrix [159].

3.4.2 Comparison of peritrophins between Dipteran species

From looking at the summary table (Table 3.3), it seems there are no significant differences between the peritrophins of the *Glossina* species. There are, however, slight differences in the numbers and types of peritrophins with *G. fuscipes* containing the most peritrophins with 44. In addition there is a huge difference between the numbers of peritrophins found in *Glossina* species to the number found in *Musca,* with more than double the number in the latter species. One possible explanation for this is the difference in host preference *G. fuscipes* has to reptiles and the feeding habits of *Musca*. Almost a quarter to a third of bloodmeals taken by *G. fuscipes* comes from reptiles, possibly more in human uninhabited regions and is this species' first host preference [163]. Whilst blood is normally sterile, reptiles are known to harbour many arboviruses and other pathogens [164]. Blood of humans and domesticated animals would be assumed to be more sterile even in areas of low vaccine coverage and little access to healthcare. *M. domestica* are mainly saprophagous (feeding on decaying organic matter) and coprophagous (feeding on faecal matter), but are also known to be carnivorous [165]. This therefore exposes them to multiple pathogens, some of which can be transmitted to humans and animals through eating food contaminated with the flies faeces or regurgitated gut contents. Due to the feeding behaviour of both *G. fuscipes* and *M. domestica*, it makes sense that these species would have a more developed defence system against invading pathogens, more so in the latter species. Moreover, there is a higher number of peritrophins found in *Drosophila* and most mosquito species than there are in *Glossina* species. *Drosophila* species also feed on decaying organic material and whilst female mosquitoes do feed on blood, both sexes also feed on nectar and plant sap; both of which are less sterile than blood. In addition, female mosquitoes of certain species are also

extremely competent vectors of arboviruses [166]. Similarly, *Stomoxys calcitrans* is an obligate bloodfeeder but do, however, take the occasional sugar meal obtained from flowers or ripe fruit [167]. They are also mechanical vectors for various pathogens including bacteria [160], viruses [168-172], protozoa [173-176] and helminths [177, 178]. In contrast to these, *Glossina species* transmit only one type of pathogen; African Trypanosomes. This, therefore suggests that peritrophins are needed not just for mechanical strength of the PM, but they have important roles to play in the defence against invading pathogens that are ingested with the meal and thus helps protects the insect in the midgut. As some peritrophins have also been shown to be involved in immune related functions in the midgut [179], it is conceivable that there are more peritrophins that also have similar roles.

Table 3.3. A summary of the properties of Dipteran peritrophins.

	Glossina morsitans	*Glossina austeni*	*Glossina brevipalpis*	*Glossina fuscipes*	*Glossina pallidipes*	*Musca domestica*	*Drosophila melanogaster*
Simple	25	29	25	31	21	57	49
Binary	1	3	3	2	5	3	4
Complex	4	2	2	3	1	17	9
Repetitive	6	6	6	8	8	25	28
Total	36	40	35	44	35	102	90
Mucins	11	13	9	12	11	32	22
PADs	33	35	26	41	34	80	80
PBDs	0	0	1	0	0	8	0
PCDs	3	5	8	3	1	14	13

3.4.3 Phylogenetic analysis of known *G. m. morsitans* peritrophins with orthologues of several Dipteran species

Previous analyses have shown that CBDs found in insect peritrophins and chitinases arose from a common ancestor and CBDs arose from chitinous networks that associated with proteins. It was suggested that ancestral insect midguts excreted mucus-like secretions. PM formation appeared due to the evolution of chitin-protein networks and chitin binding domains subsequently evolved leading to PM formation by restricted regions of the anterior midgut. Therefore the type I PM which is secreted by cells along the entirety of the midgut is thought to be the ancestral state and the type II PM evolved later on [33, 50]. The phylogenetic study here supports this theory as all genes analysed were from Dipteran species that secrete a type II PM and all were closely related. As expected all *Glossina* genes clustered together

and those genes in *Drosophila*, *Musca* and *Stomoxys* showed a higher degree of relatedness in comparison to *Glossina*. Interestingly, in some of the peritrophins analysed, both *Stomoxys* and *Musca* had multiple orthologues suggesting these peritrophins play an important role in the insects. As discussed previously, the Dipterans analysed all have different feeding habits, for example bloodfeeding in *Glossina* and *Stomoxys* vs non bloodfeeding in *Drosophila* and *Musca* and they also differ in their vectorial capacity; *Musca*, *Glossina* and *Stomoxys* transmit disease either mechanically, biologically or both, whereas *Drosophila* are not disease vectors. This suggests that peritrophins found in type II PMs may have evolved as a consequence of the feeding habits of individual insects and the likelihood of ingesting certain pathogens. Parallels between the PM of the midgut and the cuticle of the trachea and exoskeleton regarding molecular structure and function suggest these structures utilise proteins with similar properties. This is the case for CPAPs which are large proteins containing many CBDs and add to the structural integrity of the cuticle. As the cuticle is a major barrier to pathogen invasion this would suggest these proteins have similar properties which serve to act as a barrier to limit entry of pathogens whilst still being an effective interface to the external environment [180].

3.4.4 Homology modelling of Dipteran chitin binding domains

Tachycitin provided an excellent guide structure to model both the PADs and PCDs found in peritrophins identified from the tsetse PM. Structural analysis of these peritrophins also supports the model of PAD evolution as the closeness in structure homology suggests that all CBDs have a high degree of relatedness. Tachycitin was shown through nuclear magnetic resonance (NMR) spectroscopy to comprise of three anti-parallel beta sheets followed by a hevein-like fold that consists of two anti-parallel sheets separated by a hairpin loop and followed by an alpha helix region near to the carboxy terminus [181, 182]. The structures of the PADs from known peritrophins of *G. morsitans* were also predicted to follow the same general structural organisation. In the case of tachycitin, CBD_3 from GMOY002708 and CBD_32 from GMOY007191, tryptophan residues were shown to project between the disulphide bridges in the hevein-like fold [183]. This amino acid has previously been shown through NMR to have high binding affinity to GlcNAc residues, the monomer of chitin, and so this part of the PAD is most likely involved in the binding to chitin fibrils [184]. This is further supported by looking at the predicted structural models of the PCDs. These CBDs have no homology to the first 3 anti-parallel beta sheets but have extremely similar structures to the hevein-like folds of tachycitin and the PADs. They consist of two anti-parallel sheets separated by a hairpin loop, although there is no alpha-helical region near the C-terminus. In addition, each PCD contains at least one tryptophan residue. The homology models showed that tertiary structure between all CBDs was similar in the hevein-like fold and so suggests

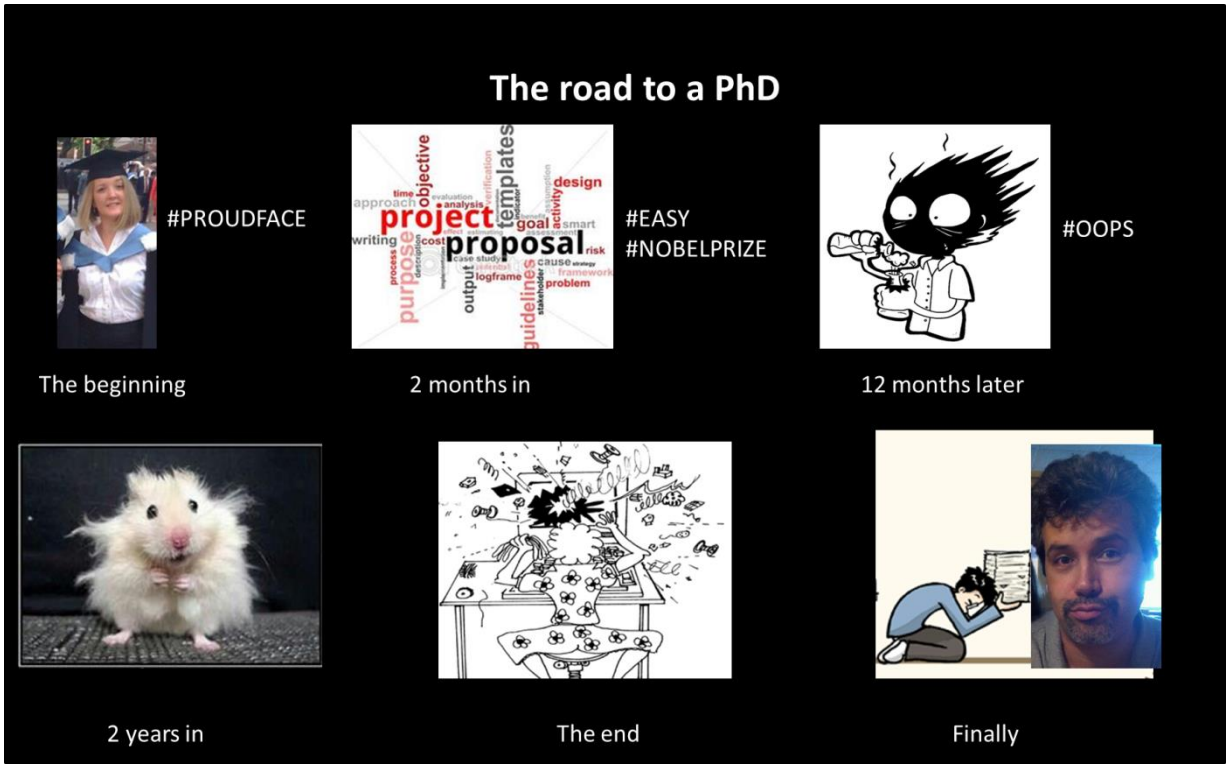
that it is this region that is important in the binding of peritrophins to chitin, whilst the rest of the protein possibly binds other proteins and carbohydrates.

3.4.5 Accumulation of metadata for the VectorBase community

Previous to the *Glossina* genome being sequenced, only two peritrophins had been identified through a proventriculus specific cDNA library. These were termed Pro1 and Pro2 as they were expressed in the proventriculus and are known to be involved in tsetse immunity. Subsequently, mass spectrometry analysis of PM derived proteins identified a further 3 peritrophins as being a part of the PM (Chapter 2). They were termed GmmPer66, GmmPer108 and GmmPer12 based on their theoretical molecular weights. However, because their orthologues in different species have slightly different amino acid sequences and differences in their post-translational modifications, the molecular weight of these orthologues are different. Therefore, to avoid confusion and to remain consistent with nomenclature, it was suggested that the term peritrophin-like-protein (PLP) should be adopted for putative peritrophins. All *Glossina morsitans* PLPs and their orthologues were added to a metadata form supplied by VectorBase (Ch3_ST1) and sent to the community portal of VB and will be included as part of the *Glossina* metadata when the geneset is next updated.

Chapter 4. Characterisation of selected *G. morsitans* peritrophins and targeting using RNA interference

It's a long, long road……..



4.1 Introduction

Species of the Glossinidae family are relatively poor vectors of pathogens in comparison to many other haematophagous insects such as mosquitoes, sandflies and ticks. Tsetse can ingest an infected bloodmeal but over half will have already cleared the infection by 4-5 days post infection, even under ideal laboratory conditions [185]. Many factors contribute to this so called self-cure including fly age and sex, fly species, trypanosome species, fly starvation status and presence of symbiotic bacteria [91, 186]. However, the actual mechanism still remains unknown, although young, teneral flies and older flies that have been starved are known to be more susceptible to infection [187]. One possibility for refractoriness or susceptibility to infection is the status of the peritrophic matrix at the time of infection. Younger flies have a shorter PM which only reaches the hindgut around 3-4 days after emergence which increases the chance of trypanosomes establishing an infection as they are less likely to be excreted quickly, whereas PM production in older flies that have been starved slows down dramatically [188]. When conditions are ideal for the fly, PM production normally occurs at a rate of 1mm per hour for the first 30 hours and also after feeding. After 30 hours post eclosion and when the fly is starved, the production rate of the PM declines significantly [188]. This would therefore almost certainly include a slower rate or disruption in the secretion of specific PM proteins. It has previously been shown that the peritrophic matrix of *G. morsitans* is comprised of many proteins (Chapter 2), including peritrophins which are integral to the PM structure [59]. These peritrophins may be a point of interaction between trypanosomes and subsequent breach of the peritrophic matrix. Studies have shown that the epimastigote forms of trypanosomes adhere to the salivary gland epithelial cells through complex membrane and cytoskeletal interactions, mediated between the parasite flagellum and the epithelial cell. It is therefore plausible that procyclic trypanosomes interact with the PM via binding of parasite ligand(s) to putative PM protein(s) [189-191]. Peritrophins may act as a receptor for trypanosomes as they are highly conserved between susceptible species (Chapter 3). Alternatively, peritrophins may act as a modulator of trypanosome infection depending on their production status. For example, establishment of infection in the fly may occur at a critical time point before production of peritrophins or before their insertion into the PM. To determine if peritrophins may play a role in susceptibility or refractoriness to trypanosome infection in *G. morsitans,* a preliminary attempt to characterise 2 peritrophins identified by mass spectrometry analysis (GmmPer66 and Pro2 - Chapter 2) was carried out using a combination of gene silencing by RNAi, expression profiling at a protein level and structural analysis by TEM of peritrophin deficient PMs after gene knockdown.

4.1.1 Glossina morsitans Proventriculin 2 (GmmPro2)

The peritrophin GmmPro2 (VectorBase ID: GMOY009587) is the most studied peritrophin of tsetse to date, mainly due to the lack of identification of tsetse PM proteins in general. It was only recently that a comprehensive study of the tsetse PM was carried out and identified 5 major peritrophins, 3 of which had not previously been identified (Chapter 2) [59]. First identified through a proventriculus-specific cDNA library [85], GmmPro2 was found to be synthesised and secreted by the proventriculus into the PM and appears to be a member of the peritrophin-15 family, inferred from homology [159, 192]. This family of proteins, present in the larval stages of Diptera, is suggested to bind to the ends of the chitin polymer within the PM thereby giving it structural support and protection from exochitinase action. In *G. morsitans,* GmmPro2 has a transcript length of 752 base pairs with 3 exons that translates to a protein of 116 amino acids with a predicted molecular weight of 12.9 kDa. It contains several cysteine residues that form a chitin binding domain (CBD) in the consensus of a Peritrophin C Domain (PCD), and the secondary structure suggests it is related to a lectin like protein called hevein (Chapter 3). It is thought to be highly modified due to the many high weight isoforms found through 2DE. At a protein level, the differential expression of GmmPro2 changes in response to fly maturation and feeding, suggesting it is not expressed consistently and is mediated by factors such as nutritional and infection status. It was suggested as a putative trypanosome receptor as studies carried out in Dr. Terry Pearson's lab (Victoria) has shown that a specific mAb for GmmPro2 (mAb4A2) significantly reduces trypanosome infection by up to 80% in comparison to control groups. In addition, it was found that GmmPro2 was overexpressed in mutant susceptible flies [193]. This suggests that GmmPro2 is a presumed receptor for trypanosomes enabling the parasites to come into close contact with the PM in order for enzymatic action to occur. However, a preliminary study has shown that injection of RNAi against GmmPro2 increases the likelihood of infection in teneral and non-teneral flies, (Fig. 4.1, unpublished), possibly by altering the structure of the tsetse PM. A more recent study showed that by reducing the expression of GmmPro2 through RNAi and causing a compromised PM, tsetse were more susceptible to trypanosome infection and were able to successfully clear bacterial infections by lack of inhibition to the fly's immune system [83]. Collectively, the results from these studies suggest that GmmPro2 is both a receptor and has a structural role in the peritrophic matrix.

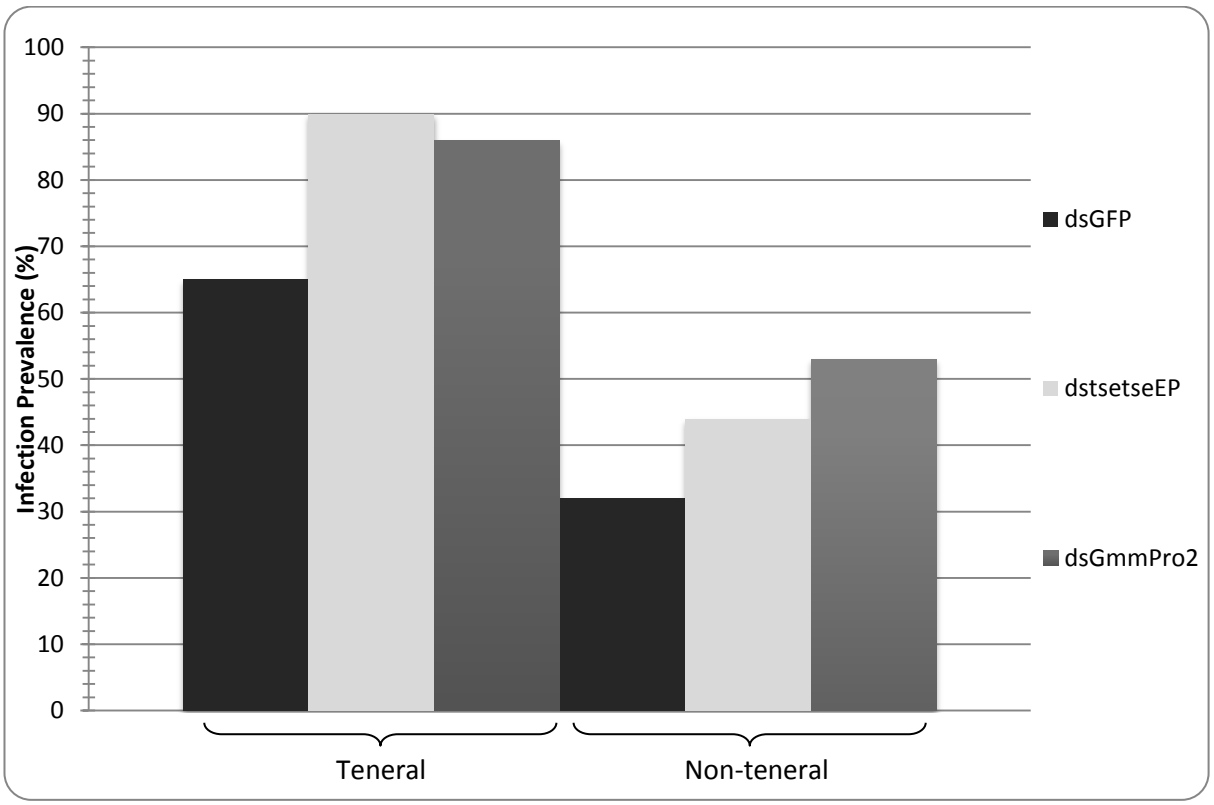


***Figure 4.1. Effect of GmmPro2 and tsetseEP knockdown on trypanosome infection rates in G. morsitans*.** *Percentage of infected flies from dsGmmPro2 injected group (n=54), dsTsetseEP group (n=35) and dsGFP group (n=35). Teneral refers to those flies that had received no blood meal before injection of dsRNA and non-teneral refers to those flies that had received one bloodmeal prior to injection*.

4.1.1.2 TsetseEP protein

A known PM protein that is also expressed elsewhere in the fly is tsetseEP (VectorBase ID: GMOY003306), which has been shown to have importance in modulating trypanosome establishment in the fly [100]. It is a protein of 306 amino acids with a large area of glutamic acid-proline dipeptide repeats that account for ~40% of the total protein sequence. This motif is also found on the surface of procyclic trypanosomes within procyclin proteins. Through RNAi it was shown that a reduction in the expression of tsetseEP caused a higher midgut infection rate compared to wild-type tsetse. In addition, recent observations have shown that when tsetseEP is knocked down in the fly, the expression of GmmPro2 is also significantly reduced (Haines, L, unpublished data), providing a useful positive control for GmmPro2 knockdown experiments.

4.1.1.3 GmmPer66

It was only recently that this peritrophin, GmmPer66 (VectorBase ID GMOY002708), was found to be a major component of the tsetse PM. It is 2125 bp in length which contains 5 coding exons that translate to a protein of 603 amino acids in length and has a predicted molecular weight of ~66 kDa. It is predicted to be highly modified and perhaps exists as isoforms as it is found at multiple molecular weights as seen by 1DE. It is the most abundant protein of the *G. morsitans* PM as identified through mass spectrometry. GmmPer66 contains 3 CBDs consisting of Peritrophin-A-Domains and has a large region of serine and threonine residues between CBD1 and CBD2 that are predicted to be heavily glycosylated and is similar to a mucin domain. It also has a smaller mucin domain between the 2nd and 3rd CBDs. Up until now there have been no studies conducted on this peritrophin.

This chapter aimed to identify the possibility of silencing genes encoding peritrophins, establish a timeframe for when the highest level of knockdown occurs, attempt to develop an antibody against the newly identified peritrophin, GmmPer66, and to identify protein expression profiles for each peritrophin. The results presented here show preliminary data with no major conclusions drawn for the function of the peritrophins. The experimental limitations and improvements that can be made are discussed, in addition to future work that should be carried out to determine the role these proteins have in the PM.

4.2 Materials and methods

4.2.1 *Glossina morsitans*

Flies were taken from an established colony of *Glossina morsitans morsitans* (Westwood) at the Liverpool School of Tropical Medicine that were maintained on sterile, defibrinated horse blood (TCS Biosciences) at an ambient temperature of 26°C±1°C and a relative humidity of 68-78%. Specific details for fly conditions are described for each experiment below.

4.2.2 dsRNA production

4.2.2.1 Plasmid identification

Templates for *tsetseEP, GmmPro2* and *GmmPer66* were available from the EST library maintained at the LSTM. They were identified by conducting a BLAST search using the cDNA sequence of each gene as identified in VectorBase against the *Glossina* EST library in NCBI and the largest ones with the most introns were used; *tsetseEP:* Tse11c09*, GmmPro2:* Tse7e09*,* and *GmmPer66:* Tse32c04*.* Regions of gene amplification were identified through sequence analysis and ran through a BLAST search at VectorBase and the non-redundant database at NCBI to ensure there were no other regions of similarity in *Glossina,* thus ensuring specificity of knockdown.

4.2.2.2 T7 primer design

Primers for RT-PCR for each gene were designed using the primer design software Primer3Plus using custom settings of optimum sequence length of 20bp, minimum of 18bp and maximum of 27bp, Tm of between 50-60°C and a GC clamp of between 40-60% that would produce a product size of 250-400bp. The T7 promoter sequence at the 5 prime terminal of the forward and reverse primers was added (Table 4.1).

Table 4.1. List of primers used for RT-PCR in the synthesis of dsRNA. *T7 promoter sequence is shown in red*

Primer name	Sequence
T7EP_F	TAATACGACTCACTATAGGGTTCTGGCAAACCCTCAAT
T7EP_R	TAATACGACTCACTATAGGGCTACGATAAATATGTCCCTCTAAT
T7GmmPro2_F	TAATACGACTCACTATAGGGAAGCTGCATTTTGTTTGATTGT
T7GmmPro2_R	TAATACGACTCACTATAGGGAGCGCTTGTTCTCTGGTTC
T7GmmPer66_F	TAATACGACTCACTATAGGGTAAAACGGAATGACTTGTGC
T7GmmPer66_R	TAATACGACTCACTATAGGGAGTGGTTCTTGGTGTATTCG
T7GFP_F	TAATACGACTCACTATAGGGACGTAAACGGCCACAAGTTC
T7GFP_R	TAATACGACTCACTATAGGGCTTGTACAGCTCGTCCATGCC

4.2.2.3 Plasmid growth and transformation

Approximately 2µl of each plasmid was transformed in 500µl of XL1-blue competent cells by incubation on ice followed by heat shock to promote plasmid uptake. Antibiotic resistance was induced by vigorous shaking in ampicillin containing lysogeny broth (LB) and plated overnight on ampicillin plates. Competent cells containing no plasmid were also plated as a control. Single, isolated colonies were cultured in LB medium with ampicillin overnight, to amplify the number of bacterial cells which were then harvested by gentle centrifugation.

4.2.2.4 Plasmid purification

The DNA plasmids were purified using the QIAprep Spin Miniprep kit (quiagen) as suggested by the manufacturer's instructions. Following purification, subsequent released DNA was eluted with warm nuclease-free water (NFW). The concentration of each eluate was measured using a Nanodrop ND-1000 spectrophotometer and plasmid (pT3T&-pac) presence and size (~3000bp) was confirmed by electrophoresis at 90 V on a 1% agarose gel.

4.2.2.5 Amplification of target DNA

An RT-PCR reaction was set up using the selected purified plasmids as template and the gene-specific primers with the T7 RNA polymerase promoter sequence for amplification of the specified gene product. PCR cycling conditions were as follows; 30 seconds for initial denaturation followed by a second denaturation step at 95ºC for 1 minute. The annealing step was carried out at 55ºC for *GmmPer66*, 51.3ºC for *tsetseEP* and 56.4ºC for *GmmPro2* for 30 seconds. The second denaturation and annealing steps continued for 34 cycles before elongation at 72ºC for 1 minute and a final elongation step of 5 minutes at 72ºC. The holding step was carried out at 4ºC. Each PCR product was visualised on a 1% agarose gel before purification.

4.2.2.6 Purification of PCR product and sequencing

The PCR products were purified using the QIAquick PCR purification kit according to the manufacturer's instructions and eluted with warmed NFW. Concentrations of samples were measured on a Nanodrop and the products visualised on a 1% agarose gel. To confirm a successful amplification of the target sequence of each gene 1ng/µl (per 100bp) of purified PCR product (total 20 µl) and 3.2pmol/µl of both F and R primers (15 µl) were sent off for Sanger Sequencing at lifesciences.sourcebiosciences.com. The received sequences were checked by conducting a BLAST search against the *Glossina morsitans morsitans* transcriptome database at VectorBase

4.2.2.6 dsRNA synthesis

The purified PCR products were used to generate dsRNA using the MEGAscript T7 Transcription kit according to manufacturer's instructions and the reaction was left to incubate overnight. The reactions were then incubated at 75ºC for 5 minutes and left to gently cool to room temperature to ensure hybridisation of RNA into dsRNA and a small amount (~1:500 dilution) of each was checked on a 1% agarose gel to examine the integrity and efficiency of the duplex formations. The addition of DNase I was carried out to ensure degradation of any template DNA still present, but did not proceed longer than 1 hour to prevent the degradation of the dsRNA products. The final products were cleaned using the Ambion MEGAclear clean up kit to remove free nucleotides, proteins, salts and nucleic acid degradation products and RNA concentration determined using Nanodrop (RNA- 40). Previous studies have shown that successful RNAi in *Glossina* usually requires each fly to be treated with ~2 µl of 5 µg/µl dsRNA (10 µg dsRNA per fly), but can vary with each knockdown due to lethal effects on the fly (for example tsetseEP). Each sample was either concentrated or diluted to an approximate concentration of 5 µg/µl (*GmmPro2*; 4.983 µg/µl, *GmmPer66*; 4.741 and tsetseEP; 2.9 µg/µl).

4.2.3 dsRNA injection in *Glossina morsitans*

4.2.3.1 Time course experiments

Male flies less than 24 hours post eclosion (h.p.e) were fed a bloodmeal and intrathoracically injected with each gene (*n*=60 each) and a GFP control 24 hours later using an established technique (Ch4_SI_1). Five flies per day from each group were sacrificed and their intact midguts snap frozen in liquid nitrogen, up to a total of 8 days (11 days for GmmPer66), and stored at -80^{o}C until use. Flies were constantly rotated throughout dissection and feeding to ensure the absence of blood in the midguts at the point of sacrifice.

4.2.4 RNA extraction from tissues

Total RNA was extracted from a pooled total of 4 midguts per group by homogenisation of frozen tissue in Trizol (Invitrogen, Paisley, UK), before the addition of chloroform. Samples were centrifuged at 4^{o}C and the aqueous phase (upper and transparent) was transferred to separate RNase-free tubes with care taken not to disturb the interphase layer containing DNA. Isopropanol was added to each sample, centrifuged at 4°C, supernatant removed and the subsequent pellets washed with 75% ethanol. After vortexing and centrifugation, pellets were briefly dried at room temperature and re-suspended in NFW at 55°C to dissolve RNA and standardised to 25ng/µl after determining concentration.

4.2.4 cDNA synthesis

Approximately 5 µg of total RNA from each sample group was used to synthesise cDNA using the Superscript III Reverse Transcriptase kit (Invitrogen) according to manufacturer's instructions. Concentrations of each sample were measured before storage at -4°C until use.

4.2.5 qRT-PCR

qRT-PCR was used to measure the relative transcript abundance of tsetseEP, GmmPer66 and GmmPro2 transcript in knockdown and control flies. Analyses per treatment were performed in triplicate using Brilliant III Ultra-fast SYBR Green master mix with the Agilent Mx3005P system, with the housekeeping genes α and β tubulin used to normalise the amount of cDNA in each sample. Primers for qRT-PCR were designed using OligoPerfect™ Designer software through Life Technologies with the following stringent parameters; primer length of 19-22bp, a GC content of 40-60%, a Tm of 60°C ± 3°C, preferentially spanning an am intron to show any genomic contamination and an ideal product size of 80-150bp. Potential primer sets were then analysed by OligoAnalyzer 3.1 software, which is a server ran through integrated DNA technologies. Primer sets were only used if the hairpin ΔG was ≤ -2 and self-dimers and heterodimers had a maximum ΔG of -5 at the 3′ end and -6 internally. The primers used can be seen in table 4.2. Data analysis was performed using the computer softwares MxPro and Microsoft Excel. Percentage knockdown level per gene was calculated using the Pfaffl method based on the efficiency of the specific primers, normalised against α and β tubulin and compared against the standard controls (tissues from dsGFP injected flies).

Table 4.2. List of primers used for qRT-PCR to measure relative transcript abundance.

Primer name	Sequence 5′- 3′
TsetseEP_F	ACCGTTCGTTTCGCTTTACTAC
TsetseEP_R	ACCTGCAGCCGTTTGACTTTC
GmmPro2_F	AAGCTGCATTTTGTTTGATTGT
GmmPro2_R	CTTTTTATTTTGCGAAGAGTC
GmmPer66_F	ACACCAGCAAAGGGGATACA
GmmPer66_R	CCTGTTTACTCGCTTGTGTTCG
αtubulin_F	TGTATGTTGTATCGTGGTGATGT
αtubulin_R	GAATTGGATGGTGCGTTTAGTTT
βtubulin_F	CCATTCCCACGTCTTCACTT
βtubulin_R	GACCATGACGTGGATCACAG

4.2.5 Antibody production for GmmPer66

The amino acid sequence of GmmPer66 was analysed in DNASTAR's protean package. Regions of hydrophobicity and antigenicity were identified and regions of similarity to other *Glossina* proteins were avoided. These included avoiding the three conserved chitin binding domains and the heavily glycosylated mucin domains. Subsequently, 3 peptides were suggested by Eurogentec, a company specialising in custom antibody production. A homology search (BLAST) was carried out in order to optimise any wanted or minimise any unwanted homology of the immunising peptides to other proteins. A 16 amino acid peptide was chosen to proceed to immunisation stage. After coupling to keyhole limpet hemocyanin (KLH), a metalloprotein used as a carrier protein from the giant keyhole limpet *Megathura crenulata*, five rabbits were immunised under the speedy programme (28 days). Pre-immunisation, the rabbits were bled and the serum was used to screen for cross-reactivity to *Glossina morsitans* midguts. Following this, 2 rabbits were chosen to continue to immunisation, each of which had 3 immunisation boosters and ELISA testing was carried out after the pre-immune and medium bleeds. Purification of the antibody occurred after the final large bleed using affinity specific IgG purification. The end products included serum from pre-immune, medium bleed anti-serum, final bleed anti-serum and purified antigen, of which all sera were screened against *Glossina* midgut tissue before use.

4.2.6 Immunoblotting of GmmPer66 and GmmPro2

4.2.6.1 Midguts from knockdown flies

One midgut from each knockdown group (i.e. GmmPro2, GmmPer66, tsetseEP and GFP) was homogenised in Laemmli buffer, boiled and an equivalent of ½ a midgut was loaded and ran on a 12.5% polyacrylamide gel (details can be found in Chapter 2). Immunoblotting was performed as in Chapter 2 with mAb4A2 used as the primary antibody for GmmPro2 at a 1:500 dilution and anti-serum from one immunised rabbit number SY6235 used as the primary antibody for GmmPer66 at a 1:100 dilution. A 1:50,000 dilution of secondary antibodies was used; goat anti-mouse for GmmPro2 and mouse anti-rabbit for GmmPer66. Blots were developed by chemoluminescence which continued for up to 18 hours.

4.2.6.2 Tissues for protein localisation

Anti-serum from rabbit number SY6235 was used to probe different tsetse tissues to identify where the protein is localised to. The following tissues and numbers of each are as follows; 10 crops, 5 PVs, 10 PMs, 1 midgut, 1 midgut minus the PM, 5 fat bodies, 5 flight muscle, 5 testes and 10 salivary glands. All tissues were harvested from 48 hour old, unfed male flies and immunoblot development continued for up to 18 hours

mAb4A2 was used to probe different tissues from 20 day old female flies using the following tissues; 6 salivary glands, 6 PVs, ½ a midgut, ~2µl haemolymph, 1 ovariole/uterus, and 1 fat body. Immunoblot development continued for up to 3 hours.

4.2.6.3 Tissues for differential expression of GmmPer66 between teneral and non-teneral flies

To see if there were differences in the expression of GmmPer66 in the midguts of flies that were unfed and fed, several tissues were harvested. 10 PMs and 2 midguts minus PMs were dissected from 24 hour unfed males and 5 day old males that had received one bloodmeal. Immunoblots were carried out using anti-serum from SY6235 and development continued for up to 15 hours.

4.2.7 Electron microscopy of *GmmPro2* knockdown flies

4.2.7.1 Structural analysis

To determine if GmmPro2 has a major structural role in the tsetse PM, midguts of flies that had been injected with dsGmmPro2 using dsGFP as a control were processed for TEM as described in detail in Chapter 5. Five flies of each group were dissected 6 days after RNAi injection as determined by the time course experiment as being the point at which knockdown is greatest.

4.3 Results

4.3.1 Antibody production of GmmPer66

Previous mass spectrometry studies revealed that the most abundant PM protein was a peritrophin termed GmmPer66 (Chapter 2). An antibody to this peritrophin was developed using synthetic peptides which were used to immunise rabbits. Identification of possible, suitable peptides was sought through sequence analysis and looking at hydrophilicity (Kyte-Doolittle plots) and antigenicity indexes (Jameson-Wolf), as well as the surface probability plots (Emini) (Fig. 2). GmmPer66 contains 3 highly conserved chitin binding domains that can be found in many other proteins such as other peritrophins, chitinases, chitin deacetylases, cuticular proteins and cuticular proteins analogous to peritrophins (CPAPs) and so have to be avoided. Similarly, the large mucin domain between CBD 1 and 2 and a smaller mucin domain between the 2nd and 3rd CBD had to be avoided due to the high degree of hydrophobicity. Previous peritrophin analyses have shown that only the CBDs of the proteins are conserved and regions outside of these are not, allowing for the design of an antibody that targets such regions. Subsequently, three peptides in the regions between amino acids 430 and 515 where there was high hydrophilicity, antigenicity and a good surface probability (Fig 4.2 and 4.3);

1. aa 450-464: C+ ETPTTKTTKIPSIKP
2. aa 462-477: IKPSLPPNTPRTTVSC
3. aa 498-513: TAPIQETTSTQDSPTT

A Cysteine (C+ at the N-term of peptide 1) was suggested as an addition to target the coupling site at the carrier protein (KLH). However, considering all of the peptides properties, it was finally recommended that the peptide most likely to produce the best immunogenicity was peptide 2. The final peptide that was synthesised for immunisations was;

H - IKPSLPPNTPRTTVSC - NH2 (16AA)

For antibody production, the best choice for an internal peptide for the C-terminus is CONH2 which mimics the uncharged peptide bond in a protein.

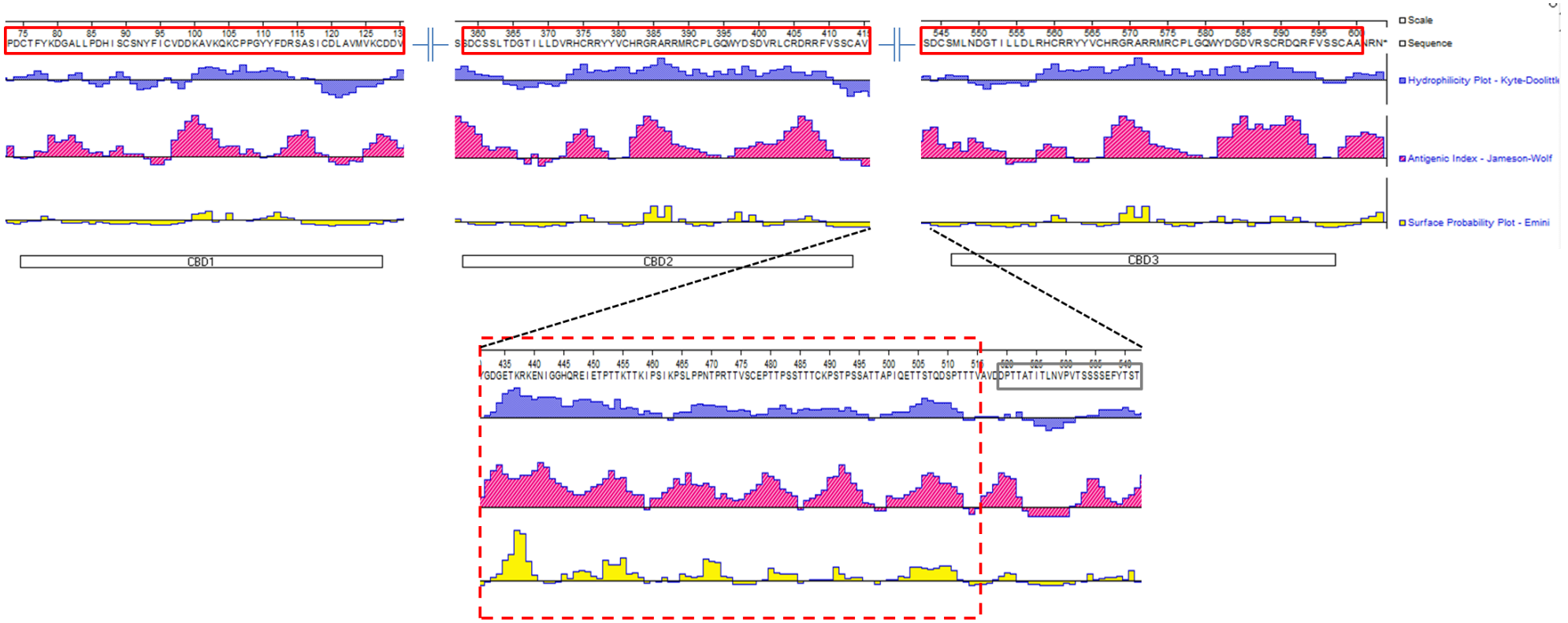


Figure 4.2. Peptide property plots of GmmPer66. *The top panel represents the 3 peritrophin-A-domains (PADs) of GmmPer66 (as shown by red boxes). Within each there are areas of high hydrophilicity (blue plots) and antigenicity (pink plots) but relatively little surface probability (yellow plots). Inset indicates the region, aa 430-515, proposed for antibody production (red dashed box) between the 2nd and 3rd chitin binding domains. The grey box indicates the small mucin domain which is unsuitable for antibody production.*

MMELGDEFPWFAEISCLVSNISESIVHITYTPYIRFWRGSIRNNMKIYLLLLFTVQSVPT
YVRAYARGLRYVPD**C**TFYKDGALLPDHIS**C**SNYFI**C**VDDKAVKQK**C**PPGYYFDRSASI**C**D
LAVMVK**C**DDVYYGETSMGHSLFGSYFPHKWIPNIFFGKKYAHDDGSVIDSPSATTTCKPN
TPSSTTTTCEPNTPPITPPITTPDPITPSSTTTTCEPNTPPITPPISTPDPNTPSSTTTT
CEPNTPPITPPISTPDPNTPSSTTTTCEPNTPPITPPIATPDPITPSSTTTSCEPSTPSS
TTTSCEPSTPSSTTTSCKPSTPSITTGDDSTIGTTAADTTTIEANSNERPLNFYRRSSD**C**
SSLTDGTILLDVRH**C**RRYYV**C**HRGRARRMR**C**PLGQWYDSDVRL**C**RDRRFVSS**C**AVSLWFL
LYSLTLMSLIYGDGETKRKENIGGHQREIETPTTKTTKIPS***IKP***SLPPNTPRTTVSCEPT
TPSSTTTCKPSTPSSATTAPIQETTSTQDSPTTTVAVDDPTTATITLNVPVTSSSSEFYT
STSD**C**SMLNDGTILLDLRH**C**RRYYV**C**HRGRARRMR**C**PLGQWYDGDVRS**C**RDQRFVSS**C**AA
NRN

Figure 4.3. Amino acid sequence of GmmPer66. *Yellow highlighted regions show the 3 chitin binding domains as indicated by the six conserved cysteine residues in each (bold and underlined). Grey highlighted regions show the putative mucin domains and the underlined peptides regions indicate the 3 suggested peptides for GmmPer66 antibody production. The bold, italicised amino acids (IKP) marks the overlap of two of the suggested peptides.*

Before immunisation of rabbits, pre-immune serum was used to identify if there was any cross-reactivity to *Glossina* midgut and PM tissues (Ch4_SF1). Subsequently, 2 rabbits were chosen to proceed to immunisation stages as their pre-immune serum caused no visible cross reactivity.

4.3.2 GmmPer66 localisation to *Glossina morsitans* tissues

Once full immunisation was complete, anti-serum from a small bleed after one booster and anti-serum from a medium bleed after 3 boosters from each rabbit was tested on an equivalent of 2 midguts each at concentrations of 1:10, 1:100, 1:500 and 1:1000 dilutions (data not shown). The best reactivity was from the large bleed anti-serum after 3 boosters from rabbit number SY6235 at a concentration of 1:100. Tissues from teneral flies (48 hours) were dissected and used for GmmPer66 expression profiles (Fig. 4.4). The banding pattern was similar in the tissues from the gut i.e. the PV, crop, PM, midgut and midgut minus the PM, with a major band across all samples with a Mr of approximately 45 kDa. Surprisingly, the tissue showing the most signal was the crop, with a banding profile that showed expression at 10-260 kDa. The PM had the highest signal of the tissues although, there were less distinctive banding patterns than those in the crop. The signal seen at a high Mr in the PM was missing in the midgut and midgut minus the PM. There was also small amounts of

signal that was detected in the fat body, flight muscle and testes, but the salivary glands doesn't appear to express GmmPer66.

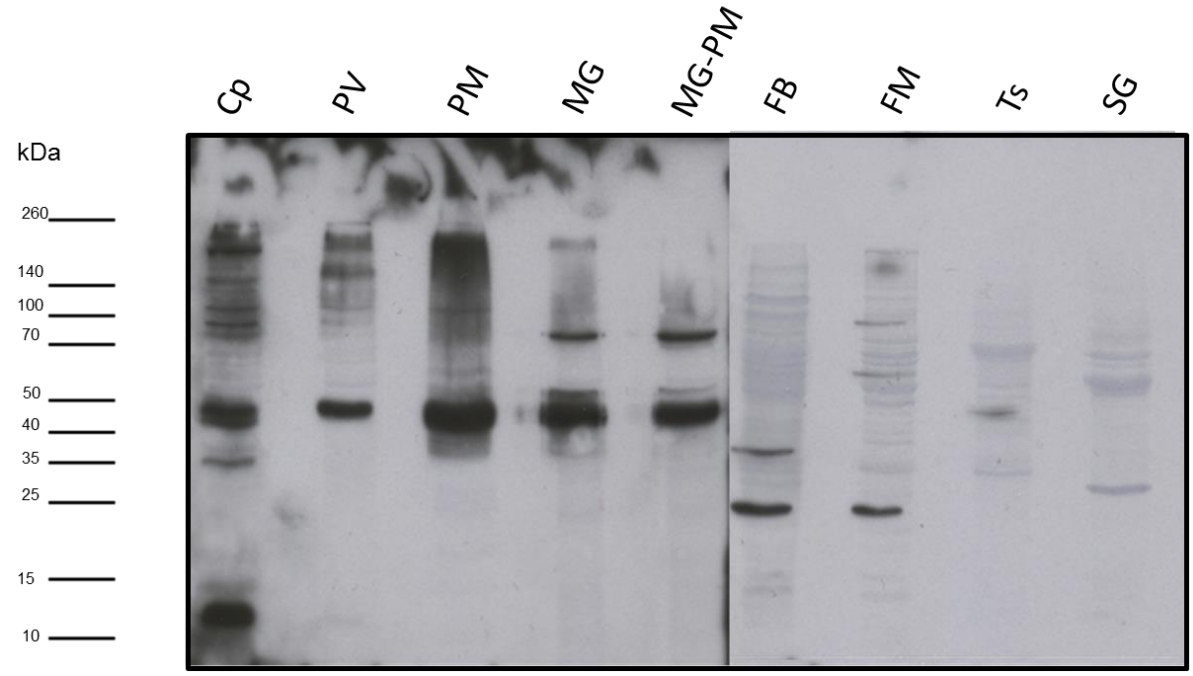


Figure 4.4. Protein expression profiles of GmmPer66 in various tsetse tissues. *Protein extracts from tissues boiled in Laemmli were used to determine protein expression in male Glossina. MWM; molecular weight marker, Cp; Crop, PV; Proventriculus, PM; Peritrophic Matrix, MG; Midgut, MG-PM; Midgut minus peritrophic matrix, FB; Fat body, FM; Flight muscle, Ts; testes and SG; Salivary gland*

4.3.3 Differential midgut and PM expression of GmmPer66 in teneral vs non-teneral flies

In order to determine if the level of protein expression or abundance changes with age or feeding status of the fly, western blot analyses was carried out on midguts (excluding crop, PM and PV) and on isolated PMs on 48 hour old males and 5 day old males that were fed once (Fig. 4.5). The analyses showed that GmmPer66 expression was highest in the peritrophic matrixes of teneral flies and lowest in the midguts of teneral flies. After 5 minutes exposure, most of the teneral PM expression profile can be seen, whereas only faint bands can be seen in teneral midguts and blood fed PMs and midguts appear after 5 minutes exposure. The major bands seen in the teneral PMs occur from 140 kDa to as little as 15

kDa, whereas there appears to be only one major band at 140 kDa in teneral midguts. Interestingly, following a bloodmeal, protein expression of GmmPer66 decreases significantly in the PM and has a slight increase in the midgut. This difference is more obvious in blots exposed for 30 minutes. The major bands in the PM of a blood fed fly appear at 140 kDa and ~37 kDa, whereas there are four major bands of a midgut from a blood fed fly; 140, ~66, ~54 and ~37 kDa. However, these results must be met with caution as the banding profile seen here does not match the PM or midgut profiles seen in Fig. 4.4 suggesting non-specific antibody binding. The implications for this are discussed further in the chapter.

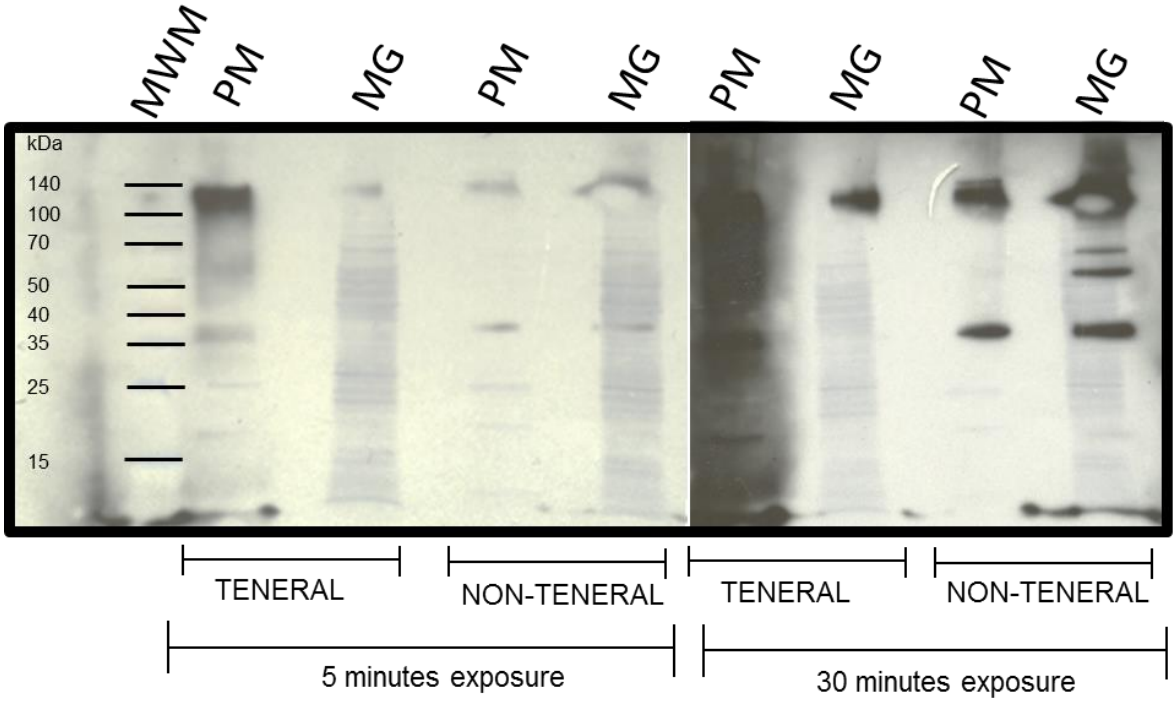


Figure 4.5. Immunoblot of GmmPer66 against several tsetse tissues in teneral and non-teneral flies. *Tissues were boiled in Laemmli buffer, separated on 12.5% acrylamide-SDS gel and blotted using anti-GmmPer66. Film was exposed to the membrane for 5 minutes or 30 minutes before development. Approximately 2 midguts (MG) minus PM crop and PV, and 10 peritrophic matrixes (PM) were loaded. Molecular weight marker (MWM) is shown on the left hand side. Blots were superimposed on nigrosine stained membranes to show protein loading and signal in relation to protein profile.*

4.3.4 GmmPro2 localisation to *Glossina morsitans* tissues

As previously stated, GmmPro2 is the most studied *Glossina* peritrophin and as such has been subject to numerous characterisation studies. Protein localisation has been shown for the crop, PV, PM and midguts of male teneral and non-teneral flies, where the expression did not change significantly between teneral and blood-fed flies. This is consistent to a transcript level analysis of GmmPro2 which found no changes to GmmPro2 transcript levels between flies before and after feeding. The only difference observed between tissues was the appearance of a lower molecular weight band in the midgut, which corresponds to the proteins predicted molecular weight of ~13 kDa. Due to male tissues having already been analysed, female tissues were analysed in order to determine any significant changes in expression profiles (Fig. 4.5). The results indicate that GmmPro2 has the same expression in females as in males, with the lower Mr band showing after a 3 hour exposure. Interestingly, GmmPro2 is expressed in the haemolymph but is not found in the salivary glands. The higher molecular weight signal was the strongest in the PV and midguts but was also quite strong in the ovarioles.

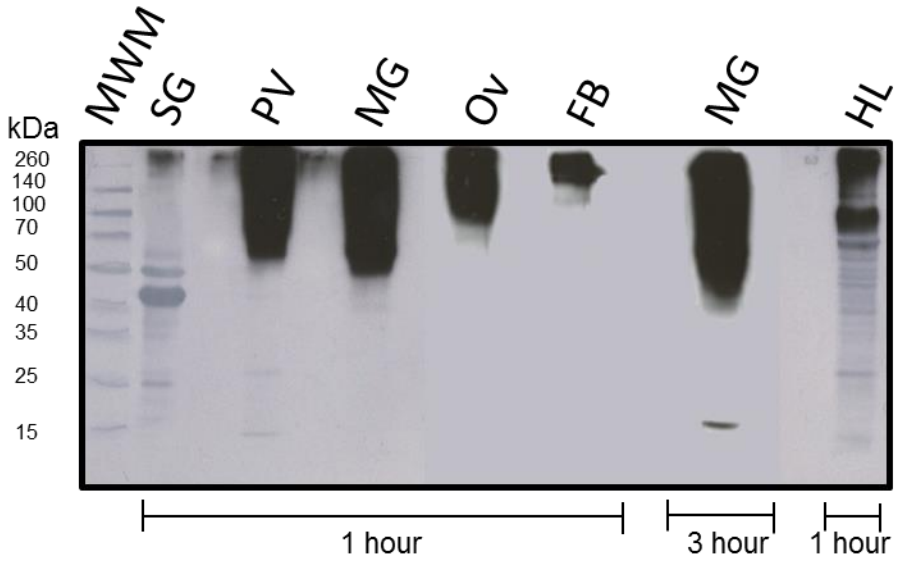


Figure 4.4. Protein expression profiles of GmmPro2 in various tsetse tissues. *Protein extracts from tissues boiled in Laemmli were used to determine protein expression in female Glossina. MWM; molecular weight marker, SG; Salivary gland, PV; Proventriculus, MG; Midgut, Ov; Ovarioles, FB; Fat body and HL; Haemolymph*

4.3.5 Effect over time of dsPer66 on the protein expression in the PMs of tsetse

Although feeding dsRNA will specifically target genes in the midgut, preliminary studies have shown injection usually works better. The latter ensure systemic down regulation, and the dsRNA is not affected by midgut enzymes. In addition, the exact quantity of dsRNA is known and can be fully delivered to the fly, whereas if the flies ingested the dsRNA they are almost certainly ingesting differing amounts. Even if the concentration was adjusted to accommodate the flies that would ingest the least, the production of dsRNA would need to be increased significantly. For these reasons, flies were injected; one group with dsGmmPer66 and a second group with dsGFP and 5 pooled PMs from each group were dissected each day up to a total of 11 days post injection. Western blot analyses were carried out on the pooled PMs for each group using the anti-GmmPer66 antibody (Fig 4.5). The results indicate that there were similar levels of protein expression for days 1-5 after injection with a reduction in GmmPer66 on day 2 in both the experimental and control group. However, on day 6 there seems to be a reduction in expression in the dsGmmPer66 group as compared to the control. This is more obvious on days 7-10 where there is very little, if any, signal, although the noise from the dsGFP on day 7 is obscuring the dsGmmPer66 signal. By day 11, protein expression is suddenly back to similar levels as day 1 in both groups. In days 1, 3 and 11, a small band can be seen at a high molecular weight (>100 kDa) in dsGmmPer66 and in day 3 dsGFP. Again, at this point it is unsure whether or not the anti-GmmPer66 antibody is specifically binding to GmmPer66 as the banding profile is different than that seen in Figs 4.4 and 4.5. If specific, the banding patterns should be the same, at least for the dsGFP control, and so no reliable conclusions can be drawn from this. In addition, for this analysis knockdown levels were not able to be verified by qPCR and so any decrease in signal from the Western blot may not be accurate. These points are expanded on in the discussion.

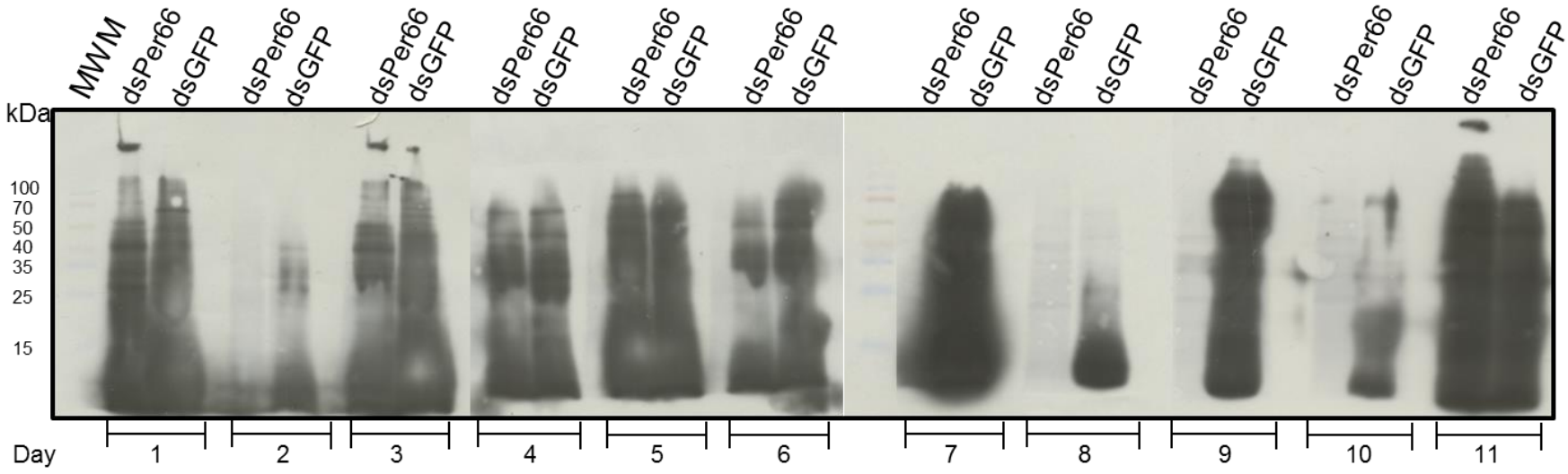


Figure 5.5. Western blot analyses of the effect on knockdown of GmmPer66 in the tsetse PM over time. *Pooled PMs were boiled in Laemmli and loaded on a 12.5% gel. Immunoblots were carried out using anti-GmmPer66. Days 1-6 were exposed for 10 seconds and days 7-11 were exposed for 30 seconds.*

4.3.6 Effect over time of dsPro2 on the protein expression in the midguts of tsetse

4.3.6.1 Transcript levels as determined by qPCR

The transcript levels of *GmmPro2* and *tsetseEP* were measured over a time course of 8 days under dsGmmPro2 and dsTsetseEP knockdown conditions (Fig. 4.6a and b respectively). When midgut transcript levels of *GmmPro2* were measured after injection of dsGmmPro2, the transcript level started to decline as early as 2 days after injection when compared to the GFP control. They continued to fall quite rapidly up to 6 days post injection and then gradually started to come back. Injection of dsGmmPro2 had no bearing on the transcript levels of *tsetseEP.* The qPCR data shows that 2 days after dsGmmPro2 injection, the transcript level of *tsetseEP* drops slightly, although insignificantly, but then increases at day 3 and remains relatively consistent. In contrast, when *tsetseEP* is knocked down in the midgut transcript levels of both *tsetseEP* and *GmmPro2* decrease. As expected, the reduction in transcript is pronounced for *tsetseEP*, with a slight, but insignificant decrease in *GmmPro2*transcript levels. There are 2 anomalies where the transcript level of GmmPro2 appear to spike, however, overall the transcripts are reduced. Transcripts for tsetseEP remain consistently low throughout the time trial and do not appear to start to increase even after 8 days.

4.3.6.2 GmmPro2 protein expression after dsRNA injection

Protein expression was measured through immunoblotting with anti-Pro2 (mAb4A2) against midguts from flies injected with dsGmmPro2, dsTsetseEP or dsGFP over a period of 8 days (Fig. 4.6c). Levels of GmmPro2 expression was consistently higher across all time points in flies that had been injected with dsGFP in comparison to those that had been injected with dsGmmPro2 or dsTsetseEP. The blots seemingly showed that protein expression decreased in flies injected with dsGmmPro2 although not at the same time points as transcript levels; protein expression decreased from day 6 onwards whereas the transcript levels started to decline after 2 days. The protein level of GmmPro2 also seemed to decrease when flies were injected with dstsetseEP, although not as much as injection with dsGmmPro2.

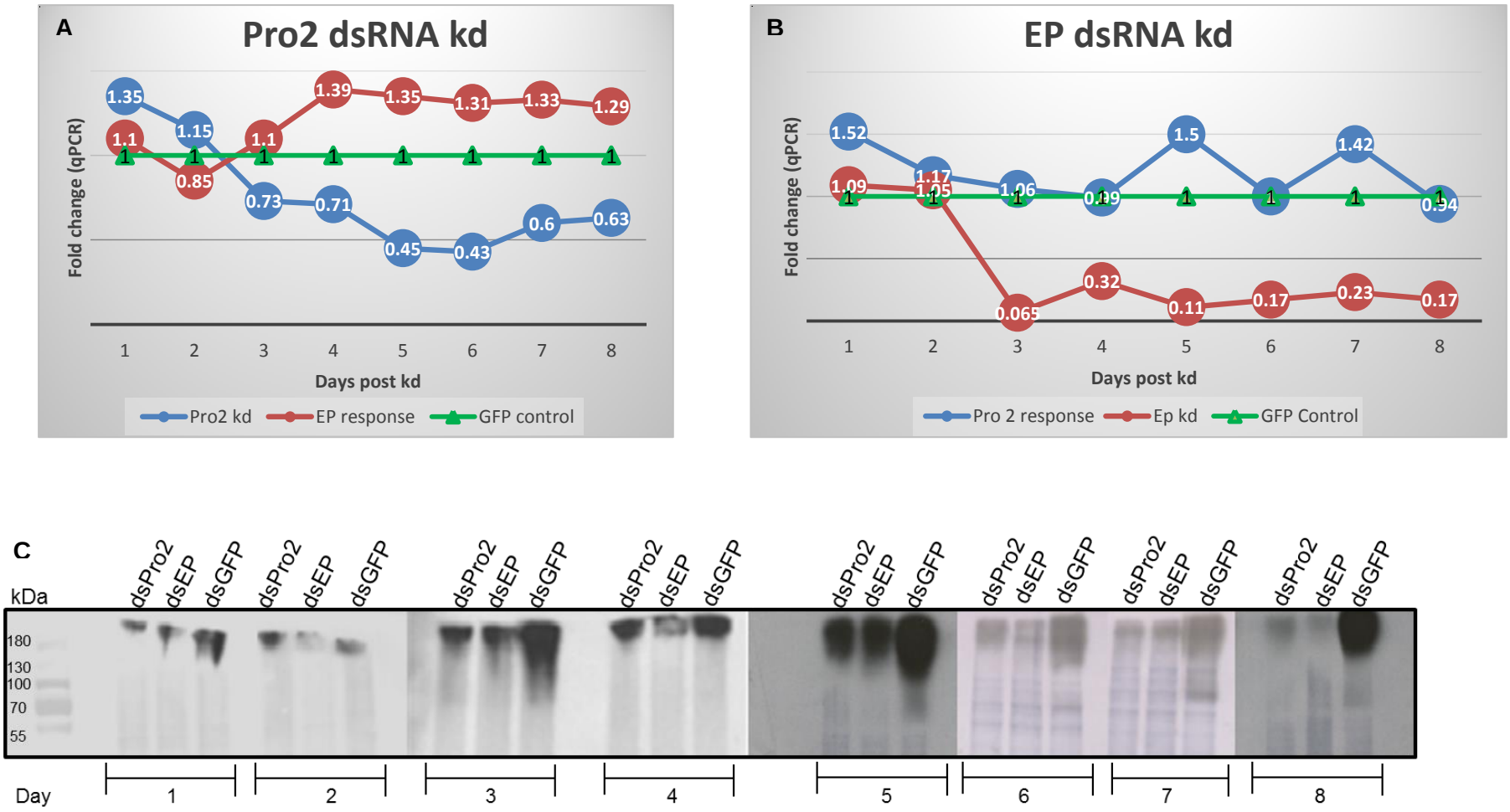


Figure 6. Immunoblot showing the protein expression of GmmPro2 in midguts from flies injected with dsGmmPro2, dstsetseEP and dsGFP over time. *Days 1 and 2 had an exposure time of 10 minutes and 3-8 had an exposure of 1 hour. Blots are superimposed over nigrosine stained membranes to show protein loading*

4.3.7 Structural analysis of the PM after knockdown of GmmPro2

In order to determine if the structural architechture of the PM changes in response to dsGmmPro2, flies were injected with either dsGmmPro2 or dsGFP and analysed under TEM (Ch4_SF2). As the transcript analysis showed days 5-6 as being the timepoint with the highest level of knockdown, TEM analysis was carried out on the midguts of flies 6 days after dsRNA injection. The PMs of flies injected with dsGFP showed the same phenotype as seen in naïve, uninfected flies (Chapter 5) and the multiple convolutions indicated that the PM was being produced in abundance and so production was normal. The PM from flies injected with dsGmmPro2 was similar in appearance to the PM of control flies, with no obvious damage observed. However, having looked at an insignificant number of flies (<5) under the same conditions, no realiable conclusions can be drawn at this point.

4.4 Discussion

4.4.1 Peritrophin expression in *Glossina*

An anti-serum containing an antibody to the major PM peritrophin of *Glossina morsitans* was developed through peptide synthesis and immunisation in rabbits. Initial analyses of tsetse tissues suggested that, at a protein level, this peritrophin was expressed not only in the PM but throughout the entire midgut and in other tissues such as the fat bodies, flight muscles and testes. This would suggest that the expression of peritrophins are not limited to the PM or tissues that synthesise the PM but may also be important components in other tissues. Although the protein has a predicted molecular weight of 66 kDa, the banding profiles are consistent with the mass spectrometry analyses of the tsetse PM (Chapter 2) in which GmmPer66 was found at differing weights between 10 and 260 kDa. Whilst the first indication that an antibody is specific for its target protein is the presence of a single band at its known or predicted molecular weight, presence of multiple bands or bands not at the proper molecular weight could represent the same target protein at different post-translational modification status, breakdown products, or splice variants. Sequence analysis of GmmPer66 shows this protein has 3 CBDs interspersed with two mucin domains as predicted by the high probability of *N*-glycosylation sites (Chapters 2 and 3). It is therefore possible that the banding pattern as seen by western blot is due to the high variability of the protein at different glycosylated states. However on subsequent blots, the banding pattern was different to that which was observed previously under the same experimental conditions. Initially it seemed as though the relative expression of this protein in both the PM and midgut changes between teneral and blood fed flies. In young, unfed flies, the expression of GmmPer66 looked to be highest in the PM but then declined following a blood meal and seemed to be expressed more in the midgut. Although possible, the obvious difference in banding patterns between the two analyses makes the specificity of the antibody to GmmPer66 seem unlikely and the results from this analysis are not conclusive. To ensure antibody specificity, a number of techniques should be adopted. For example, a blocking peptide can be used. Blocking peptides are the sequences used to generate the antibody and, during production, are incubated with the antibody in great excess. Antibodies can then be used with and without the blocking peptide to probe tissues known to produce the target protein. If specific, the antibody should not be able to bind to the tissue as the blocking peptide would have saturated the antibody binding sites and so there would be a significant reduction in binding. However, this only demonstrates that the antibody is specific for the immunogen from which it was generated; it does not prove selectivity of the antibody, only that the antibody is non-specific if there is binding observed when in the presence of the peptide.

One possibility for the lack of specificity may be in the methods used for the western blots. Tsetse tissues were ran under reducing conditions which alters the conformation of the entire protein and so antibody binding may be affected by the specific folding of the peritrophin. To see if this is the case, native-PAGE should be carried out in the first instance. Production of recombinant GmmPer66 is another possibility to ensure antibody specificity. Although costly, this method can reliably produce high protein yields which can be used to validate antibody binding through ELISA.

Given that GmmPer66 is the most abundant PM protein as determined by MS analysis, it would be worthwhile to improve the methods used here and continue towards a goal of evaluating the function of this peritrophin.

An antibody (mAb4A2) to GmmPro2 had previously been developed and screened against tsetse midguts and other tissues. Immunoblotting against male tissues revealed the presence of GmmPro2 in the crop, PV, PM and midgut. A complimentary analysis of GmmPro2 expression in female tissues was carried out which showed similar results. GmmPro2 was seen to be expressed in the tissues of the midgut and PM, but surprisingly in other tissue and haemolymph. This suggests that GmmPro2 is not only a protein limited to the PM but is secreted to other tissues of the fly. Its role is suggested to bind the ends of chitin fibrils in the PM, however as chitin is also found in other tissues such as the cuticle, GmmPro2 may play a role in modulating chitin during growth and development. The presence of GmmPro2 in the midgut is consistent with previous studies where it was found most prominently at a high molecular weight of >170 kDa. However, a longer exposure time (~ 3hours) reveals a smaller molecular weight band that corresponds to the predicted molecular weight of the amino acid sequence, which is only seen in the midgut.

4.4.2 Effect of peritrophin knockdown on protein expression

4.4.2.1 GmmPer66

To see if the protein expression of GmmPer66 changes over time in response to lower transcript abundance, gene knockdown was performed on flies over a course of 11 days. After receiving one dsRNA injection, flies were observed over the time period for signs of atypical phenotypes. However, there seemed to be no obvious deviations. There was no increase in mortality or inability to digest the bloodmeal, which is a common occurrence in tsetse who have received lethal doses of dsRNA. Similarly, the PMs of GmmPer66 showed no obvious fragility when dissecting. This indicates that the protein has no major role in the structure, or more likely given its abundance in the PM, the dose of dsRNA was too low for

any effects to be seen especially if initial transcript levels and turnover were high. There did seem to be a slight knockdown of GmmPer66 as indicated by the lack of signal in in the PM of dsGmmPer66 knockdown flies 7-10 days after dsRNA injection. However, given that the antibody to GmmPer66 seems unspecific, this is probably down to a number of issues unrelated to the effect of knockdown. For example, baseline protein levels can be variable between individual flies and the protein sources used in this experiment came from the equivalent of 2 midguts and 10 PMs. For reliability, protein levels should be measured and equal loading of protein carried out. Another indication that the antibody is unspecific is the different banding pattern seen in dsGFP to that seen in teneral and blood fed flies (Figs 4.4 and 4.5). As dsGFP is used as a control, there should be no effect observed in the protein expression and so therefore should be the same as a fly that had not received an injection. In addition, this experiment represents just one biological replicate of dsGmmPer66 knockdown and transcript levels were not validated by qPCR due to contamination of the gut by blood. As a result, no major conlclusions can be drawn. To overcome this, transcript levels of GmmPer66 should first be determined using whole flies minus the midgut in control flies over the course of 2 weeks, thus establishing a baseline level for normal transcript levels before experimental conditions are carried out.

4.4.2.2 GmmPro2

Knockdown of GmmPro2 was achieved similarly to previous studies. However, the main difference in this study was that a time line of knockdown was accomplished. The qPCR data suggests that knockdown starts to take effect as soon as 2 days after injection, with the highest knockdown achieved 5-6 days after injection. This has implications in previous studies showing infection phenotypes in GmmPro2 knockdown flies, where knockdown in teneral and non-teneral flies showed a significant increase in trypanosome prevalence in the midgut compared to control groups. However, these results should be interpreted with caution as the time of highest knockdown was not established prior to infecting with trypanosomes. One interesting result in this study is that the lower molecular weight band of GmmPro2 is lost in GmmPro2 and tsetseEP knockdown flies, yet the higher molecular weight bands remain. Whether or not this has implications in parasite establishment remains to be elucidated.

4.4.3 Effect of peritrophin knockdown on PM structure

In an attempt to characterise the role of GmmPro2 in the structural integrity of the PM, TEM analysis was carried out on midguts after dsGmmPro2 treatment. As shown in previous GmmPro2 knockdown studies, loss of GmmPro2 transcripts leads to an increase in trypanosome infection in the fly. Since the opposite was expected to happen if peritrophins

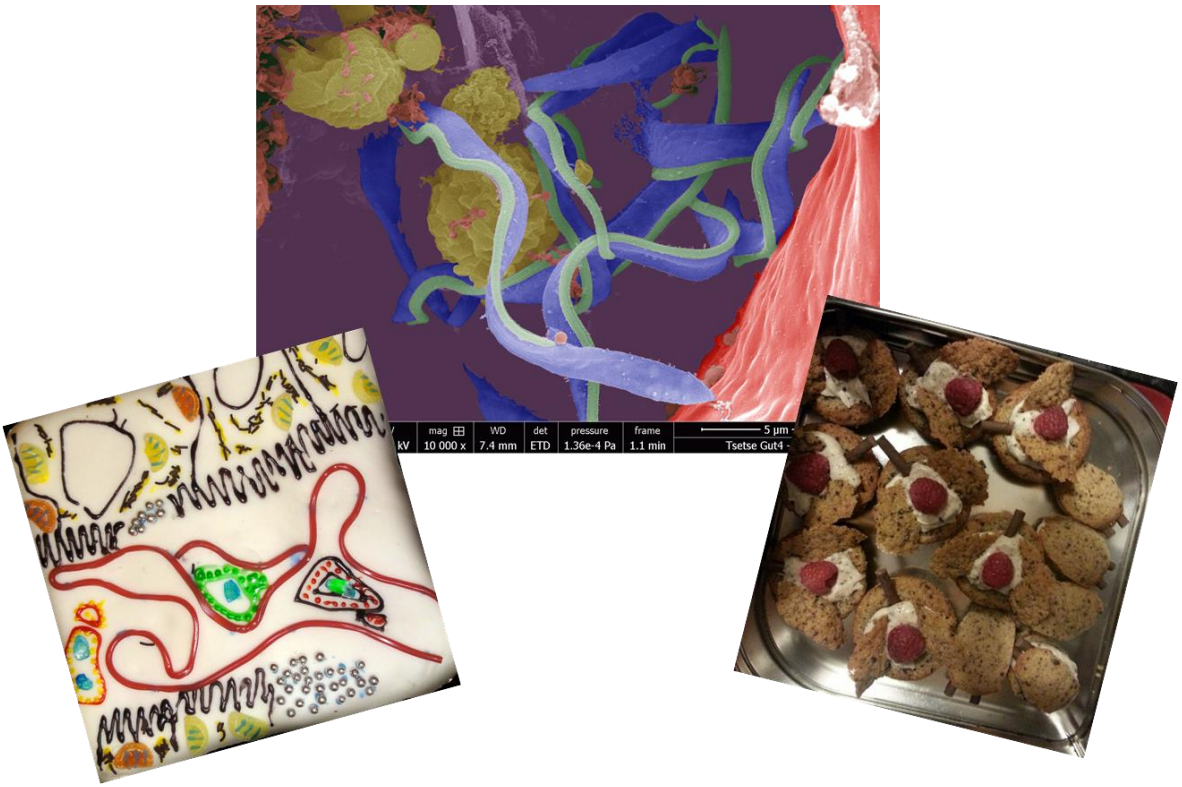
acted as receptors of trypanosomes, one suggestion for this phenotype was that knockdown led to a compromise in the PM integrity, allowing for an easier passage of parasites through the PM. It was hypothesised that by reducing the transcript levels and thus the protein expression of GmmPro2, led to a disruption in PM architecture. The current study aimed to look at the PM of a knockdown fly to see at an ultrastructural level if there are any structural changes to the PM. The results shown here suggest that even at the highest point of knockdown, there are very little structural changes, if any, to the PM as compared to a wild-type PM. Having looked at an extensive number of TEM images of naïve, refractory and infected flies, it is tempting to say that under knockdown conditions there seems to be an increase in the amount of secretions that can be seen in both the lumen and ectoperitrophic space, suggesting that the PM is being rapidly re-modelled, perhaps in response to the knockdown treatment. In addition, there seems to be an absence of the 3rd PM layer which may also reflect the effects of the knockdown, but, even in wild-type PMs, this 3rd layer is not always present (Chapter 5). However, caution must be adopted when making this statement as an insignificant amount of flies were analysed (<5 different individuals) and so this could be an artefact. Moreover, transcript levels were not verified and so GmmPro2 may not have been knocked down in those flies processed for TEM.

In order to identify the spatial distribution of GmmPro2 in the PM, immuno-TEM was attempted under different fixation methods (data not shown). However, the first attempt was unsuccessful and due to time constraints, it was not possible to repeat this experiment. The major difficulty in using this technique is having a good enough fixation of tissue that the cells are still identifiable, without damaging the epitope (i.e. GmmPro2), in which case the antibody will be unable to bind. Once this technique has been optimised, and once an effective antibody has been developed for GmmPer66, immuno-TEM should be adopted to look at the distribution of these peritrophins in the PM and PV in naïve, refractory, infected and knockdown flies. This may give an idea into the function of these peritrophins.

In summary, this study has shown the first attempts to characterise the major PM peritrophin GmmPer66 and has established a timeline of transcript abundance in knockdown GmmPro2, which will be useful for future infection studies. It may be that due to the high turnover rate of the PM in flies, knockdown will always be only partial and so the role of peritrophins will not be fully understood. In addition, if peritrophin integration into the PM occurs before knockdown effects take hold, then the knockdown may only affect further secretions into the already mature PM and other tissues.

Chapter 5. Visualising the crossing of the *Glossina morsitans morsitans* peritrophic matrix by *Trypanosoma brucei*

When science meets art…….



5.1 Introduction

The developmental cycle of African trypanosomes in their tsetse host (*Glossina* spp.) has been well documented ever since Colonel Sir David Bruce discovered the connection between tsetse and trypanosome transmission [194] [195]. In the case of *trypanosome brucei* spp., only a very small percentage (typically ~2%) of wild caught flies are found to have a mature salivary gland infection due, in part, to the number of bottlenecks to infection that the fly exerts [196] [197]. Initial establishment involves the transformation of parasites from blood stream forms into procyclic forms, that are infectious only to the fly, and this change occurs quite quickly after ingestion of an infected blood meal. Trypanosomes are acellular throughout their life cycle, both in the vertebrate and tsetse hosts. Within the fly they are in constant contact with multiple types of tissue and must undergo a series of directional migrations that are critical for survival and development. One major, outstanding question that still exists within the tsetse-trypanosome field is how these parasites are able to firstly establish a midgut infection within the fly. For this to happen, trypanosomes must escape the hostile midgut environment where they would be exposed to lectins, serum complement and components of the fly immune system, among others [198] [199]. For 100 years it has been suggested that trypanosomes must develop in the ectoperitrophic space (ES) (the region between the gut epithelial cells and the peritrophic matrix (PM)), before maturation and migration to the salivary glands. When the fly ingests an infected bloodmeal, it is immediately passed to the crop, where some parasites may be able to survive but cannot establish themselves elsewhere. Digestion rapidly starts to take place, with crop emptying occurring as fast as 20 minutes after feeding, and diuresis starting almost immediately [200]. In parallel, the ingested bloodstream trypanosomes need to quickly shed their VSG (Variable Surface Glycoprotein) coat and replace it for one made of GPEET-procyclins during transformation into the procyclic form, which is the first step in midgut establishment. The next step is to seek out a suitable environment for development and migration to the proventriculus (PV) (the site of PM synthesis). The “safe place” for this cell division and maturation is thought to occur in the ES where digestion is not an active process and where the parasites are in contact with a constant rich supply of low molecular nutrients from the gut epithelium [40]. Three possible theories for how trypanosomes escape the lumen and enter the ES exist (Fig 5.1). The first is that trypanosomes circumnavigate the open end of the PM in the hindgut and travel anteriorly through the ES towards the proventriculus and once mature, penetrate the newly secreted, softer PM (i.e. from ecto- to endoperitrophic) in the region of its formation [201] [202]. It was demonstrated that trypanosomes are capable of penetration of the fluid-like newly synthesised PM and also of the more developed form in the roof of the PV [203] [204] . Once penetration has

occurred, they then develop into proventricular forms which results in the complete molecular restructure of the parasites into one long and one short epimastigote. The short ones are able to then migrate to the salivary glands via the hypopharynx and gland ducts where they firstly attach to epithelial cells and undergo division to transform into the free swimming mammalian infectious metacyclic form. It was assumed that this journey from midgut to salivary glands by way of migration around the open ended PM was so complicated and perilous that it was the reason why so few flies were found to be infected even in endemic areas [205]. This theory was widely accepted for many years and it wasn't until after Wigglesworth described in detail the formation of the PM (known then as the peritrophic membrane) [206] that Johnson and Lloyd first mentioned the PM in relation to trypanosome development in the fly [90]. Subsequent studies showed the PM was a continuous tube all the way to the hindgut and trypanosomes could be found either side of the PM; in the lumen (endoperitrophic space) and in the ES. Buxton was the first to suggest that the PM was important in determining infection rates of tsetse by *T. brucei*. It was observed that flies fed within 48 hours of emergence were much more susceptible to infection than those flies who fed later (which was stated was due to 'some biochemical condition') and this was later demonstrated by Wijers [207]. The susceptibility of younger flies to infection was later termed the teneral (young and unfed fly) phenomenon. A second theory was that trypanosomes are able to take advantage of tears in the PM that may occur after a large bloodmeal acquisition. In younger flies, with a shorter PM it could be argued that a large meal at this age could be damaging to the PM especially when the PM is immature. Indeed, when describing the formation of the PV, Wigglesworth noted that the PM is disrupted halfway between the anterior midgut and mid midgut and suggested this may be a way of trypanosome invasion into the ES. Although, this could, in part, explain the low infection numbers in tsetse and suggest a theory for the teneral phenomenon, there is no evidence that the PM is vulnerable to breaks or tears in this way. With the second theory ultimately dismissed, research focussed again on the migration of trypanosomes from the anterior gut to the hindgut and back again, however, several further studies contradicted the theory of trypanosome development by way of circumnavigating the PM in the hindgut. The hindgut was shown to be extremely hostile with pH increasingly more acidic (from pH 7.2 to 5.8) and digestive enzymes ever more present which would further complicate the already hazardous journey. One of the PMs functions is to create a barrier and compartmentalise the bloodmeal keeping it away from the gut endothelial cells to prevent damage. An open ended PM, even for a few days, would expose the gut to not only the bloodmeal and everything ingested with it but the fly's digestive enzymes and immune components. Studies on teneral flies and flies given a small bloodmeal not long after emergence showed that the PM was a closed sac as long as it did not reach the rectal spines in the hindgut

[188]. The rate of synthesis of the PM was calculated at 1mm per hour for the first 30 hours and after feeding then more slowly thereafter as long as the fly remained unfed, reaching the rectal spines 3-4 days after emergence. This study suggested that trypanosomes penetrate the more fluid-like, newly secreted PM near to the PV where it would be in a high proportion in young flies due to the fast PM synthesis rate. In older flies with a longer and more mature PM, the softer region would only be in the immediate vicinity of the PV and so would prevent trypanosome penetration of the PM in this way.

Further evidence for PM penetration began to emerge and the theories that trypanosomes could invade the ES through a ruptured PM or via migration to the posterior open distal end of the PM were consequently dismissed. The more recent third theory is that trypanosomes directly and actively penetrate the mature PM and migrate to the ectoperitrophic space. The invasion of the ES by trypanosomes and their development within this space was studied by sections under light microscopy and it was found that within 30 minutes of an infected blood meal, trypanosomes could be found in both the lumen and the ES [208]. This was later supported by electron microscopy evidence where parasites could clearly be seen in both the ES and the lumen. Direct penetration was first proposed by Ellis and Evans in 1977 after they suggested that trypanosomes have a more developed “penetrative power” than previously thought after they viewed trypanosomes residing within the layers of the PM and they concluded that what they were witnessing was the penetration of the PM by parasites in order to cross from the lumen into the ES [89]. They also argued that trypanosomes could penetrate a more mature PM and not just the newly secreted portion as the parasites they were seeing was from the middle of the midgut rather than the anterior midgut near to the PV. In addition they showed EM evidence that certain trypanosome species such as *T. b. rhodesiense* could penetrate the midgut cells of the fly and reach the haemocoel [209]. This suggested mechanism has been the most accepted route of initial trypanosome migration in tsetse ever since. The complex lifecycle of trypanosomes, particularly in regards to early midgut establishment has had much interest over the past 100 years but it is obvious that there are still many aspects of their interaction and development in the fly that remain to be elucidated.

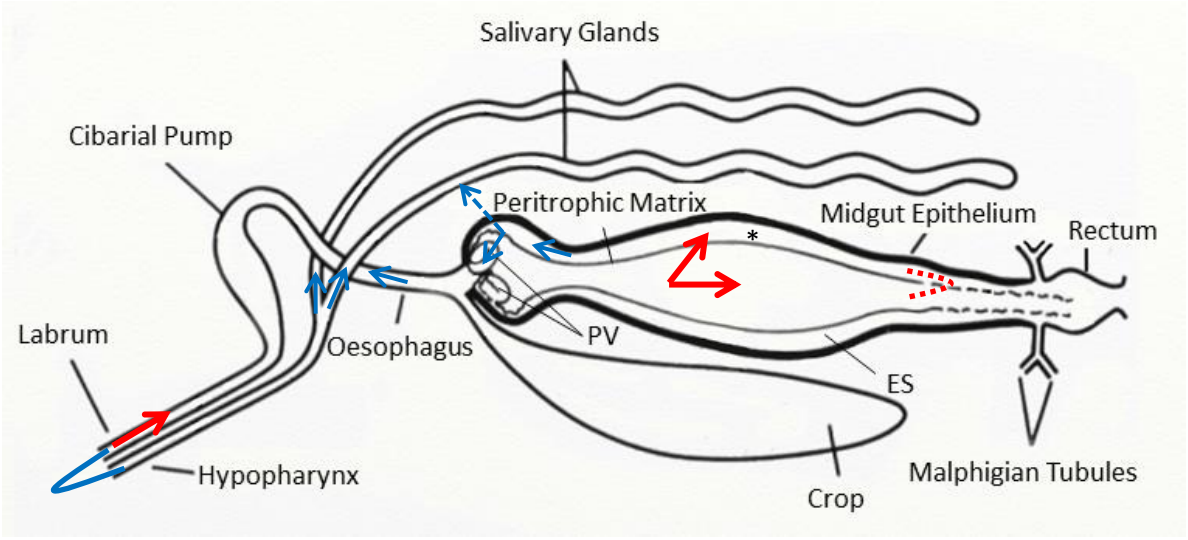


Figure 5.1. Schematic of several of the proposed routes that trypanosomes take during their migration in the tsetse to become established in the salivary glands. *Red arrows indicate trypanosomes ingested with the bloodmeal where they go into the gut. From here it was proposed that they gained access to the ectoperitrophic space by circumnavigating the ripped end of the PM (dashed red line) or by directly penetrating the PM. The asterisk indicates the ES where trypanosomes develop into PV forms. The blue arrows indicate the direction the trypanosomes move to get to the salivary glands; either through penetration of the midgut cells and via the haemolymph or back down the oesophagus and up the hypopharynx to the salivary glands (most accepted route). Adapted from http://aparasiteworld.blogspot.co.uk/2010/01/trypanosomiasis.html*

The outstanding question is, if penetration occurs, how do trypanosomes degrade the PM in order to cross through to the ES. The pore diameter of the tsetse PM are approximately 9nm, which allows, for example, the filtration of enzymes from the epithelial cells through to the lumen and water absorption in the opposite direction, but too small for trypanosomes (several microns long) to pass through. This suggests that these parasites secrete an enzyme or enzymes to digest the PM in order to permit PM crossing. A related species, *Leishmania* spp. of the Class Kinetoplastida and phylogenetically close to African trypanosomes, secrete chitinases which completely disintegrate the PM of various sandfly vectors allowing the parasites to attach to the midgut epithelial cells and establish midgut infections [46]. Likewise, the malaria causing *Plasmodium* species also secrete chitinases which quickly degrade the PM which allows a midgut infection to persist in the mosquito [210]. However, analyses of the African trypanosomes show that degradation enzymes such as chitinases are not expressed in the genome of any *T. brucei* complex species or *T. congolense*. However, there are certain proteases that are expressed only by the procyclic

trypanosomes which may act to degrade the PM in the absence of chitinases. In order to understand the process of trypanosome crossing, a closer look and a recap of the tsetse PM architecture must be carried out.

5.1.1 Tsetse Peritrophic Matrix

Tsetse are unusual within the insect taxa in that they secrete a type 2 PM throughout the whole of their life cycle. The type 2 PM is secreted by a group of specialised cells in the PV and is continually secreted as an unbroken concentric sleeve or sleeves which extends along the entire length of the gut and is always present. In contrast, type 1 PMs are produced by the midgut epithelial cells in response to an ingested meal and is usually degraded after digestion has occurred (or by pathogens present within the meal). Many insect vectors produce a type 2 PM at the larval stage of development and subsequently switch to produce a type 1 PM upon reaching the adult stage. These type 1 PM producing adults are usually more permissive vectors of disease. This can be explained partly by the way in which the type 1 PM is produced and also because the type 2 PM is a much more highly complex and organised structure. Indeed, it is not uncommon that type 2 PMs can present up to 5 concentric sleeves in certain species [33]. Electron microscopy studies show the tsetse PM to be comprised of 3 layers (PM1, PM2 and PM3), each differing in thickness and composition (Fig. 5.2) [78]. Biochemical studies show that on the whole, the PM is comprised of chitin, glycoproteins and glycosaminoglycans (GAGs). Whilst all 3 layers are predicted to contain N-acetyl-*D*-glucosamine (GlcNAc) residues based on lectin recognition), they are mostly abundant in PM1 and PM2. PM3 has been shown to have similar properties to that of mammalian kidney glomerular basement membrane and may be important in the filtration properties of the tsetse PM [61].

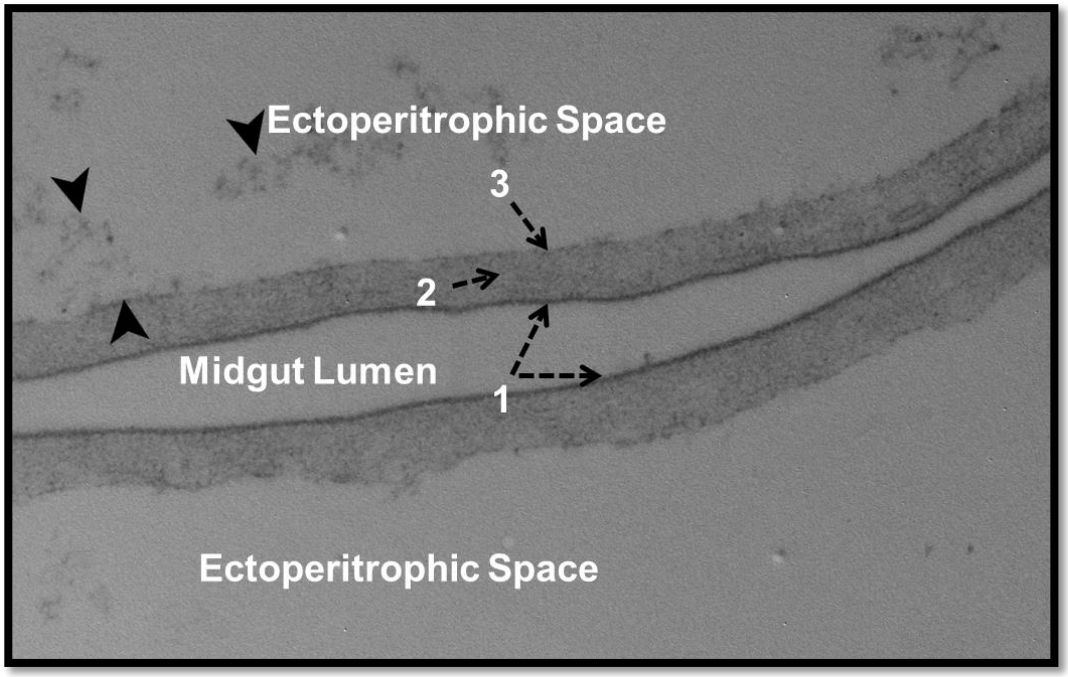


Figure 5.2. An electron micrograph clearly showing the 3 layers that comprise the tsetse PM. *(1). The first layer, an electron dense layer facing the lumen of the gut that mainly consists of mucopolysaccharides (2). The second layer, the thickest of all three layers and (3). The third and outermost layer, this is slightly electron dense and faces the ectoperitrophic space. The formation of the 3rd layer can be seen as material presumably originating from the epithelial cells is deposited (arrowheads).*

To understand the biology of how trypanosomes interact with and potentially traverse the tsetse PM, a number of microscopy techniques were adopted to visualise this phenomenon. The migration of GFP expressing trypanosomes has been reported in regards to the time course of trypanosome migration through the tsetse fly [185]; however the main PM crossing event has never been described. The aim of this study was to re-examine the route in which trypanosomes migrate within their tsetse hosts, focussing on the developmental cycle in the gut of the fly. In contrary to what has previously been reported, there was no direct evidence that trypanosomes actively cross a mature PM. Instead, it is suggested that the PM acts as a “sticky trap” that envelops parasites which are subsequently excreted when the PM reaches the rectal spines in the hindgut and that the entire lifecycle prior to salivary gland migration occurs in the proventriculus.

5.2 Materials and methods

5.2.1 Tsetse flies

Flies were taken from an established colony of *Glossina morsitans morsitans* (Westwood) at the Liverpool School of Tropical Medicine that were maintained on sterile, defibrinated horse blood (TCS Biosciences) at an ambient temperature of 27°C±2°C and a relative humidity of 65-75%. Experimental flies were collected at <24 hours post eclosion (P.E) and offered a bloodmeal every 2 days before being starved for 72 hours in preparation for dissection at 5, 8 or 11 days. All flies used in these experiments were males.

5.2.2 Trypanosome strains

Two salivarian *Trypansoma* (*Trypanazoon*) *brucei brucei* strains were used in this study. TSW196 blood stream form (BSF) trypanosomes were used for the TEM, SEM fluorescent and confocal microscopy experiments. J10 (*Trypanosoma brucei brucei* MCRO/ZM/73/J10) Green Fluorescent Protein (GFP) expressing BSF were used for the light sheet and confocal microscopy experiments. Flies for infection experiments were collected at <24 hours post-eclosion and were given their first blood- or serum-meal containing approximately 4 x 10^6 parasites. Any flies that did not feed were removed the following day and conditions before sacrifice are the same as described above.

5.2.3 Transmission Electron Microscopy

5.2.3.1 Embedding

Midguts were dissected in ice-cold fixative (0.2M cacodylate, 4% paraformaldehyde (PFA), 2.5% glutaraldehyde (GA), 3% sucrose, pH 7.4), transferred to 3cm^3 petri dishes containing fixative, sealed with parafilm and incubated on ice for an hour. The tissues were then washed twice in ice-cold 0.1M cacodylate buffer containing 3% sucrose (pH 7.4) for 2 minutes and left in1% osmium tetroxide for an hour at room temperature. The osmium was aspirated off and a copious amount of ice-cold 0.1M cacodylate buffer was added three times for 30 minutes then washed twice in distilled water for 30 minutes. Midguts were placed in 30% ethanol for 30 minutes before 0.5% uranyl acetate in 30% ethanol was added to cover the tissue. Midguts were then subject to a series of 10 minute ethanol washes increasing from 30-80% and left for 30 minutes in 100% ethanol. Graded hard embedding resin 182 (TAAB) was mixed in a 1:1 ratio with 100% ethanol and left on midguts overnight to ensure infiltration of tissue. This was replaced with fresh 100% resin for 30 minutes and placed in an oven at 60°C for 48 hours to cure.

5.2.3.2 Cutting and Sectioning

Once hardened, the midgut containing resin was removed from the dish using liquid nitrogen to help separate the two, and rough cut with a razor and hacksaw to approximately 18mm length by 8mm width, leaving the midgut to run orthogonally through the resin. As the region of interest is just behind the proventriculus (PV) in the thoracic midgut, this was the start of sectioning. Glass knives were prepared using a knife cutter which were mounted in a Leica EM UC6 microtome to rough cut an inner square of 150x150 µm around the midgut and an outer square of 500x500 µm. A diatome 45° diamond knife was used to slice 8-10 sections at a 70nm depth (giving alternating gold/silver appearances) which were then collected on freshly prepared Pioloform®-coated 200 (for midguts and PVs) or 100 (for PVs) mesh nickel grids. Sections were left to dry for at least an hour before post-staining.

5.2.3.3 Post-staining of grids

Grids were placed sample side down in drops of uranyl acetate (5% w/v in 30% ethanol) for a minute before being washed in distilled water 3 times for 20 seconds. They were then placed in 50% lead citrate and washed for 20 seconds in distilled water three times and left to dry.

5.2.3.4 Grids

0.3% Pioloform® (polyvinyl butyral) diluted in chloroform was used to coat clean glass slides. The film was separated from the glass slide, floated in distilled water and grids placed on the film. A clean slide was used to collect the grids and left to dry for 1-5 days in a sealed container before use.

5.2.4 Scanning Electron Microscope

5.2.4.1 Preparation and staining

Isolated chitinase treated PMs or infected/naïve midguts were dissected in ice cold Phosphate Buffered Saline (PBS, 140 mM NaCl, 1 mM KCl, 6 mM phosphate buffer, pH 7.4). The proventriculus (PV) from midguts was isolated using a fine razor blade and either laterally or sagittally sliced. They were then transferred to poly-L-lysine coated coverslips and fixative added for one hour on ice before being rinsed twice in ice-cold 0.1M cacodylate buffer containing 3% sucrose (pH 7.4) for 2 minutes and left in 1% osmium tetroxide for an hour at room temperature . They were washed 3 times for 5 minutes each and dehydrated in an ascending series of ethanol (30-80%) and then twice in 100% for 10 minutes. Samples were then placed in a critical-point dryer before being transferred to a pin mount and sputter-coated with a thin layer of gold/palladium and viewed in an FEI Quanta 250 SEM. Gatan digital micrograph software was used for alignments and conversions to TIFFs.

5.2.5 3View®

5.2.5.1 Staining

Midguts were prepared and stained for SEM and 3View® reconstruction using a method based on the protocol of Deerinck et al 2010 [211]. Briefly, midguts were dissected in ice-cold fixative (0.1M cacodylate, 2% paraformaldehyde (PFA), 2% glutaraldehyde, 3% sucrose, 2mM calcium chloride pH 7.4) and washed 5 x 3 minutes in 0.1M cacodylate buffer pH 7.4 with 2mM calcium chloride prior to staining with reduced osmium tetroxide (2%) containing 1.5% potassium ferrocyanide in 0.1M cacodylate buffer. Midguts were washed 5 x 3 minutes in distilled water and incubated in the mordant thiocarbohydrazide (TCH) for 30 minutes before a further 5 x 3 minutes in distilled water. A second osmium (2%) staining was carried out at room temperature for 40 minutes, midguts washed 5 x 3 minutes in distilled water before incubation in 1% uranyl acetate overnight at 4°C. A final wash of 5 x 3 minutes in distilled water was carried out and samples were stained in warmed lead aspartate for 30 minutes before dehydration in graded ethanol 30, 50, 70, 90% for 5 minutes each then 2 x 100% ethanol for 5 minutes. For the infiltration step, graded hard resin 812 (TAAB) at a 1:1 ratio with 100% ethanol was applied to each midgut and left overnight. The resin was removed and followed by infiltration with resin:ethanol (2:1) for 1 hour, (3:1) for 1 hour and 2 x 30 mins of 100% resin. Finally midguts were embedded in fresh, 100% resin and left to cure for at least 48 hrs before rough cutting and sectioning as described previously. Due to the heavy staining used, there was no need to post-stain the samples following this stain. All samples prepped for TEM were viewed at 100 KV in an FEI Tecnai G2 Spirit and all micrographs were taken using either an Olympus megaview 3 or a Gatan orios camera with megaview or gatan digital micrograph software respectively.

5.2.5.2 Increased staining for better contrast

To help with increased contrast and better infiltration for 3D reconstruction, the above protocol for 3View® staining was slightly modified. Midguts were dissected in ice-cold, modified fixative (0.1M cacodylate, 2% PFA, 2% GA, 3% sucrose, 0.1% tannic acid pH 7.4) for 30 minutes on a rocker. The part of interest (PV or anterior gut) was cut away with a fine razor blade and transferred to fresh ice-cold modified fix and left for 90 minutes on ice, then washed 3 times for 3 minutes in 0.1M cacodylate buffer pH 7.4 with 2mM calcium chloride at room temperature. For reduced osmium staining, tissues were transferred to bijou bottles prior to staining with reduced osmium tetroxide (2%) containing 1.5% potassium ferrocyanide in 0.1M cacodylate buffer. Midguts were washed 5 x 3 minutes in distilled water and incubated in TCH for 20 minutes before a further 5 x 3 minute washes in distilled

water. A second osmium (2%) staining was carried out at room temperature for 30 minutes, midguts washed 5 x 3 minutes in distilled water before incubation in 1% uranyl acetate overnight at 4°C. A final wash of 5 x 3 minutes in distilled water was carried out and samples were stained in warmed lead aspartate for 30 minutes before dehydration in graded ethanol 30, 50, 70, 90% for 10 minutes each then 4 x 100% ethanol for 10 minutes and 2 x 10 minutes in 100% propylene oxide. A number of infiltration steps were carried out in graded hard resin 812 (TAAB) and 100% propylene oxide; 1:3 for 4 hours, 1:1 overnight, 2:1 for 1 hour, 3:1 for one hour, 4:1 for one hour and finally 3x1 hour of 100% resin. Finally midguts were embedded in fresh, 100% resin and left to cure for at least 48 hrs and then prepared for 3D reconstruction by gluing a small square of embedded sample onto a pin mount.

5.2.5.3 3D reconstruction

Samples that were mounted on pin heads were placed in a microtome inside the SEM. For the midgut reconstructions, 100nm thick sections were taken and each one was scanned up to a total of 474 slices of which the first 200 were taken for reconstruction. For PV reconstructions, the thickness was reduced to 40nm thick sections and 3 regions of interest were scanned, all of which were 458 slices. The reconstruction was carried out on the first 400 slices. Gatan digital micrograph software was used for alignments and conversions to TIFFs and the reconstructions were carried out using Bitplane Imaris version 8.1.

5.2.6 Chitinase assay

Two different bacterial chitinases with different enzymatic properties were used in feeding assays and to incubate isolated PMs. An exochitinase from *Streptomyces griseus* was used at a concentration of 0.05874 U/100µl and a combined endo- and exochitinase from *Trichoderma viride* was used at a concentration of 0.0158 U/100µl. Flies of 24-48 hours old were fed each chitinase in a serum meal and dissected 72 hours later or teneral flies were dissected and their isolated PMs were incubated overnight with each of the chitinases. PMs were then either stained with WGA-rhodamine (see below), and viewed under fluorescence or were prepared for SEM (see above).

5.2.7 Wheat Germ Agglutinin stains

Isolated peritrophic matrixes from naïve and infected flies at either 5, 8 or 11 days post eclosion/post infection were dissected in PBS and transferred to poly-L-lysine covered slides. PMs were then either incubated in rhodamine labelled Wheat Germ Agglutinin (WGA) or fluorescein labelled succinylated WGA at a 20 µg/ml concentration for 5 minutes,

washed once in PBS and covered with a coverslip sealed with clear nail polish. PMs were viewed immediately under a fluorescent or confocal microscope.

Rhodamine labelled WGA was also fed to flies at a final concentration of 20 µg/ml in SDM-79 serum. Flies that did not take the meal were removed the following day and the remainder were dissected on day 5, 8 or 11 days post eclosion/infection. Whole guts were dissected and mounted under a bridge slide in PBS and viewed by fluorescent microscopy.

5.2.8 Fluorescent Microscopy

Isolated PMs or whole guts were viewed under 20x, 40x and oil immersion 100x objectives giving an overview or close-up respectively using a Nikon Labophot fluorescent microscope equipped with a Nikon Coolpix 995 (Nikon Corp, Tokyo, Japan). All data acquisition was processed with NIS-Elements v4,3 software and edited in Photoshop.

5.2.9 Confocal Microscopy

Isolated PMs were dissected and transferred to poly-L-lysine covered slides. For fixed mounts, PMs were fixed in 4% paraformaldehyde, washed in PBS and incubated for 10 minutes in WGA. DAPI (4', 6-diamidino-2-phenylindole) was added for 2 minutes to flies that were infected after washing off the WGA with PBS. After a final wash in PBS, most remaining liquid was aspirated off and then the PMs were mounted with mowial 4-88® containing 2.5% DABCO (1, 4-diazabicyclo-[2, 2, 2]-octane). For live mounts, isolated PMs were processed as above, minus the fix. After all incubations and washes, a drop of SDM-79 medium was added to each PM, a coverslip applied and sealed around the edges with clear nail varnish. Images of GFP constructs, WGA and DAPI were taken using either a Nikon A1R confocal microscope at room temperature with a 60× 1.4 NA objective or on a Zeiss LSM 510 confocal microscope at room temperature with a 100x oil immersion with identical laser settings (GFP: excitation 488 nm, emission 500–550 nm, WGA: excitation 561.4 nm, emission 570–620 nm, DAPI: excitation 405, emission 410 -460). Pinholes were set between 2-70µm and data capture and extraction was carried out with Slidebook digital microscopy software v6 or Zeiss LSM510 v5.0 operating system respectively.

5.2.10 Light Sheet-based Fluorescent Microscopy

5.2.10.1 Live imaging of GFP J10 trypanosomes inside the tsetse gut

Male flies that were <24 hrs old were fed an infectious SDM-79 media meal containing approximately 4 x 10^6 parasites J10 (*Trypanosoma brucei brucei* MCRO/ZM/73/J10), supplemented with rhodamine labelled WGA at a final concentration of 20 µg/ml and DRAQ5™ at a final concentration of 10 µg/ml. Intact midguts containing the PV and crop were dissected after 48 hours and placed in a drop of 1% low-melting agar (42°C). A

plunger was placed into a 10-50 µl glass capillary and used to suck up the midgut, posterior end first. After 10 minutes, the sample was placed into the Zeiss Z.1 light sheet and guided until the PV could be seen. The gut was then extruded anteriorly into the sample chamber containing media deficient of phenol red and cell tracking was initiated for image acquisition. Images were processed with Zen 2 core software

.

5.2.10.2 Imaging isolated infected PMs

PMs were dissected from 11 day infected flies and given a WGA-rhodamine/DRAQ5™ supplemented media meal 72 hours prior to sacrifice. They were then fixed in 100% methanol and processed for LSFM as above.

5.3 Results

The assumption that African trypanosomes are able to actively penetrate the mature PM during midgut establishment is one that has been established for 40 years and only a handful of studies subsequently have re-visited the idea. Tracking GFP expressing trypanosomes over many days within the tsetse midgut has shown that by day 5, flies either exhibit an established infection and trypanosomes can be seen in the ES from day 6 onwards or they have cleared the infection completely. Although this study shows GFP trypanosomes lining up at the edge of the PM with the assumption they are sensing or binding to the PM, the actual penetrating or crossing event has not been captured. With this assumption in mind and the advancement of higher quality microscopy techniques in recent years, this present study aims to clarify the events surrounding PM crossing and initial midgut establishment of *T. brucei* in its vector *Glossina morsitans morsitans*. The time-points of 5, 8 and 11 days post infection were chosen to coincide with the well-established, known timeframe of trypanosome development in the midgut.

5.3.1 Wheat Germ Agglutinin as a tool to stain the tsetse PM

Because the PM is transparent and hard to visualise, it is sometimes difficult to see the way in which trypanosomes are interacting with it; even those parasites that express GFP. To overcome this, the PM was stained with Wheat germ agglutinin (WGA). WGA is a lectin that preferentially binds to N-acetyl-D-glucosamine (GlcNAc) and sialic acid residues. GlcNAc is the monomeric unit of chitin which is a major component of bacterial cells walls, insect cuticle and insect PMs. The succinylated WGA is different from the native form in that it doesn't bind sialic acids. This is important to distinguish the binding affinity to GlcNAc rather than sialic acids that are present on red blood cells for example. This ensures that what the lectin is binding to is actual components of the PM and nothing else. In addition, incubating WGA and succinylated WGA with chitin hydrolysate prior to incubation of PMs prevents WGA from binding to the PM (Appendix: Fig. A1). At a higher dilution of 1:4 of chitin hydrolysate, neither the WGA nor succinylated WGA showed binding, whilst at a more dilute concentration of 1:40, WGA was still able to bind, albeit at a reduced affinity, whereas succinylated WGA failed to bind the PM at both concentrations. However, it is important to note here that although the tsetse PM is believed to contain chitin, the total proportion or distribution throughout the PM is unknown and the WGA could be binding to other PM components, such as glycosaminoglycans (GAGs), which contain GlcNAc residues. Nevertheless, what the WGA is binding to is not important here, it is simply a useful tool for labelling the PM for better visualisation of trypanosome-PM interactions. Feeding tsetse

with WGA and also incubating the isolated PM is sufficient to stain the entire length of this tissue (Fig. 5.3). Feeding WGA is important for viewing live trypanosomes within a whole gut as it doesn't bind or label any other tissues and isolating the PM is important for visualising parasites that are bound or interacting in some way to the PM. After establishing that WGA stains PM components, the succinylated WGA was no longer used for further experiments as the fluorophore it emits is in the same emission spectra as the strain of GFP expressing parasites used. The native WGA lectin coupled to rhodamine emits at a different spectral range than the GFP parasites so can easily be distinguished. It was also determined that fixing tissue before staining and vice versa didn't affect the staining of PM tissue.

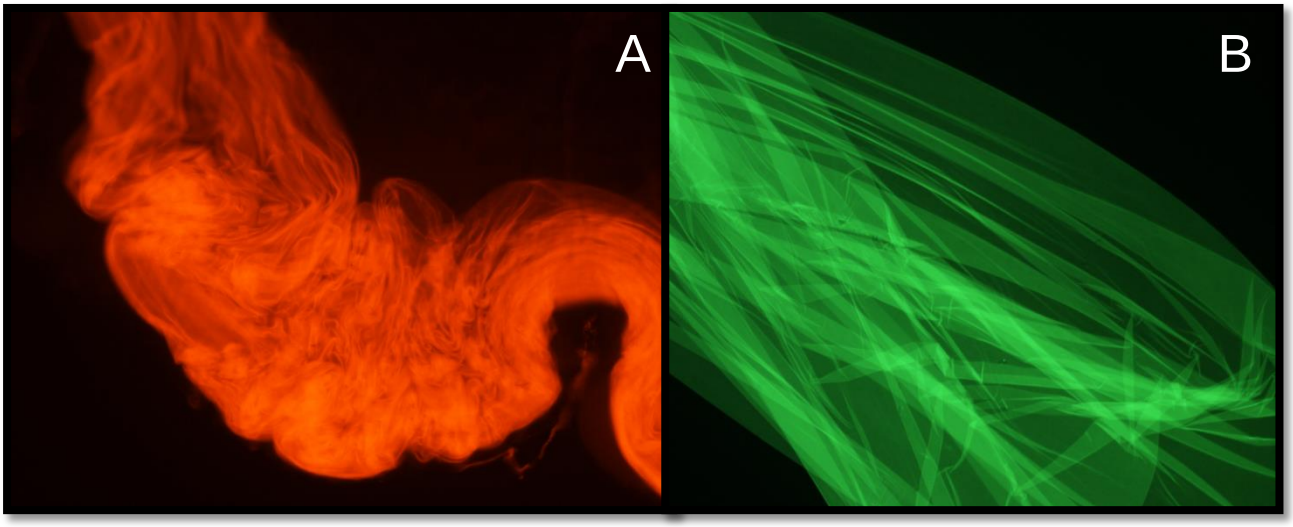


Figure 5.3. WGA staining of* G. morsitans *PM *(A) shows WGA rhodamine staining of the PM from a whole gut mount from a fly that was fed WGA. (B) shows succinylated WGA fluorescein staining of an isolated tsetse PM.*

5.3.2 Chitinase Assay

To see if chitin is a major component of the tsetse PM, a chitinase assay was carried out on isolated PMs which were then analysed for any damage that may be attributed to the activity of the enzymes on chitin present in the PM. The chitin component of the tsetse PM may be involved in regulating trypanosome establishment in the tsetse midgut. Although African trypanosomes do not express chitinases, tsetse flies can harbour up to 3 different symbiotic bacteria at a time, of which, at least one is known to secrete chitinase. The secondary symbiont, *Sodalis* glossinidius, has previously been shown to confer tsetse susceptibility to trypanosome infection although the exact mechanism is unknown. It is

however speculated that they secrete chitinases in order to release GlcNAc residues for use as their energy source, and as they have a close association to the tsetse PM, it is thought that these residues are PM derived. The GlcNAc residues accumulate in the larvae and upon emergence, inhibit trypanocidal lectins and thus allow establishment of trypanosomes in younger flies [212]. As the fly ages and the population of *Sodalis* increases, the store of residues declines and lectins are able to efficiently kill any trypanosomes ingested. Another suggested role of *Sodalis* in potentiating infection in the fly is alteration of the tsetse PM architecture through digestion of the chitin within this structure. In order to determine if chitinases excreted by *Sodalis* have a role in regulating the susceptibility of tsetse to trypanosome infection by way of PM alteration, exogenous chitinase from two different bacteria was used in both feeding assays and to incubate isolated PMs. As *Sodalis* has multiple chitinases, including exo- and endo-chitinases, the chitinases chosen for the assay was based on their ability to exhibit activity on multiple sites within the chitin molecule (Fig. 5.4). Chitinase activity from isolated exochitinase from *Streptomyces griseus* involves a 2-step process. Firstly, chitodextrinase-chitinase removes chitobiose (*N*-acetyl D-glucosamine disaccharides) units from the chitin polymer starting at the non-reducing end of the molecule, and then glucosaminidase-chitobiase cleaves these dimers into monosaccharides. The activity of chitinase from *c*involves the same 2-step process as that exhibited by *S. griseus*, but also includes endo-chitinase activity, whereby the enzyme can either randomly cleave internal points in the chitin molecule, or will cleave first into chitobiose units and then liberate into monosaccharides.

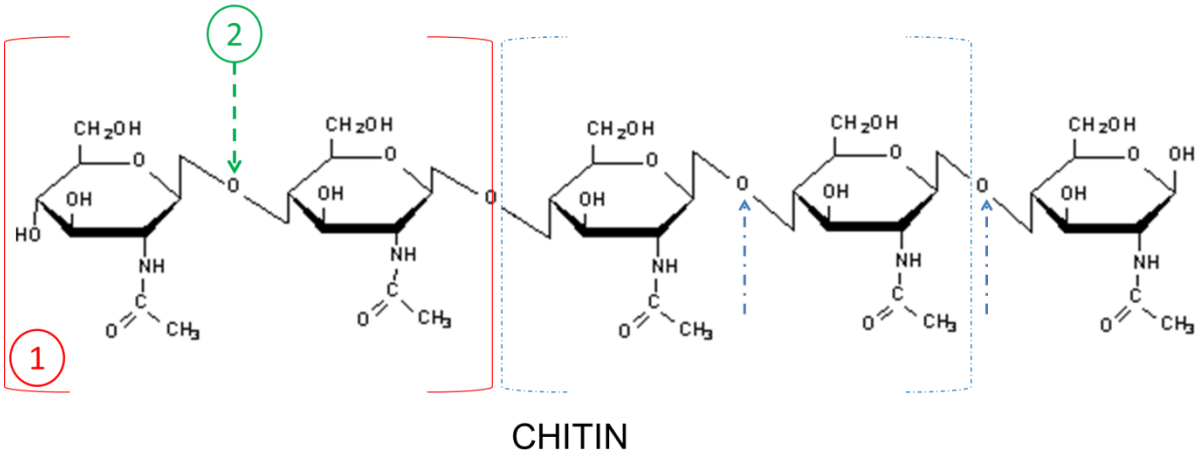


Figure 5.4. A schematic showing the chitinase activity from 2 different bacteria on chitin. S. griseus *displays exo-chitinase activity whereby chitin is initially cleaved into disaccharide units of N-acetyl D-glucosamine (chitobiose; 1, red brackets) and then into monosachharides; 2, green arrow.* T. viride *exhibits both exo and endo-chitinase activity,*

the latter acting on random internal points in the chitin molecule (blue arrows), or can also cleave chitin into chitobiose units (blue brackets).

As WGA is able to stain the PM, it was assumed that any damage caused by chitinase activity would show up under fluorescence as a reduction in or lack of staining would indicate that the PM is damaged in some way. Results show that whilst the chitinase from *S. griseus* appear to have no effect on the PM integrity after both feeding and incubation, the enzyme from *T. viride* appears to have a slight effect on the architecture under both experimental conditions (Fig. 5.5). However, this effect was only mild and it did not have a detrimental effect on the digestion of the bloodmeal nor the longevity of the fly. It would seem that the synergistic action of an exo- and an endo-chitinase is more effective at disrupting the tsetse PM than either chitinase alone and it is interesting to note that *Sodalis* express both types of chitinase.

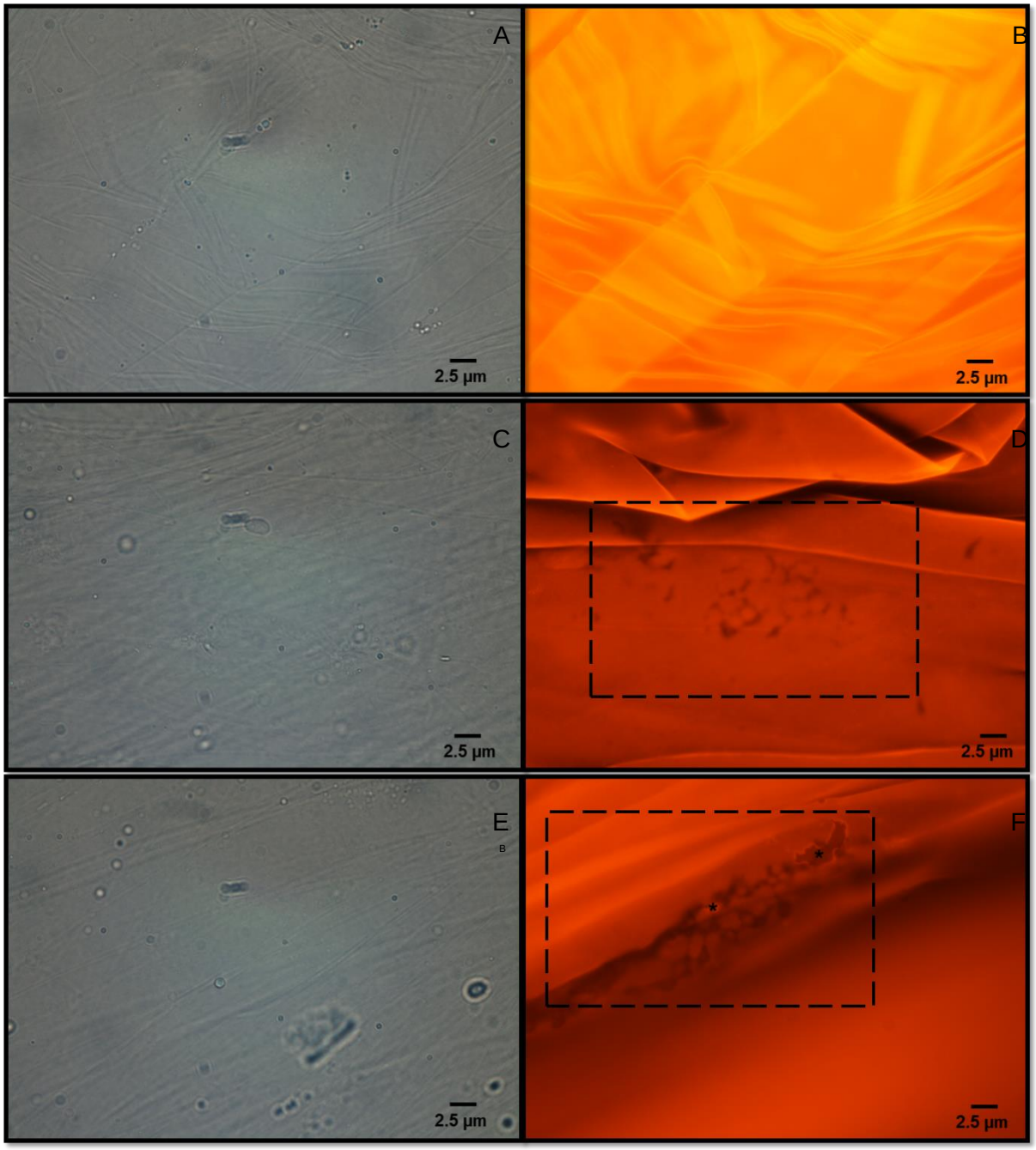


Figure 5.5. PMs from flies after either feeding chitinases or incubating isolated PMs with chitinases. *(A) Brightfield image of a PM from a fly that was fed chitinase from S. griseus. (B) Fluorescent image of the same area of PM in A, showing uniform staining of WGA across the PM. (C) Brightfield image of a PM from a fly that was fed chitinase from T. viride. (D) Fluorescent image of the same area of PM as C. A reduction in the staining in the WGA (dashed black box) suggests that there has been damage caused to the PM from the chitinase action. (E) Brightfield image of an isolated PM that was incubated overnight in chitinase from* T. viride. *(F) Fluorescent image from the same area of PM as E. A similar pattern to that in the PM shown in D, where a lack of staining (dashed black box) indicates chitinase damage. (*) indicates dissection damage. 100 x oil*

5.3.3 SEM of isolated PMs after chitinase treatment

Although the results from the chitinase assays showed a reduction in staining in rhodamine when viewed under fluorescent microscopy, it was difficult to determine whether this was a genuine effect as a result of chitinase action on the PM components due to the sporadic nature of the reduction in staining. In addition, any effect from the action of the chitinase secreted by *S. griseus* may have been too small to be seen under fluorescent microscopy. To overcome this, scanning electron microscopy (SEM) was carried out on PMs isolated from flies and subsequently incubated with chitinases overnight. As suggested from the chitinase assays which indicated that feeding chitinase or incubating isolated PMs with chitinase gave similar results, it was decided that the SEM experiment should go ahead after incubation of PMs rather than after feeding. Dissection after feeding needs 3 days after the meal to ensure full digestion has occurred whereas incubation can be done overnight. The results of the SEM analysis showed similar results to the chitinase assays (Fig. 5.6.) where there seems to be slight and sporadic damage in the PMs incubated in chitinase from *T. viride.* Unfortunately, the PMs incubated with chitinase from *S. griseus* had come off the coverslips and the samples were lost so were unable to be viewed. However, the collective results from all chitinase experiments suggest that either chitinases have very little effect on the tsetse PM possibly due to the chitin being inaccessible to the enzyme, or that the tsetse PM contains very little chitin in comparison to other vectors and insects in general.

Fig. 5.6a-b. SEM micrographs of PMs that were dissected from two teneral, 3 day old flies. There doesn't appear to be any damage to the PMs. Both 50,000 x

Fig. 5.6c-d. SEM micrographs of PMs from two different flies that were fed 0.580 U/ml of chitinase from* T. viride *and dissected after 3 days. Slight damage can be seen on parts of the PM (dashed box), but not all over (*). Sometimes the damage is greater and clear holes can be seen (arrowheads). Both 50,000 x

Fig. 5.6e-f. SEM micrographs of PMs from 3 day old flies that were isolated and incubated in 0.580 U/ml of chitinase from* T.viride. *Shows similar to results to the feeding assays where slight and sporadic damage can be seen in some parts of the PM (dashed boxes) but again, not all over (*). E; 50,000 x, F; 10,000 x

Fig. 5.6 g-h. SEM micrographs of PMs from 3 day old flies that were isolated and incubated with 0.580 U/ml of chitinase from* T. viride. *Parts of the PM appear to be similar to the PMs from control flies where there appears to be no damage. G; 50,000 x, H; 25,000 x

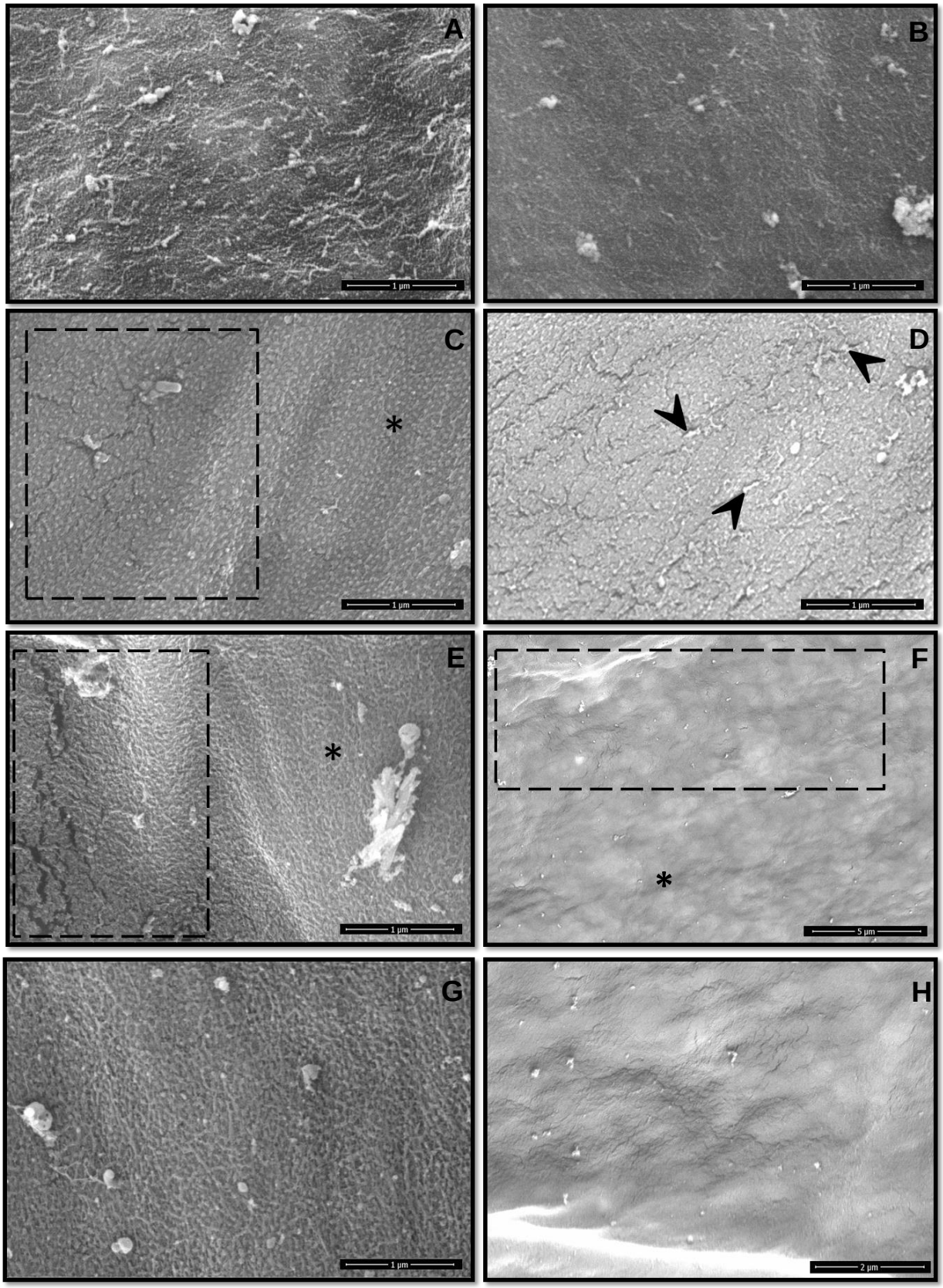


Figure 5.6. A selection of scanning electron microscopy images that show the effect of a chitinase on tsetse PMs compared to control PMs. *Images taken after incubating isolated PMs with a chitinase from* T. viride. *Details of each image are shown in the text above.*

5.3.4 Fluorescence Microscopy of PMs from naïve and infected flies

Having determined that chitinase activity causes relatively little or no disruption of tsetse PM architecture, the next step was to see if trypanosomes themselves caused any degradation to the PM. Incubation of isolated PMs from naïve flies and also from refractory flies (i.e. those flies that were given a bloodmeal but subsequently cleared the infection) showed a uniform staining of WGA along the entire length of the PM. There were however, parts of the PM that displayed variability in signal brightness that could be seen in PMs from naïve, refractory and infected flies that were usually tract-like in appearance running parallel and almost always occurred towards one end of the PM (Appendix: Fig. A2). No obvious damage to the PM was observed under brightfield. It is likely that these tracks are more frequent towards the posterior end where the PM is older and possibly more degraded or the increase in brightness of the signal could be due to accumulation of midgut components being laid down on the PM as it matures.

Other irregularities that were observed in PMs from naïve, refractory and infected flies was obvious mechanical damage from dissections (Appendix: Fig. A3). The edges of the rips or tears had an increase in signal, possibly due to more GlcNAc residues being exposed.

Taking into account all the inconsistencies and obvious damage from dissections, PMs from infected flies at 5, 8 and 11 days post infection (d.p.i.) showed clear holes where trypanosomes could be seen inside. Most holes appeared bigger than the parasite itself, suggesting they are able to move inside the PM and create space around them (Fig. 5.7.).

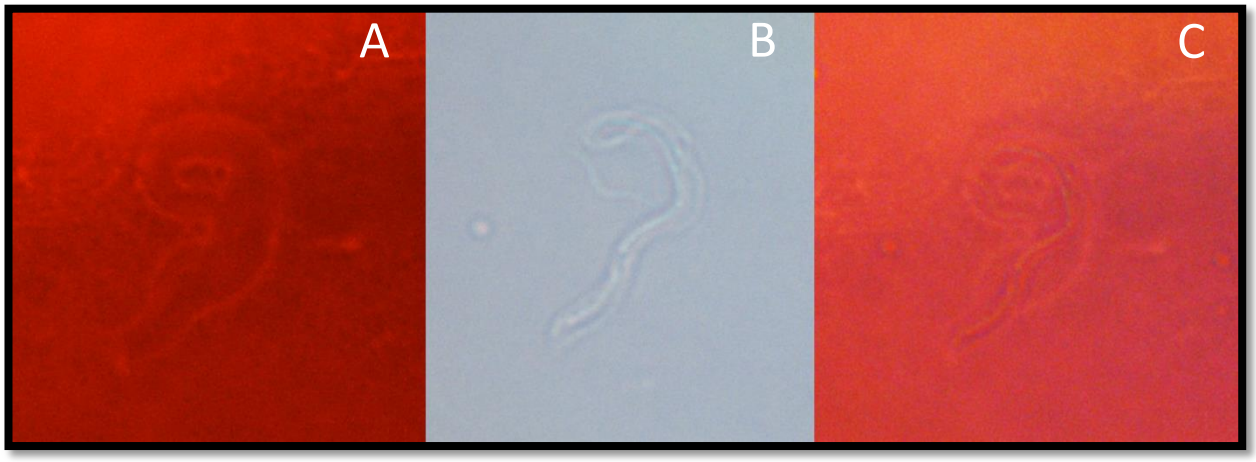


Figure. 5.7. A montage of a PM from a TSW196 infected 8 d.p.i. fly. *(A) Fluorescent image of the PM showing a depression in the rhodamine staining in the profile of what appears to be a parasite. (B) A Brightfield image of a trypanosome. (C) An overlay of A and B showing the depression in the rhodamine staining is larger than the parasite. 100x*

All holes examined from all ages of infected flies contained a trypanosome or multiple parasites and no holes were observed that appeared to be empty. Viewing isolated PMs in live mounts showed trypanosomes that presented in a number of ways; trypanosomes were either motile and not stuck in any way to or in the PM, or parasites could be seen with their posterior ends stuck in the PM and their flagella (anterior end) flapping about, or the parasites appeared immotile inside holes and were assumed dead, or parasites were observed moving inside holes. In correlation with previous studies, at 4 d.p.i. large numbers of parasites were observed in flies infected with TSW 196 and GFP expressing J10 parasites, or flies showed no sign of infection suggesting the crucial time point of midgut establishment is within the first 96 hours of a fly taking an infected bloodmeal.

5.3.5 Confocal Laser Scanning Microscopy

In order to determine if what is seen under normal fluorescence microscopy are trypanosomes within the PM layers, crossing the PM or simply adhered to the surface, confocal laser scanning microscopy (CLSM) was carried out on isolated PMs. This technique gives an advantage for obtaining high resolution optical images with depth selectivity by optical sectioning. To better visualise the trypanosomes interacting with the rhodamine stained PM, GFP expressing J10 *T. b. brucei* were used. Using CLSM, it was clear to see the highly folded nature of the PM when adhered to glass slides. This made it somewhat difficult to differentiate those trypanosomes that were lying inside folds of the PM from those that were truly between the PM layers. Nevertheless, using orthogonal projections which show three, mutually perpendicular planes; XY, XZ and YZ, it was clear to see those trypanosomes inside the PM layers as the rhodamine signal could clearly be seen above and below the parasite.
CLSM confirmed with greater clarity what was seen under normal fluorescent microscopy, and the two strains of parasite used showed similar characteristics. Video evidence showed those parasites that were stuck to the PM at their posterior end whilst the anterior end was free and the flagella could be seen highly animated (Appendix: Fig. A4.).

This phenomenon could be seen at all time points of infected flies observed. However, this posteriorly orientated adherence was more common in flies at 4 and 5 d.p.i. Those flies dissected and examined at 8 and 11 d.p.i. more commonly presented trypanosomes that were clearly inside holes within the layers of the PM. The parasites were seen to be moving inside these holes when viewed in live assays, or could be seen immotile and presumed dead. In addition, trypanosomes were mainly long and thin (similar to those forms most

notably found in the proventriculus), but they could also be found in rounder holes curled up (Fig. 5.8.).

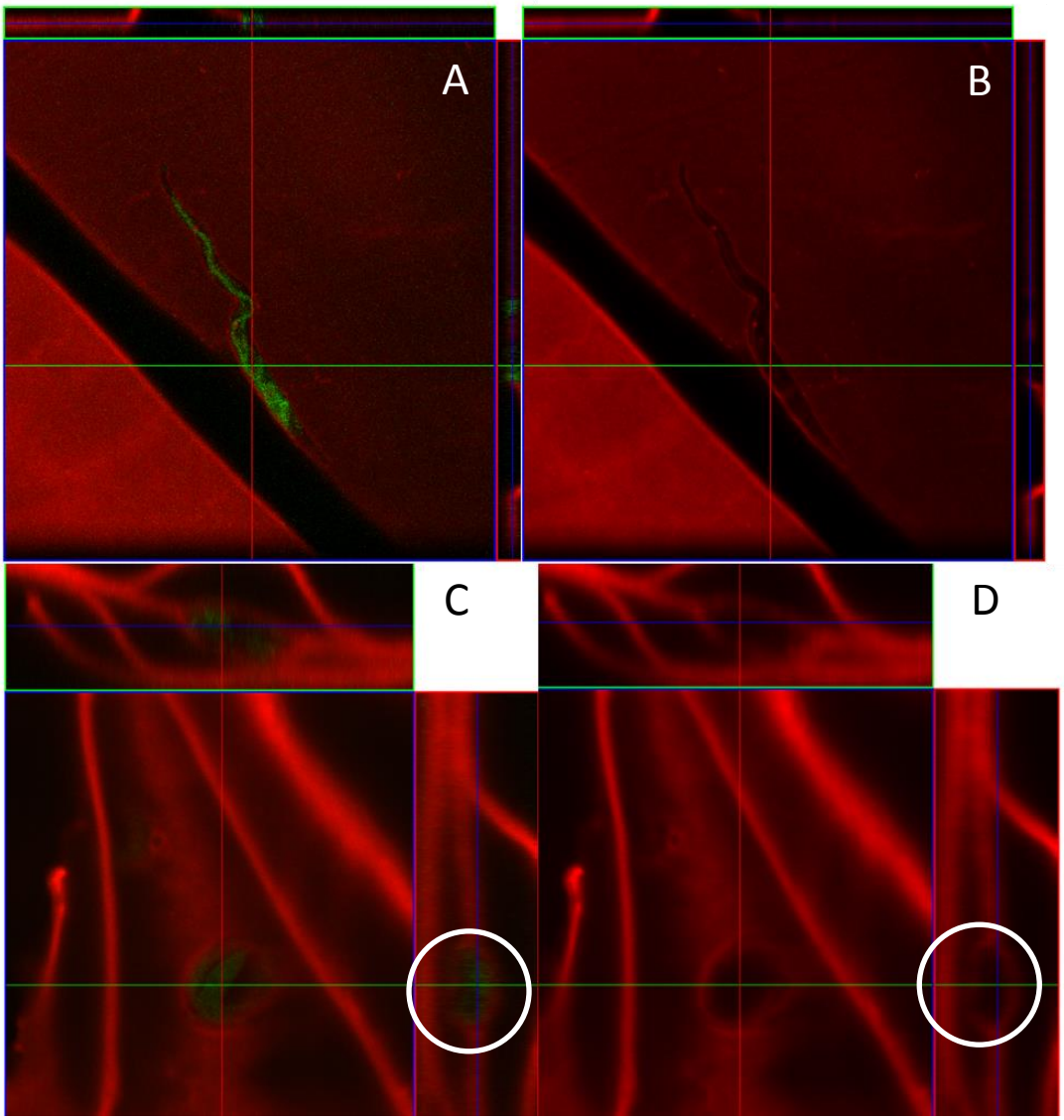


Figure. 5.8. Images of rhodamine stained PMs taken under LSCM. *(A) Shows a long, thin GFP expressing parasite. (B) Shows the clear hole in the PM when the laser for GFP is switched off. (C) Shows a rounder hole with a curled up trypanosome inside. (D) When the laser for the GFP is switched off, the hole in the PM is clear. It is also evident from the XZ and YZ planes (white circles) that the rhodamine signal is only emitting above and below the parasite. 100x*

Holes were found throughout the entire length of the PM, usually containing a single parasite as shown in figure 5.8. However, occasionally multiple parasites could be seen in a single hole (Fig. 5.9.). Holes were most noticeable where the PM had adhered flat to the coverslip and there were less folds to be seen.

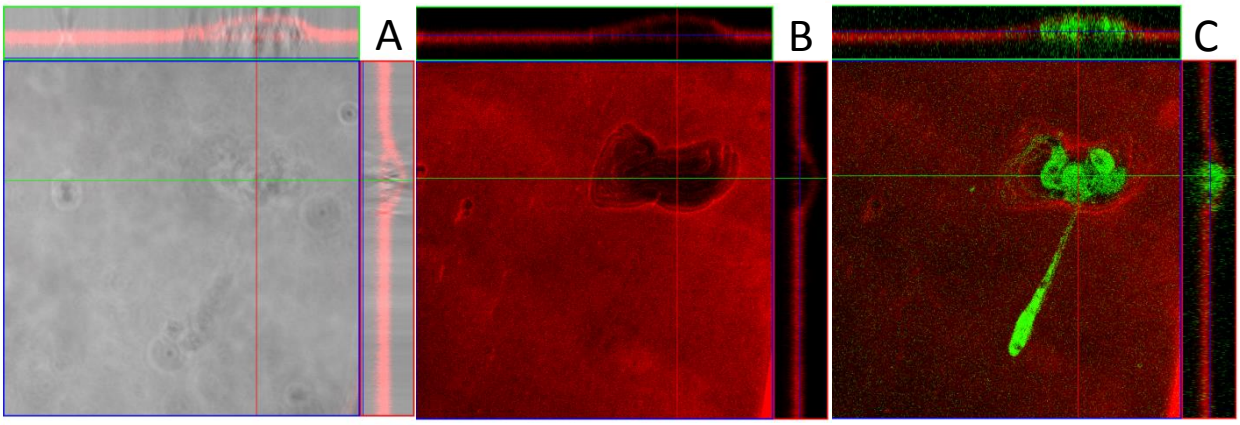


Figure 5.9. Images of an isolated PM from an 8 d.p.i. fly *(A) Looking at the XZ and XY planes shows a clear hole in between the layers of the PM when the GFP signal is turned off. The parasites are visible under brightfield. (B) When the GFP and brightfield signals are turned off, the hole in the PM is clearly visible from the reduction in rhodamine emission. (C) When the GFP signal is turned on, the bundle of GFP expressing parasites can be seen within the hole. There is a single trypanosome projecting from the hole that doesn't seem to be within the layers of the PM. 100x*

A large proportion of trypanosomes were morphologically similar to proventricular forms. It may be argued that these forms were simply adhered to the surface of the PM after rupture of the PV, however they still existed even after extensive washing of the PM and they could also be found within the layers of the PM. The use of DAPI (4',6-diamidino-2-phenylindole) improved visualisation of morphology through the staining of the nuclei and kinetoplasts of the cells. It was evident that there were trypanosomes at varying stages of their lifecycle (Fig. 5. 10.).

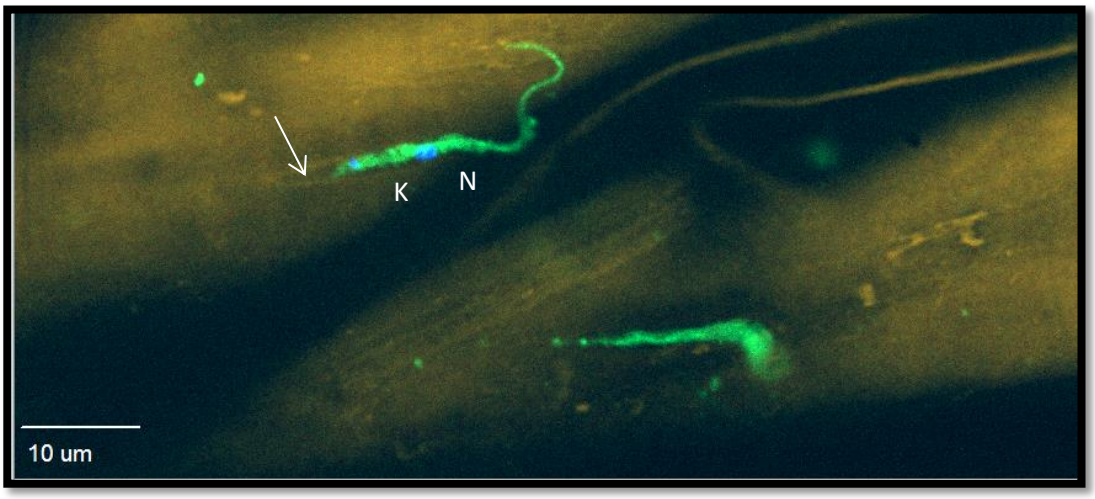


Figure 5.10. Image taken with a spinning disk confocal microscope of an isolated PM from an 11 d.p.i. fly. *Shows a trypanosome more morphologically similar to a*

proventricular form as shown by the distance between the nucleus (N) and the kinetoplast (K). The kinetoplast is at the posterior end of the cell. The arrow indicates a hole in the PM which the parasite is lying in. 63 x oil

Through 3D reconstruction of images taken with the spinning disk confocal, it was easier to see parasites with different morphologies lying within the PM layers (Fig. 5.11.). As trypanosomes transform from procyclics into proventricular forms they become elongated along the anterior end whilst the posterior end becomes increasingly short and fat. On average, they increase in length from 16 µm to up to 30 µm before continuing to grow to 40 µm once they are fully mature PV forms.

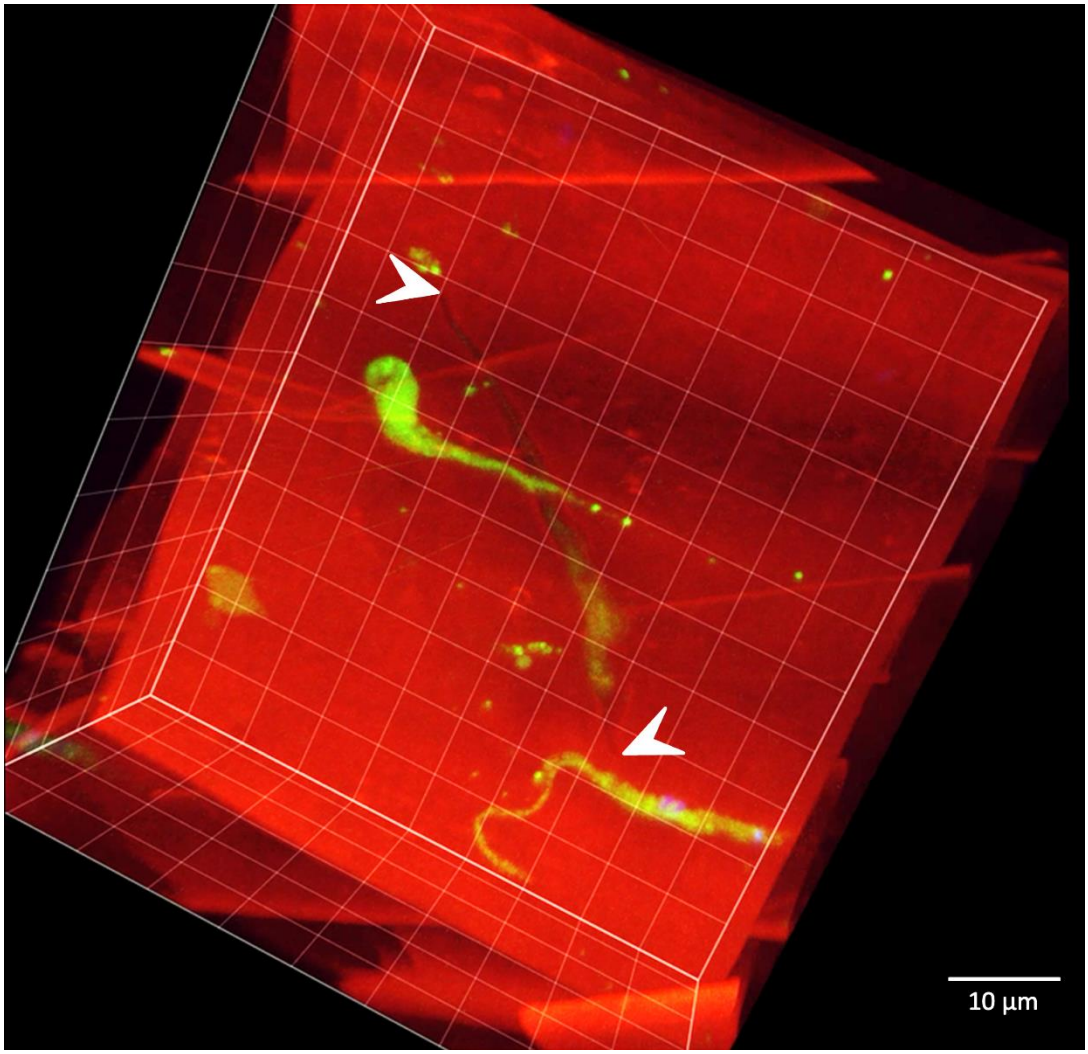


Figure 5.11. Screen capture of a video from a 3D reconstruction of isolated PMs from an 11 d.p.i. fly. *A GFP expressing trypanosome can be seen clearly within the layers of the PM lying in a hole (white arrowheads). The length of the parasite indicates that it is a PV form (~35 µm). 63 x oil*

All the collective evidence from both fluorescence and confocal microscopy indicate that whilst trypanosomes are visible both on the surface of the PM and residing between the PM layers, there is no suggestion at this level that trypanosomes are penetrating the PM. Whilst the lack of rhodamine signal above and below parasites in some PMs indicate that trypanosomes are between the PM layers, multiple analyses of PMs of flies at different time points have shown no evidence that the parasites transversely penetrate as a lack of rhodamine signal all the way through the PM where parasites were located could not be seen. In addition, when there are parasites that are adhered to the PM rather than in between the layers, there is no reduction in rhodamine signal and no holes are present. All holes seen contained one or multiple parasites and there were no holes that appeared to be empty. In order to see the PM trypanosome interactions at an ultrastructural level and to complement the fluorescence and confocal microscopy evidence, a combination of transmission electron microscopy (TEM) and scanning EM (SEM) were used.

5.3.6 Transmission Electron Microscopy of tsetse midguts and hindguts

Ultrastructural analysis of the tsetse PM has previously been carried out and has shown to consist of three layers, each differing in thickness and composition as previously discussed. To ensure reproducibility, TEM of naïve flies was carried out in order to determine the correct region of interest. The initial region of interest was parallel to Ellis and Evans analysis; the midgut anterior to the bacteriome. This was confirmed by the presence of specific types of epithelial cells that can be found there (Fig. 5.12). As tsetse are heamatophagous, they ingest large quantities of blood at intermittent intervals and so the anterior part of the midgut is specifically designed to cope with any sudden changes due to the distention and pressure exerted by the ingested bloodmeal. In addition, this part of the tsetse midgut is responsible for the rapid transfer of water in the bloodmeal from the midgut lumen to the haemolymph and so have ultrastructural specialisations associated with this. Cell-cell junctions in this part of the gut are numerous as they are responsible for retaining epithelial cell integrity when distorted by the large bloodmeal and by the unfolding of the basement membrane; termed basal infoldings. Once the region of interest was determined through identification of these specialised cells, TEM was carried out in this area of all infected and naïve flies at all time points.

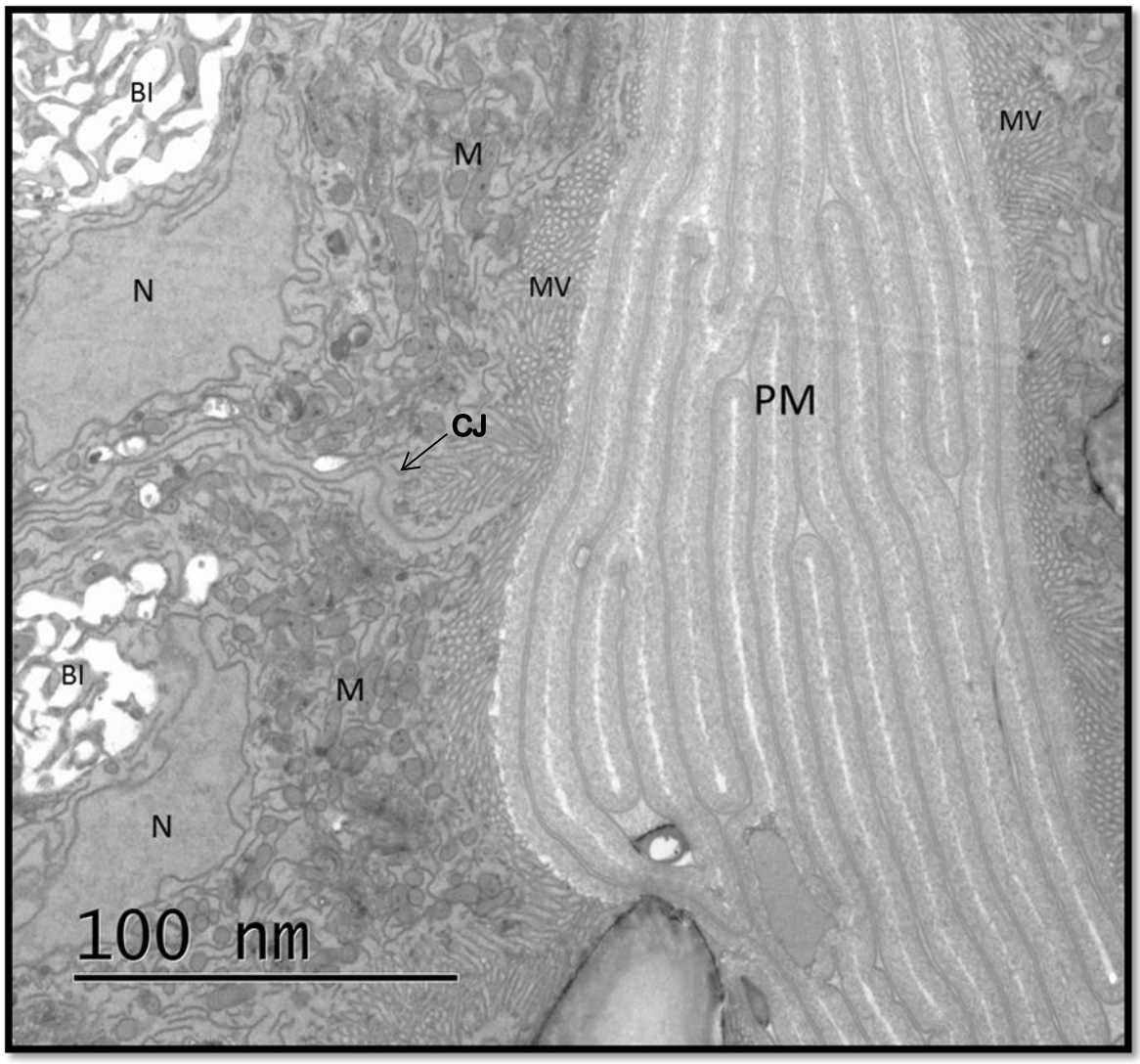


Figure 5.12. A section of a midgut from a 5 day old naïve fly. *The epithelial cells are typical of those found in the region of the midgut anterior to the bacteriome. BI: Basal infoldings, CJ: Cell junction, M: Mitochondria rich region, MV: Microvilli, N: Nucleus, PM, Peritrophic matrix. 43000 x*

The PMs of naïve and refractory flies were consistent in their thickness and appearance. Very rarely was any damage seen and on the two occasions that inconsistencies were observed, the PM was completely transversely broken which was attributed to mechanical damage (Fig. A5). This damage may have occurred at any stage in the processing but was most likely caused from initial midgut dissections. Thus, it was with confidence that variations of the PM in infected flies were believed to have been caused either by the trypanosomes directly or by the fly in response to an infection or, alternatively, as a consequence of the physiological adaptations of the fly during infection.

Trypanosomes can be identified under TEM by their distinctive microtubule formation of their cytoskeleton and flagella (Fig. 5.13.). The cytoskeleton of *T. brucei* largely consists of microtubules which lie directly under the plasma membrane in the form of a subpellicular corset and are important in growth and cytokinesis. The flagella of trypanosomes consist of a typical core axoneme arranged in a 2+9 configuration, which lies adjacent to the paraflagellar rod.

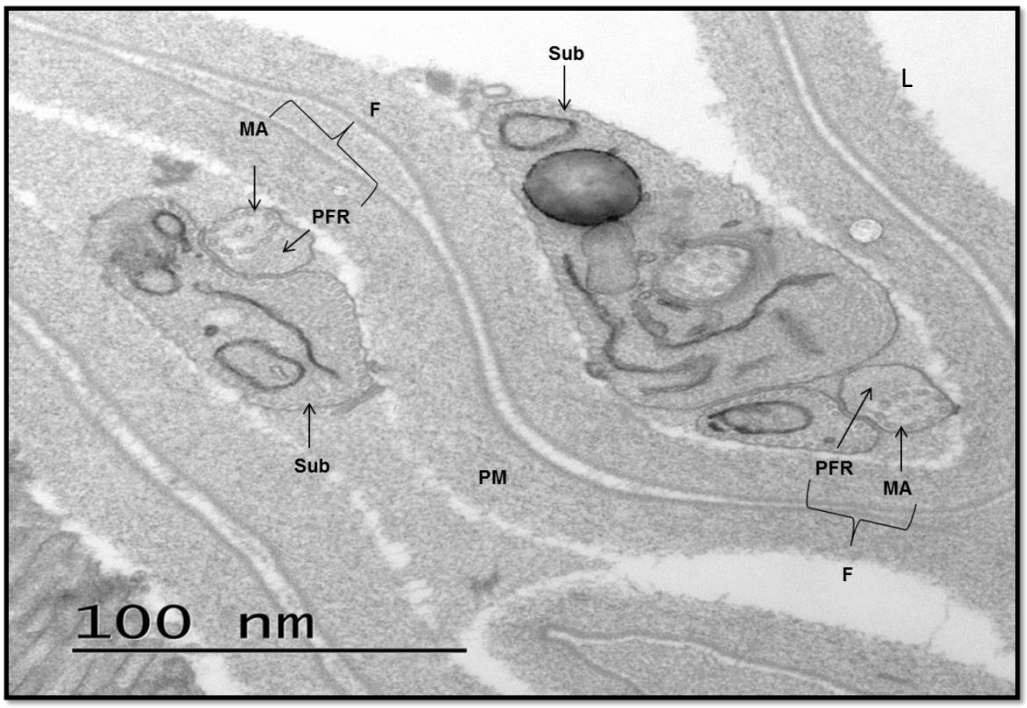


Figure 5.13. Typical structure of a wild type* T. brucei *as seen in a fly gut under TEM. *Trypanosomes can be seen here in the ectoperitrophic space. Cross sections of the microtubules in the main cell body can be seen in the subpellicular corset (Sub). The flagellum (F) is made up of a microtubular axoneme (MA) and a paraflagellar rod (PFR). PM: peritrophic matrix. Taken from a fly at 11 d.p.i. 65,000 x*

In infected flies at 5, 8 and 11 d.p.i. trypanosomes could be seen in both the lumen (=endoperitrophic space) and the ectoperitrophic space (ES). At this level it is impossible to tell whether or not the trypanosomes are healthy and can replicate or whether they are moribund, or indeed at what stage of the lifecycle they are in. However, these findings corroborate those that were previously seen in that at an early stage of infection trypanosomes can be found still contained inside the lumen or that they are presumed to have crossed the PM at this stage as they can now also be found in the ES. In most PMs from infected flies, damage is a common recurrence and the more typical type of damage tends to be separation which occurs as an ‘unzipping’ of layers between PM1 from PM2 (Fig. 5.14a.). Usually one or more trypanosomes can be found within this separation (or objects that look similar to a parasite and may be a lysed trypanosome (Fig. 5.14b.)) and

only very rarely can damage be seen that looks unassociated with a parasite (Fig. 5.14c.). Sometimes the parasites can be seen tightly contained between PM1 and PM2 (Fig. 5.14d.), but in a lot of instances the parasites are between PM1 and PM2 surrounded by excess space between the two layers (Fig. 5.14e.). Occasionally, variations of the PM damage can be seen where the parasites seem to be splitting PM2 down the middle rather than ‘unzipping’ PM1 from PM2 (Fig. 5.14f), and only on two occasions (albeit from the same fly) during analyses has a complete degradation to PM1 been seen (Figs. 5.14g and 5.14h.). Multiple parasites can sometimes be found between the PM layers, from a few cell bodies (as seen in Fig. 5. 14d.), to huge numbers that form a cyst-like structure Fig. 5.14i.). These cysts are more commonly found in older infected flies but can occasionally be found in flies at 5 d.p.i.

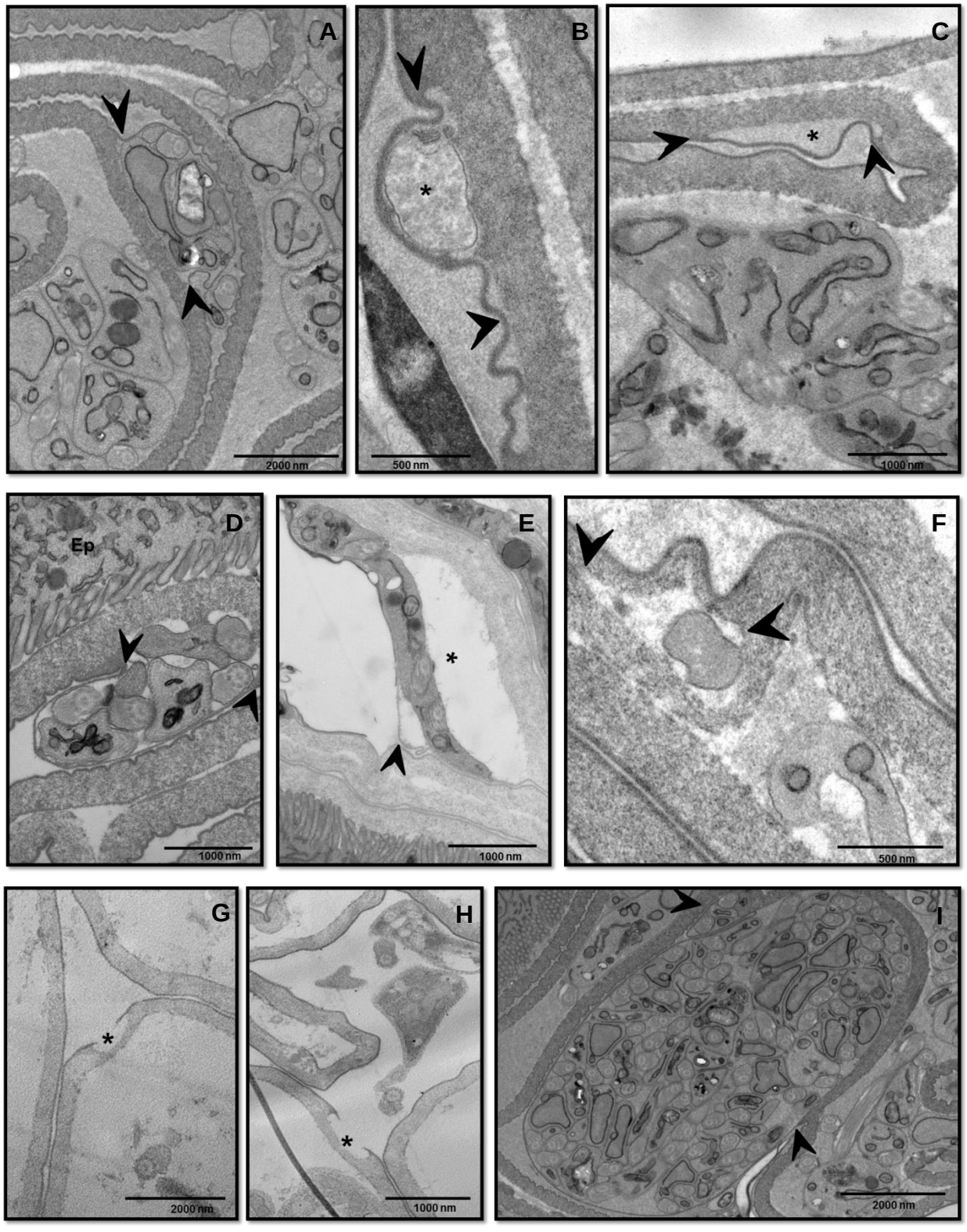


***Figure 5.14. A selection of electron micrographs showing the types of damage seen to the PM when infected with TSW196* T. brucei.** *Images were taken from flies that were 5, 8 or 11 d.p.i. Details of each image are shown in the text below.*

Fig. 5.14a. The typical damage found to PMs from infected flies is the separation of PM1 from PM2. Usually the electron dense first layer doesn't break, and can be seen peeling away from the second layer (arrowheads). Image taken from a fly at 11 d.p.i. 16500 x

Fig. 5.14b. Characteristic damage of the first and second layers of the PM separating from each other (arrowheads), but an object inside that is atypical of what is normally seen (*). This is probably a lysed trypanosome. Image taken from a fly at 8 d.p.i. 43000 x

Fig. 5.14c. The space between PM1 and PM2 after separation (arrowheads) is usually filled and seldom are there empty spaces like the one seen here (*). Image taken from a fly at 8 d.p.i. 26500 x

Fig. 5.14d. Usually the trypanosome(s) are tightly engulfed between PM1 and PM2 (arrowheads). Image shows 3-4 trypanosomes within the two layers. Ep: Epithelial cells. Taken from a fly at 11 d.p.i. 26500 x

Fig. 5.14e. Occasionally, trypanosomes are not tightly packed between PM1 and PM2. The layers have separated by quite a distance (arrowhead on PM1) and there is a lot of empty space within the hole (*). Taken from a fly at 5 d.p.i. 26500 x

Fig. 5.14f. Sometimes the damage to the PM is not the separation of PM1 from PM2, but the damage is to the second layer (arrowheads). This image shows a single flagellum splitting PM2. Taken from a fly at 8 d.p.i. 43000 x

Fig. 5.14g-h. Only twice has it been shown that a complete degradation of PM1 (*) has occurred. These images were taken from the same fly at 11 d.p.i. g: 16500 x h: 26500 x

Fig. 5.14i. In older flies at 11 d.p.i, and less commonly in flies at 5 and 8 d.p.i., multiple parasites can sometimes be seen between PM1 and PM2 (arrowheads showing PM1 at the point of separation) forming huge cyst-like structures like the one seen in this image. Note that the electron dense layer remains unbroken. Taken from a fly at 11 d.p.i. 16500 x

5.3.7 Serial Block-face Scanning Electron Microscopy and 3D reconstruction

In the most common damage where PM1 separates from PM2, the first electron dense layer remains intact and is never broken (apart from the two occurrences shown) even when there are huge numbers of parasites present between PM1 and PM2. These data validate was has previously been shown in that trypanosomes can be perceived to be crossing a more mature PM. However, as TEM only gives a detailed 2-Dimensional picture, it is impossible to say if the trypanosomes are actively crossing the PM transversely. To better see this event, serial block-face scanning electron microscopy (SBSEM) was used on samples that could give a good 3-Dimensional reconstruction. SBSEM involves the use of an ultramicrotome mounted inside the chamber of an SEM. The SEM images the surface of the block-face of the sample and after imaging the microtome is used to take a section. This process is automated and can generate thousands of images in perfect alignment so long as the sample remains undamaged by the beam of the SEM.

Whilst most schematics depicting the PM within the context of the tsetse gut represent a straight tube running parallel to the gut, the PM is in fact a highly convoluted structure. The mature PM is usually twice as long as the gut itself and the hyper-folds are more pronounced in flies that have completely digested their bloodmeal. For this reason, flies were chosen for 3View® analysis depending on the results from the TEM. Those guts that had a spaced out PM were disregarded for SBSEM as it would be likely to incur more beam damage and thus give poor resolution and decrease the chances of getting a good reconstruction. Similarly, those guts that displayed an extremely compacted PM were also disregarded due to the high probability of poor resin infiltration throughout. All of these guts were, however, analysed under TEM. Flies chosen for SBSEM and 3View® analysis were dependent on having a good contrast under TEM, little evidence of poor resin infiltration and enough material per section.

One of the more interesting phenomena were the huge cyst like structures that could be found more commonly in older (11 d.p.i.) flies. This has never been reported before and presented an opportunity to provide insights as to what is happening within the PM layers. It is not known if trypanosomes are able to replicate here and thus are creating these cysts or whether this is a mechanism of the PM for eradicating the trypanosomes from the fly gut. A gut from a fly at 11 d.p.i was chosen (Fig. 5.15) for the reasons stated above.

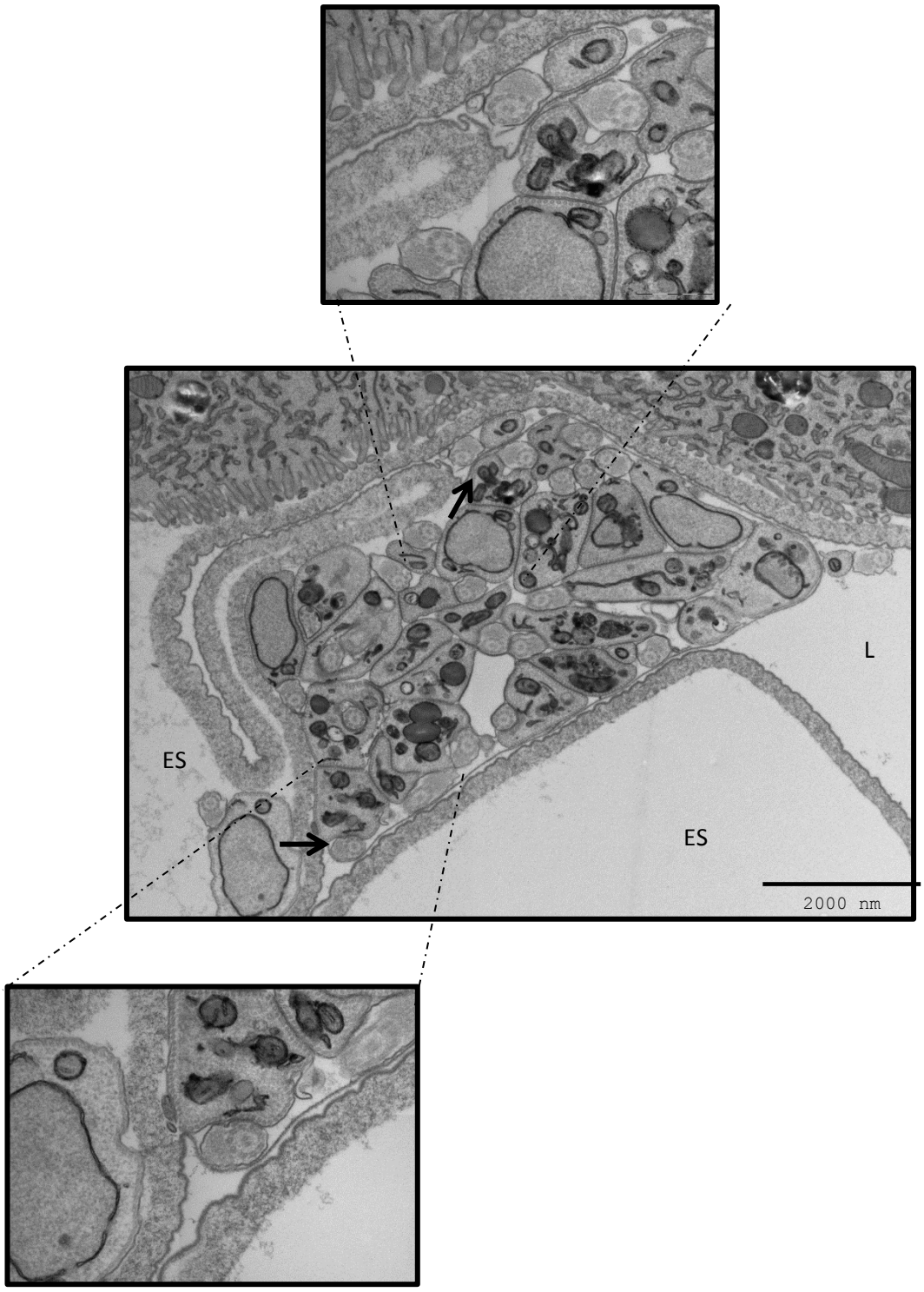


Figure 5.15. A 'cyst' of TSW196 trypanosomes in-between PM1 and PM2 at 11 d.p.i. *Inset shows a close up of where the electron dense luminal layer is being peeled away from the 2nd layer (arrows on main image) as the trypanosomes reside inside. However, both layers remain unbroken. ES: Ectoperitrophic Space, L: Lumen.*

By using SBSEM, it was possible to obtain over 500 serial sections of the sample shown in figure 5.15 (Ch5_S1), and then reconstruct this particular area of interest (Ch5_S2) by 3D tomography. With the assumption that trypanosomes cross the mature PM it was surprising to see that all parasites seen were contained within the PM and there was no indication that they were crossing nor dividing (Fig. 5.16.).

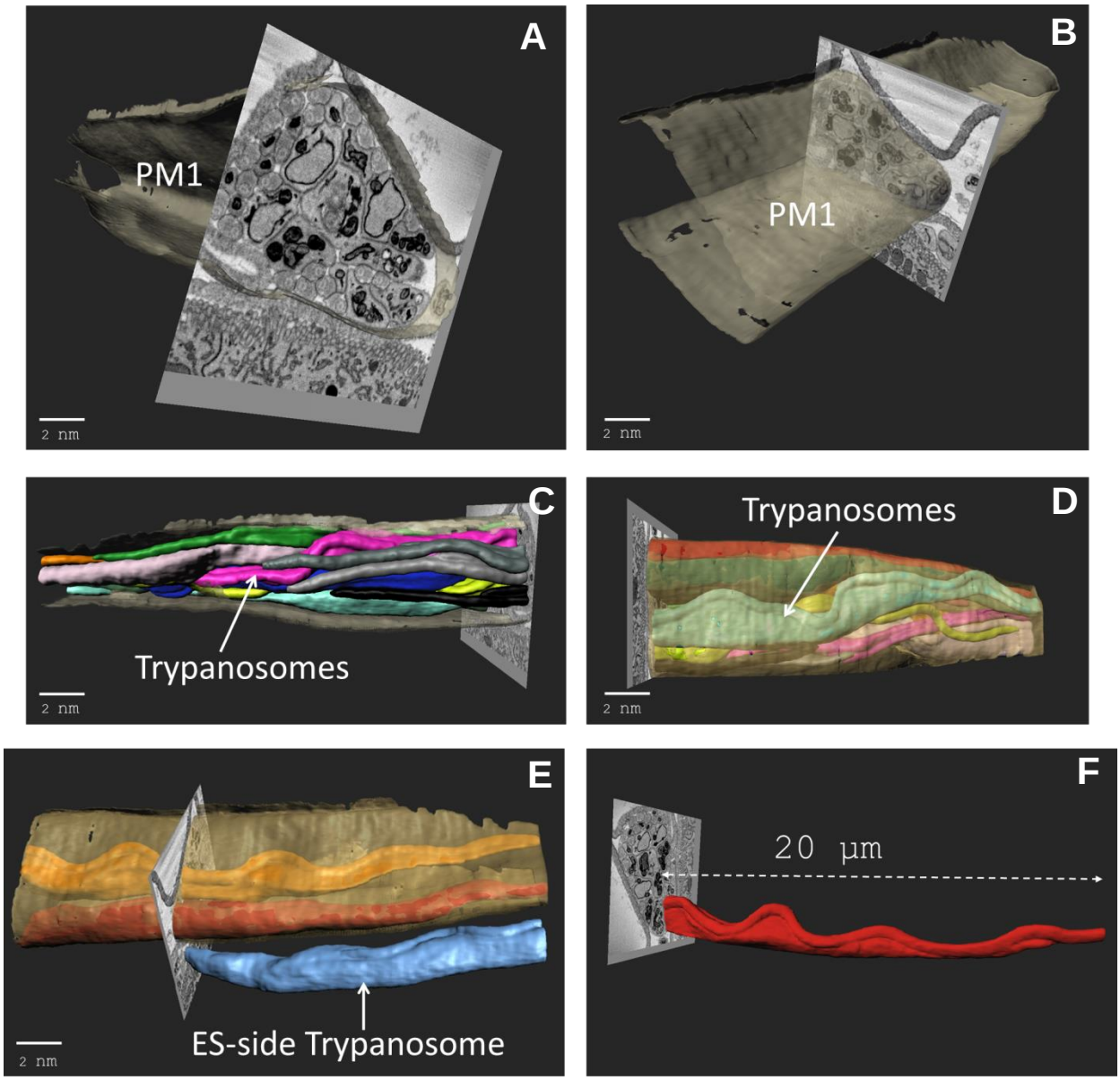


Figure 5.16. Stills taken from a 3D reconstruction of the tsetse midgut anterior to the bacteriome from an infected fly at 11 d.p.i. *(A) SBSEM slice merged with manual segmentation to illustrate the 1st electron dense layer of the PM. (B) A different point of view as that in A, showing no breaks or damage to PM1. (C) SBSEM slice merged with manual segmentation to show the multiple parasites seen between PM1 and PM2. (D) A reverse view of that shown in C, showing that the parasites are contained within the 1st and 2nd PM*

layers as seen by PM1 surrounding the trypanosomes. (E) SBSEM slice merged with manual segmentation showing a parasite in the ectoperitrophic side that is not interacting with the PM in any way. (F) A measurement of one of the trypanosomes contained within the cyst giving an indication of life-stage.

The results strongly indicate that trypanosomes don't appear to cross the PM at this point. There is no evidence that the parasites are able to cross a more mature PM nor do they appear to be dividing here. This suggests that the trypanosomes appear to be at a dead-end and that the tsetse PM acts as a sticky-trap engulfing any parasites that are caught within it. As discussed previously, the PM is synthesised at a steady rate of approximately 1mm per hour. This is true for newly emerged flies for up to 30 hours post emergence and for a short period following feeding. For flies that remain unfed following emergence the rate of PM synthesis reduces significantly after 30 hours and only reaches the hindgut after 3-4 days [188]. It could therefore be assumed that the PM functions as a conveyor belt and carries trapped parasites in the PM along the entire length of the gut and excretes them with the PM via the rectal spines in the hindgut. This implies a race against time for the trypanosomes in which they must escape the confines of the PM to the ectoperitrophic space. However, as breaks in the PM where trypanosomes occur have not been observed and reconstruction shows no transverse crossing of the PM, it is more likely that the PM encapsulates the parasites and excretes them without the trypanosomes being able to escape. Previous analyses using GFP expressing trypanosomes have shown that there are very few parasites that can be seen in the hindgut of flies dispelling the theory that trypanosomes by-pass the PM via circumnavigation around the open ended PM in the hindgut. However, as GFP expression is more obvious in actively metabolising cells, those parasites that are unhealthy and have a slower or abnormal development may show a reduction in GFP expression so may have been missed in fluorescent assays. In order to see if trypanosomes can be found in the hindgut in cyst-like structures as seen in the midgut, TEM analysis was carried out (Appendix: Fig. A6). Trypanosomes could be seen only in the ectoperitrophic space and were never observed in the lumen. The PM was seen to be thinner in this region of the gut which contained signature epithelial cells that are found in the hindgut and are associated with nutrient absorption. This region was also seemingly dehydrated demonstrated by the shrunken PM which obscured the lumen. Although no cysts were observed containing parasites, there was multiple damage to the PM consisting of a separation between PM1 and PM2. This is similar to what has been seen in the midgut although the holes in the PM in the hindgut were mostly empty. In addition, trypanosomes were seen either very rarely looking like "healthy" cells but most commonly could be seen to be abnormally shaped and some with multiple flagella.

5.3.8 TEM analysis of the tsetse proventriculus

One theory that was never further investigated was that trypanosomes cross the immature, freshly secreted PM near to the PV. Also, in younger flies the PM is secreted at a faster rate than in older flies so would be more fluid-like. Often in experiments, the PV is not looked at until at least 10 days after infection as this is when the PV is packed full of proventricular form trypanosomes. However, as there was no evidence of trypanosomes crossing further down the midgut, it was decided that the PV should be investigated at 5 and 11 d.p.i. time points to see if the trypanosomes cross the PM at its point of production. There have been very few studies on the tsetse PV so it is worth a recap of the literature in order to understand the PV morphology and its role in PM production (Fig. 5.17.). Its basic structure, as seen in a longitudinal plane consists of the oesophageal duct, the crop duct and the start of the midgut. The funnel shaped invagination of the foregut is covered by a chitinous intima and conceals an annular pad of high columnar epithelial cells which indicates the origin of the midgut.

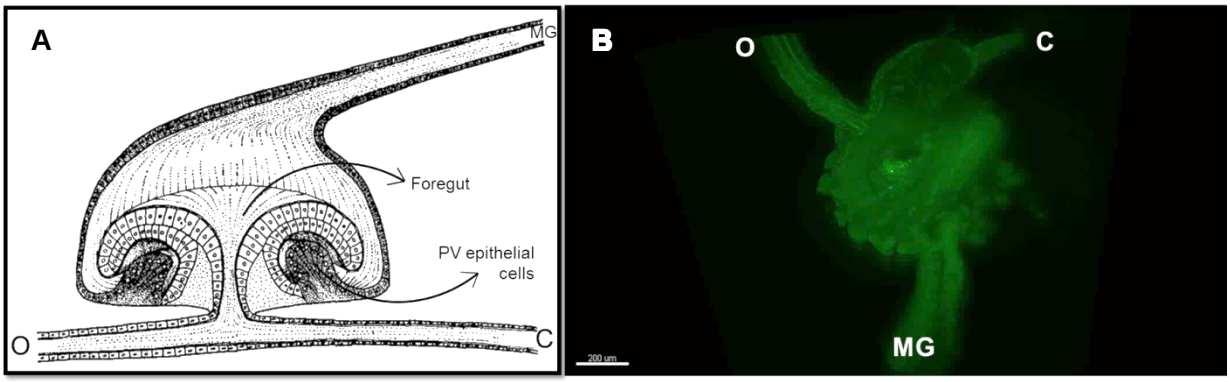


Figure 5.17. Structure of the tsetse PV. *A; Schematic of a tsetse proventriculus as seen in a sagittal plane. Adapted from [206]. B; A tsetse PV taken from the maximum projection view as seen under light sheet microscopy. Fluorescence is auto-fluorescence from the tsetse tissue. O; Oesophagus, C; Crop duct, MG; Midgut.*

The PM is secreted by the large columnar cells that mark the beginning of the midgut. There are three cell types that the annular pad consists of, each one morphologically and functionally distinct. The first cell types are confluent with the foregut cells and have very small and sparse microvilli (Appendix: Fig. A7b.), whereas the second cell type have slightly larger and more numerous microvilli and are in close proximity to the foregut cells (Appendix: Fig. A7c and d); both are responsible for the secretion of the first electron-dense PM layer. The third cell type of the annular pads present a large presence of endoplasmic

reticulum and contain large amounts of microsomes which appear to be mainly in the vicinity of the microvilli. The third cell type is presumably responsible for the secretion of the amorphous, thicker, electron-opaque second PM layer. PM1 is secreted first and PM2 is laid on top before they are 'pressed' together in the junctions of the annular pads and foregut invagination. PM3 is not always present along the entire length but when it is seen, it looks as though it is a definite part of the PM structure and can sometimes be seen to being laid down on the more mature PM (Appendix: Fig. A8). This occurrence can be seen in naïve, refractory and infected flies at all time points observed.

In infected flies it was surprising to see that by 5 days post infection, the PV was packed with trypanosomes (Appendix Fig. A9 and Fig. 5.18). Parasites could be seen in the spaces between the annular pads and the invaginated foregut although at this time point they were not seen to be penetrating the PV cells (Fig. 5.19). They were also observed in the ecto-peritrophic side of the PM and in some cases were extremely close to the epithelial cells of the PV. Parasites could sometimes be seen between PM1 and PM2 and the PM was also seen to be broken in certain places. However, it is hard to tell whether this was due to the tissue preparation or in fact if the trypanosomes themselves are actively breaking the PM or it is due to the sheer force of parasites acting on the fragile, immature PM. In addition, all of the midguts that were examined after seeing trypanosomes in the PV showed a midgut infection just as what has been seen previously. The PV shown in figure 5.18 was subsequently processed for SBSEM at the areas of interest shown (Ch5_S3) and (Ch5_S4). A partial reconstruction of some of the flagella that are in close proximity to the chitinous foregut was carried out (Ch5_S5 and Fig. 5.20).

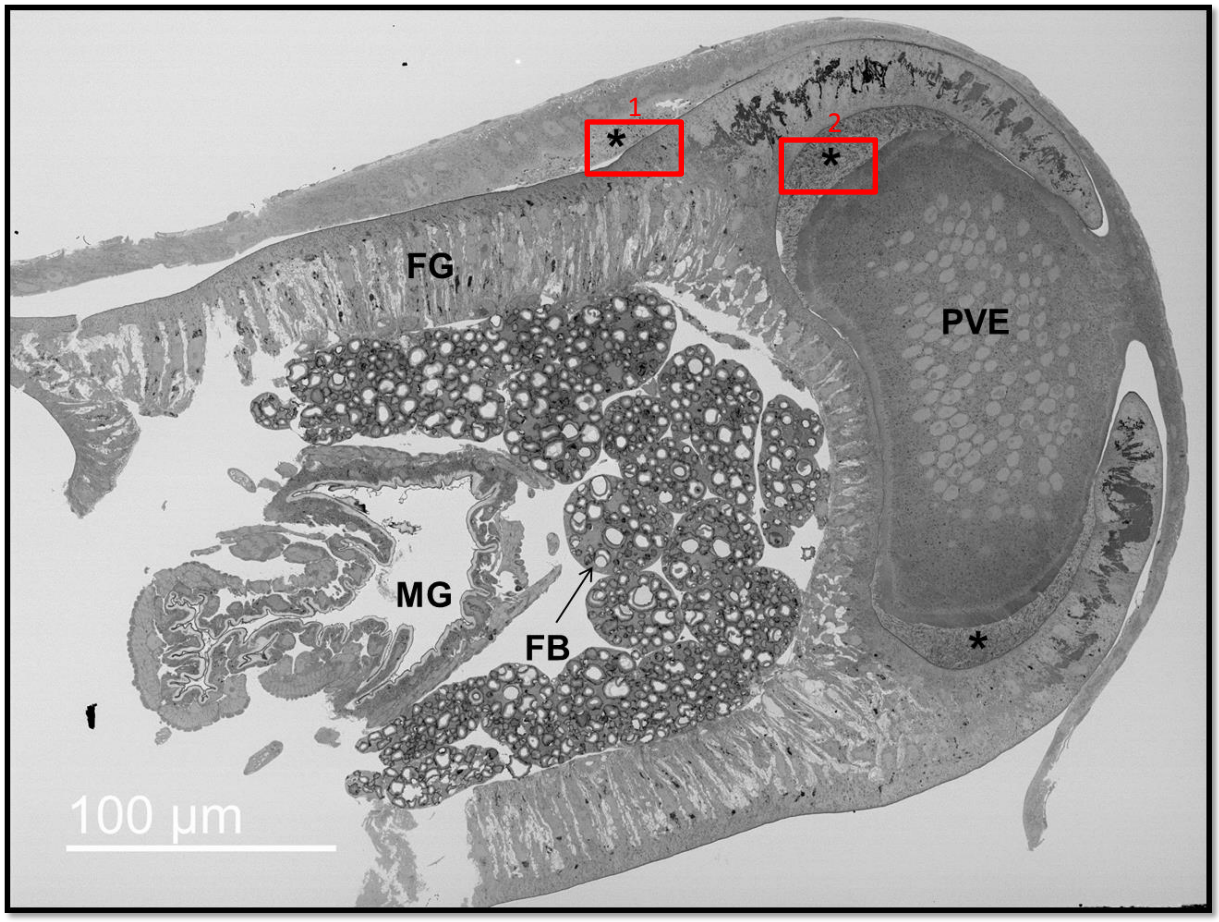


Figure 5.18. An SEM micrograph showing the overview of an infected PV at 5 d.p.i. in a transverse plane. *Trypanosomes (*) can clearly be seen occupying the spaces between the foregut and the annular pad of the PV. The red boxes indicate the regions of interest (ROI) that were subject to SBSEM. FB; Fat bodies, FG; Foregut, MG; Midgut, PVE; Proventricular epithelial cells.*

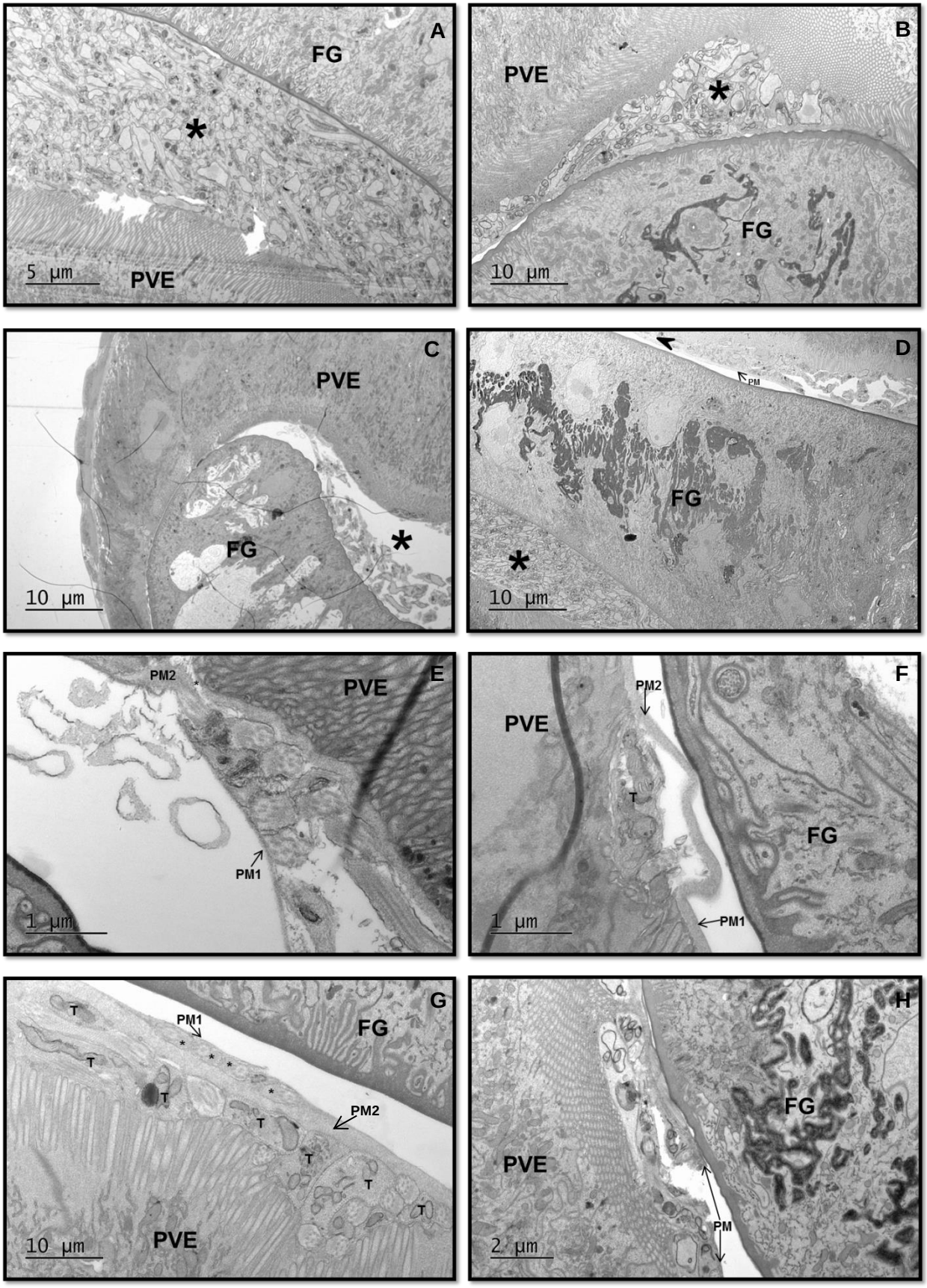


Figure 5.19. A selection of TEM micrographs of an infected tsetse PV taken at 5 d.p.i.

Details of each are shown in the text below.

Fig. 5.19a. At 5 d.p.i. the PV can be seen to be heavily infected. Trypanosomes (*) can be seen in the space between the foregut and the PV epithelial cells.

Fig. 5.19b. Trypanosomes (*) can be seen near to the epithelial cells of the PV annular pad but are not seen to be penetrating them.

Fig. 5.19c. In a heavy PV infection, the majority of the trypanosomes can be found in the vicinity of the type I and II cells of the PV, but some are found in the bigger space between the foregut and the type III cells of the PV (*).

Fig. 5.19d. Trypanosomes can be seen at this stage already in the Ectoperitrophic space side of the PM. The PM appears broken at this point. The arrowhead represents the continuation of the PV epithelial cells that mark the start of the midgut. Parasites (*) can also be seen in the space between the foregut and PV epithelial cells.

Fig. 5.19e. Trypanosomes can be seen in between the layers of the PM. The flagella can be seen between PM1 and PM2 and the two layers appear to have separated quite a distance from each other. The PM also appears broken (*).

Fig. 5.19f. Trypanosomes (T) can be seen already in the ectoperitrophic side of the PM with no obvious damage to the PM.

Fig. 5.19g. The flagella (*) from multiple parasites can be seen in between the layers of PM1 and PM2 with very little separation to be seen. Trypanosome bodies (T) can be seen next to the epithelial cells of the PV.

Fig. 5.19h. Trypanosomes can be seen next to the PV epithelial cells and the PM surrounding them appears broken.

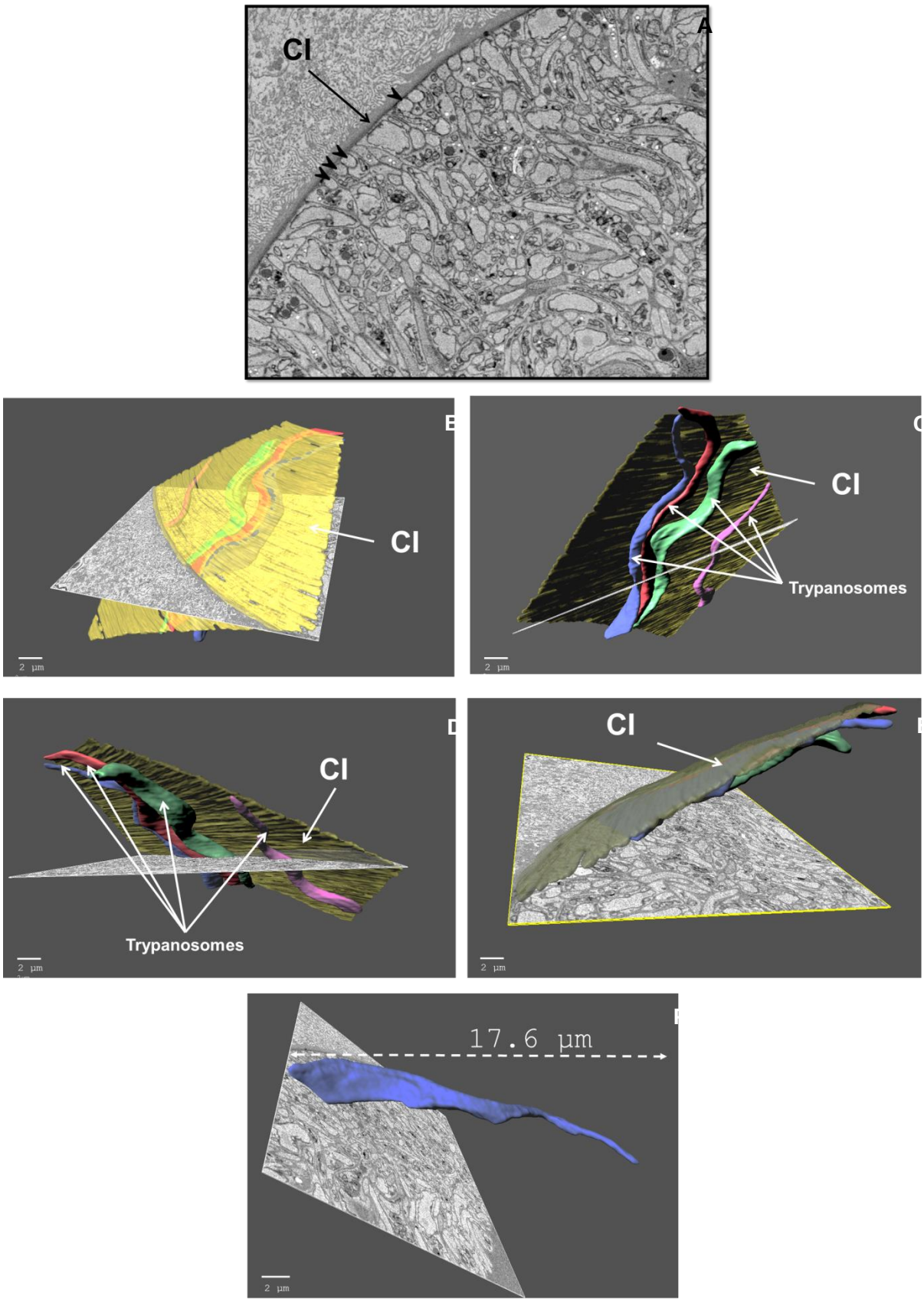


Figure 5.20. Stills taken from a 3D reconstruction of the tsetse proventriculus from an infected fly at 5 d.p.i. *(A) A slice taken from a series of 458 slices by SBSEM showing*

the structures that were processed for manual segmentation and 3D reconstruction. Arrowheads are the flagella from four individual trypanosomes and the arrow indicates the chitinous intima of the foregut cells. (B) SBSEM slice merged with manual segmentation to illustrate the undamaged chitinous intima (CI). (C) A reverse view of that in A showing trypanosomes against the CI but no penetration appears to be occurring. (D) Unlike in the cyst, trypanosomes don't appear to be facing in one direction. (E) The cuticle intima of the foregut cells surrounds the trypanosomes. There is no indication of attachment between the parasites and the foregut cells. (F) A measurement of a trypanosome giving a indication as to what life stage the parasite is in.

It appears that at 5 d.p.i. parasite are invading PM, sometimes getting stuck between the PM layers and sometimes are able to successfully to cross to the ectoperitrophic space. At this point, neither the PV epithelial cells of the annular pads nor the cells of the invaginated foregut appear to be penetrated by parasites. In order to see if trypanosomes stay within the space between the foregut and annular pads throughout infection or if they penetrate the epithelium, TEM was carried out on PVs at 11 d.p.i Fig. 5.21).

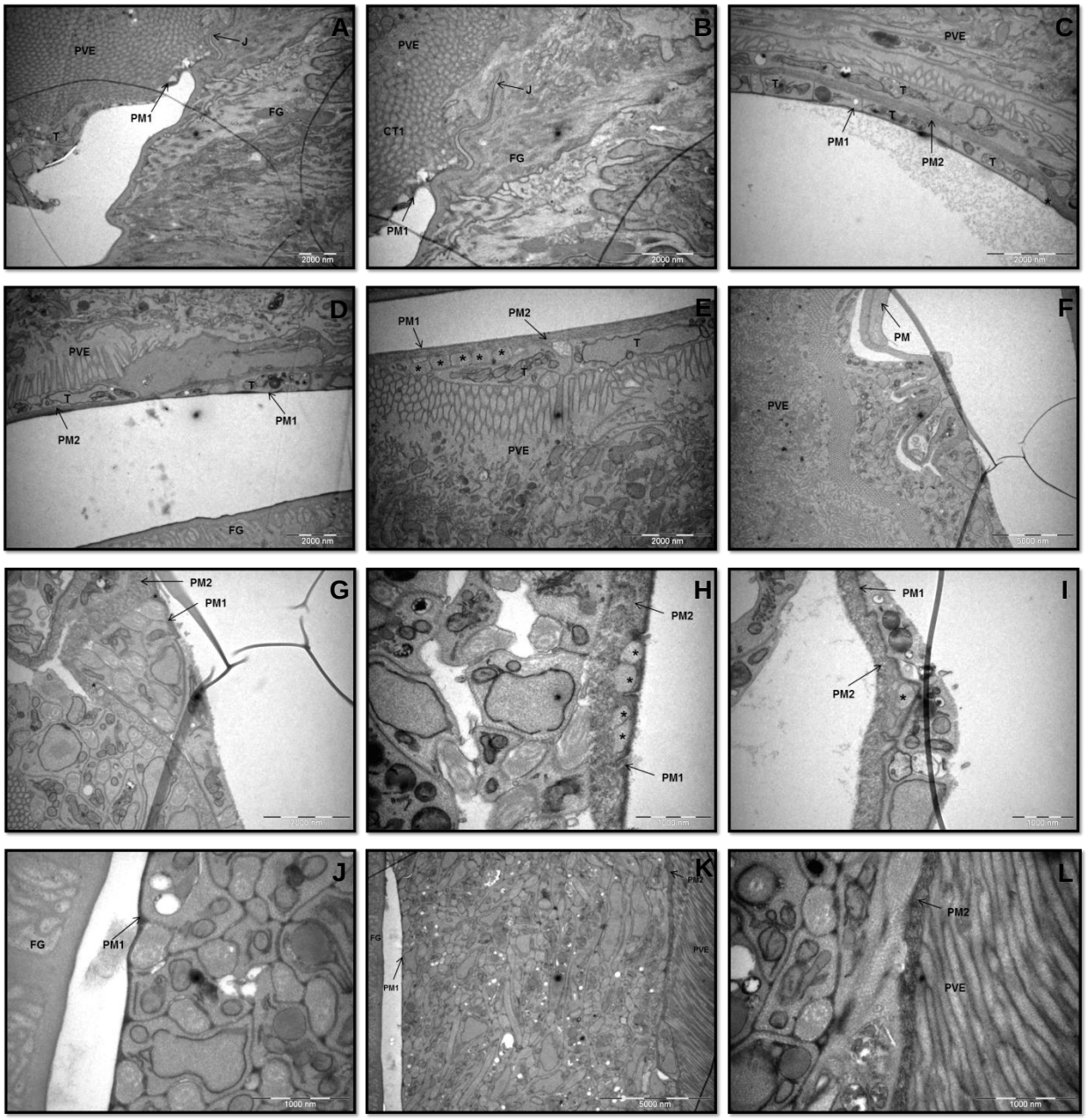


Figure 5.21. A selection of electron micrographs showing trypanosome interactions in the PV from a fly at 11 d.p.i. *Details of each image are in the text below.*

Fig. 5.21a. The junction (J) between the foregut (FG) and the proventricular epithelial cells (PVE). PM1 can be seen delaminating from the cuticular portion of the foregut and extending along the PVE where the first cell type is located. Trypanosomes (T) can be seen between PM1 and the PVE. 8200 x

Fig. 5.21b. A close-up of the junction (J) between the PVE and the FG. The junction can be seen extending slightly between the two cell types and PM1 can be seen extending backwards to cover the type 1 cells (CT1) of the PVE. 11200 x

Fig. 5.21c. Trypanosomes (T) can be seen between PM1 and PM2 and also between the PM and the PVE. The separation of PM1 from PM2 is indicated by an *. 11200 x

Fig. 5.21d. Similar to that seen in 22c, trypanosomes (T) can be seen both within the 1st and 2nd PM layers and between the PM and the PVE. 8200 x

Fig. 5.21e. Trypanosomes (T) are between the PM and the PVE. Multiple flagella can be seen (*) in close proximity to the PM with only two parasite bodies in view. 11500 x

Fig. 5.21f. Trypanosomes are located on both sides of the PM with no breaks or damage to either PM1 or PM2. The parasites seem to fit around the contours of the PM. 4200 x

Fig. 5.21g. A cyst-like structure can be seen containing multiple parasites, which closely resembles the pockets of trypanosomes most commonly found in the midgut of the fly. PM1 and PM2 can clearly be seen to be separated but remain unbroken. 11500 x

Fig. 5.21h. Trypanosomes are visible between the PM and the PVE. There are four flagella-like structures in tandem that can be seen between PM1 and PM2 and PM2 appears to be more unstructured around these flagella-like structures. 20500 x

Fig. 5.21i. A trypanosome can clearly be seen on the luminal side of the PM (i.e. between the foregut cells and the PM) and appears to be in intimate contact with the PM. A portion of what is presumably the trypanosome (*) can be seen permeating between PM and PM2. 16500 x

Fig. 5.21j. A close up of PM1 showing parasites contained within this layer which remains unbroken. The foregut can be seen with no trypanosomes between it and the electron dense PM layer. 26500 x

Fig. 5.21k. A view of the masses of parasites that can be seen between PM1 and PM2. Both layers are unbroken and there are some parasites that are outside of the PM and lie between the PVE and the PM. 8600 x

Fig. 5.21l. A close up of PM2 showing it lying between the trypanosomes and the microvilli of the PVE. 26500 x

Because it is difficult to see at TEM level how the parasites are interacting with the PV tissue and also to see spatially where the parasites lie in relation to the PM, scanning electron microscopy (SEM) was carried out on uninfected, refractory and infected PVs at 5 and 11 d.p.i (Fig. 5.22).

Trypanosomes could be seen embedded into the surfaces of the tissues via their flagella and in some cases their posterior end.

5.4.9 SEM analysis of tsetse PV

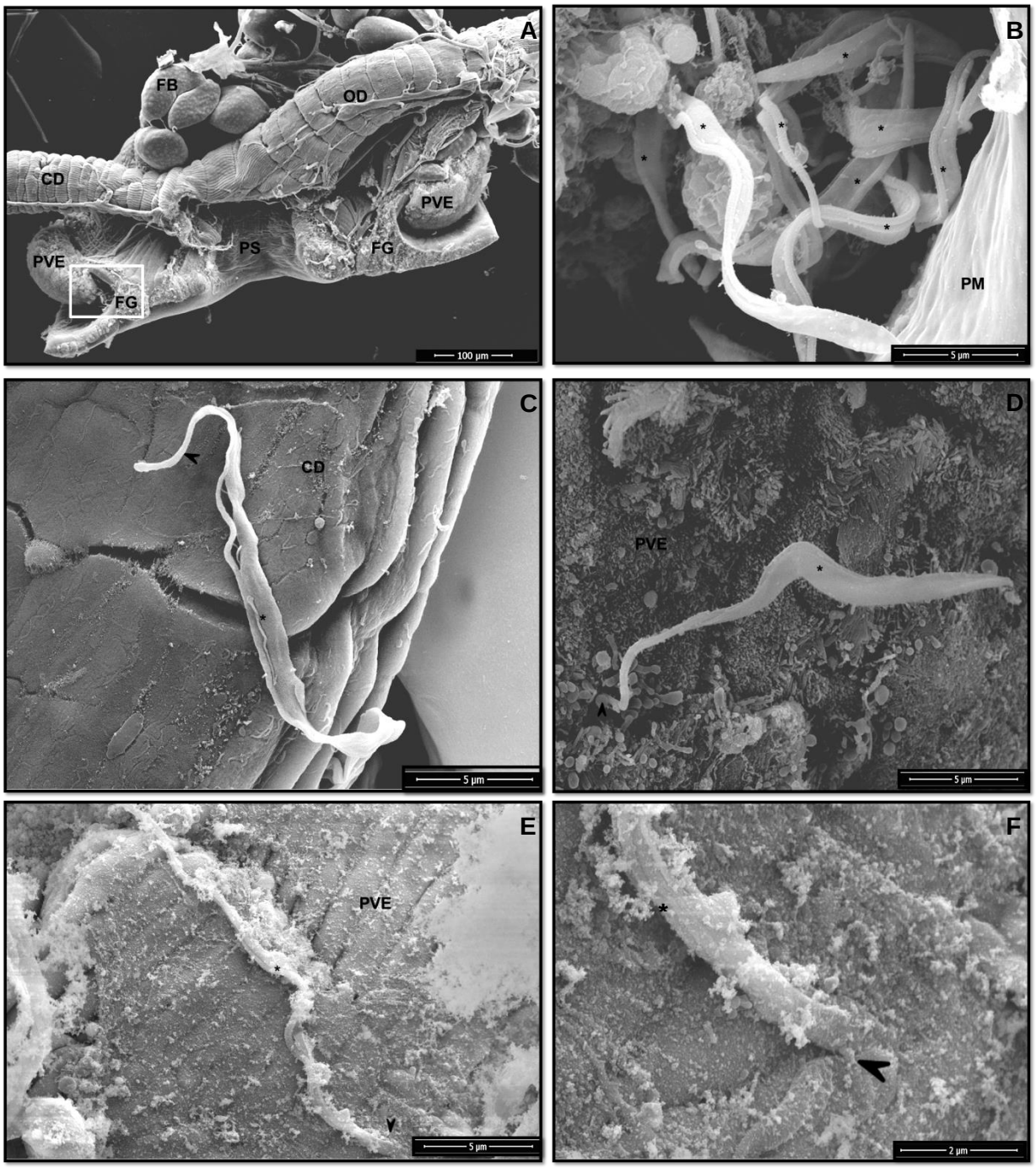


Figure 5.22. A selection of SEM micrographs showing trypanosomes in the PV from a fly at 11 d.p.i. *Details of each can be found in the text below.*

Fig. 5.22a. A sagittal view of the tsetse PV. The foregut (FG) cells and the PV epithelial (PVE) cells are evident. The crop duct (CD) and oesophageal duct (OD) can be seen evidenced by their rings of thick circular muscle. The proventricular sinus

(PS) can be seen as the midgut is missing. PV-specific fat bodies (FB) are also present. The white box indicates a close up view as seen in B. 400 x

Fig. 5.22b. The region of interest as shown in A. Trypanosomes (*) can clearly be seen in the space between the foregut cells and the epithelial cells of the annular pads. The PM is also evident. 10000 x

Fig. 5.22c. A long trypanosome (*) can be seen in within the muscle tissue of the crop duct. The posterior end is in the muscle wall whilst the anterior end is free. The flagella tip (arrowhead) can be seen. 9000 x

Fig. 5.22d. A trypanosome (*) can be seen in the PV epithelial cells with the flagella (arrowhead) seeming embedded into the tissue of the annular pad. 8000 x

Fig. 5.22e. A long PV form trypanosome (*) in the PV epithelial cells. The posterior end of the trypanosome seems embedded in the tissue (arrowhead). 7546 x

Fig. 5.22f. A close up of the posterior end of the trypanosome as seen in 22e. The posterior end of the parasite is seen embedded into the tissue of the annular pad. The asterisk indicates the insertion of the flagellum. 25443 x

5.5 Discussion

One of the suggested functions of the tsetse peritrophic matrix is as a barrier to pathogen invasion. For 40 years it has been assumed that *T. brucei* species initially develop in the tsetse midgut lumen before crossing the peritrophic matrix in the first step of an epic migration to the salivary glands. It was presumed that the parasites cross near the anterior of the gut due to unfavourable conditions in the posterior region. However, multiple studies have failed to pinpoint the exact moment in which the trypanosomes cross the PM. One such study involved tracking GFP expressing trypanosomes over many days within the tsetse midgut and showed that by day 5, flies either exhibited an established infection where trypanosomes could be seen in the ES from day 6 onwards or they had cleared the infection completely. Although this study showed GFP trypanosomes lining up at the edge of the PM with the assumption they are sensing or binding to the PM, the actual penetrating or crossing event was not captured. As crossing or penetration is a precondition of adherence and such attachment receptors or ligands have failed to be identified within the tsetse PM (chapter 2), this study aimed to determine the missing link in trypanosome development by re-visiting their migration through the fly midgut.

5.5.1 PM staining reveals holes in the PM of infected flies

Staining of the tsetse PM by WGA-rhodamine allowed better visualisation of trypanosome interaction in the midgut. Through fluorescent microscopy on isolated PMs, it was clear to see that there were holes in the PM made obvious by the lack of rhodamine emission where the parasites were found. Confocal microscopy helped to verify that these were holes and not due to the trypanosome being a physical barrier to the rhodamine emission by showing rhodamine emission above and below the parasite. This suggested that the trypanosomes were inside the layers of the PM, although there was no evidence that they had got there through penetration of the tissue as the parasites were contained and there were no other obvious damage or empty holes along the entire length of the PM. In addition, there were no irregularities to the PM staining in the absence of trypanosomes suggesting that trypanosomes are not secreting proteases or lyases that may disrupt the architecture of the PM in order to gain access to the ectoperitrophic space.

5.5.2 TEM analysis of infected flies shows damage to the PM

Initial electron microscopy studies on the anterior midgut posterior to the bacteriome showed similar evidence to what has previously been recorded. In fly guts taken from flies at 5, 8 and 11 d.p.i., trypanosomes could be seen in both the lumen and the ES, with the proportion on each side of the PM seemingly similar. In addition, trypanosomes could be

seen inside the layers of the PM with a separation of PM1 and PM2 being the most common. This has previously been recorded and was thought to be the point of PM crossing by trypanosomes, However, from all sections viewed, there is no evidence that the parasites are able to transversely cross the PM due to lack of damage through all 3 PM layers. It may be that in a few instances where parasites can be seen in PM2, that they are able to escape from this layer into the ES. However, this damage was seldom seen and there was no evidence of trypanosomes emerging from PM2. This type of damage either probably occurs only by chance or, because PM2 consists of secretions from the midgut epithelial cells, is able to repair itself after trypanosomes have broken through. This may explain why this type of damage is rarely seen as it would need to be occurring at the point of fixation and would need to be sectioned at that immediate point. In addition, sometimes multiple parasites were seen between PM1 and PM2, ranging from 2-3 cell bodies up to many that formed cyst-like structures. In all but two instances where parasites were found between layers, neither layer of the PM was broken completely, rather they seemed to engulf the trypanosomes. In those instances where PM1 was seen to be completely broken may be explained by poor fixation as judged by the condition of the midgut cells in these samples. Following these initial trials, the fixation and staining methods were significantly improved and a complete break of PM1 was never again seen. There are other discrepancies with infected gut tissue that have not been previously recorded but are worth noting here (Ch5_S6). Some may be attributed to how the sample was fixed and subsequently processed, however it is not certain if these findings are true in some flies or artefacts of the methodology used. Even though great care was taken to process samples in the exact same way, there may be variations between different sample batches. Nevertheless, reporting these sporadic inconsistencies will add more strength to the majority of what is seen as being real occurrences between trypanosome-PM interactions.

5.5.3 Trypanosomes do not cross a mature PM

It was initially presumed that the trypanosomes may be dividing between the two layers of the PM as well as in the ES. Here the parasites would be protected from the midgut enzymes and immune system components and there was no evidence that they could divide in the lumen of the gut. It is still unknown if those trypanosomes seen in the lumen of the midgut are viable and are able to continue development or whether only those that are in the ES are proliferative. Previous studies have shown that those trypanosomes seen in the ES are longer and thinner than those seen in the midgut lumen [213]. However, because only a small proportion of holes seen within the two layers contained more than one parasite, it would be more likely that these parasites are trapped and are moribund. Additionally, in most cases there was no evidence that the trypanosomes could burst out

through PM2 and PM3 due to a lack of transverse damage, nor were any holes found that appeared empty in this region of the midgut. The evidence presented by 3View® analysis and 3D reconstruction supports this.

5.5.4 Parasites are eliminated from the fly after becoming trapped in the PM

To see if the PM was trapping and excreting the trypanosomes through the rectum, the hindguts of infected flies were looked at. The damage to the PM in which PM1 and PM2 were separated was obvious and looked similar to those found further up the gut. However, the holes that sill remained were either always empty or contained debris that was presumably a trypanosome, unlike those previously seen which in the majority of cases almost always contained at least one visible parasite. This could suggest that the parasites have undergone apoptosis (programmed cell death) by this point or alternatively because the PM is more mature and rigid in structure that the holes left between the layers from trypanosomes that may have escaped are unable to close and the layers unable to re-join. The trypanosomes in this region also appeared very different to those seen in the anterior gut. They looked abnormally round in comparison to the more triangular profile of those seen in the anterior gut. In addition, in some sections there appeared to be a distinct lack of flagella that were associated to any cell body and in others, trypanosomes appeared multi-flagellated. Multi-flagellation occurs due to problems with cytokinesis and the cells do not divide normally. Because cell body duplication is always preceded by flagella duplication, abnormal division results in numerous flagella that are associated with a single cell. Some of the trypanosomes seen resemble mutant parasites that have been RNAi induced to knockdown genes associated with cytokinesis [214]. These cells form aggregates of cells arrested in late stage cytokinesis. Moreover, trypanosomes seen in the hindgut appeared to have many more glycosomes and acidocalcisomes than their counterparts in the anterior gut, suggesting that these cells are abnormal. Glycosomes are essential in trypanosomes and are responsible for regulating the glycolytic pathway. Studies have shown that a decrease in the number of glycosomes leads to apoptosis through osmotical frailness [215] and it's possible that the cells found in the hindgut are over-compensating and over-expressing glycosomes in response to the change in environment. Acidocalcisomes in kinetoplastids are involved in osmoregulation and have been shown to induce autophagy in *T. brucei* in response to stress [216]. The increase in numbers of this organelle may be a coping mechanism due to the unfavourable conditions in the tsetse hindgut. In refractory and uninfected flies, the PMs from the hindgut were relatively normal looking with no breaks or damage seen. (Ch5_S7). The thickness of the PM was the same as PMs seen in the anterior gut of uninfected and refractory flies, suggesting that once the PMs in infected flies become damaged, the damage is irreversible and the PM is unable to repair itself at this

point. In all gut analyses, the thickness of the PM was consistent and there were no obvious changes in thickness in infected flies compared to refractory or uninfected flies. This suggests that PM architecture is not affected by trypanosome infection.

5.5.5 Trypanosomes need to penetrate the immature PM in the PV to establish midgut infection

As the evidence suggested that PM invasion by trypanosomes did not occur through the mature PM in the anterior part of the midgut, focus shifted to examining the PM at its point of production in the proventriculus. Here, trypanosomes could be seen already in the space between the epithelial cells and the PM in what would become the ectoperitrophic space. This is observed in PVs from both 5 and 11 d.p.i. flies but with differences in the manifestation of the trypanosomes in the PV. One of the more striking differences between an infected PV at 5 d.p.i and one at 11 d.p.i is that the PM in the 5 d.p.i PV seems more fragile. Parasites could already be seen in both the ES and the lumen and also between the layers. However, in some instances the PM was transversely broken and trypanosomes could be seen between and near to these breaks. The parasites seemed to fill the entire PV cavity between the foregut and the annular pads whereas in the 11 d.p.i, trypanosomes were neatly contained within the PM and were lying adjacent to the PV epithelial cells. Very few trypanosomes could be seen in the lumen and there were none that were in proximity to the cuticular intima of the foregut in all sections viewed. In addition, no breaks or transverse damage could be seen in any 11 d.p.i. samples viewed. At 5 d.p.i., flies show two clear phenotypes; those that have a huge parasite burden and those that show no sign of infection. At this point in the infection, trypanosomes can be seen everywhere in the PV, occupying and filling all available spaces. The huge number of parasites could be attributed to the initial bloodmeal acquisition or could be due to the trypanosomes replicating. By 11 d.p.i., the fly has either got the infection under control or more likely that the trypanosomes themselves have regulated the infection. It is known that procyclic trypanosomes regulate their numbers in the fly through programmed cell death (PCD) which is beneficial to the parasite both in terms of maturation and keeping the fly alive long enough to complete their lifecycle [217]. Due to the age of the flies, they would have received several more bloodmeals than the flies at 5 d.p.i and so the PM at this point is more mature than in a 5 d.p.i. fly even this close to its point of production and is able to contain the trypanosomes without breaking. Those trypanosomes that were seen near to the cuticular intima of the foregut cells in the flies at 5 d.p.i. are rarely seen at 11 d.p.i. These trypanosomes were probably from the initial bloodmeal and would have quickly died or been excreted by the fly, and the ones that are able to stay in the ectoperitrophic space are those that go on to mature. Overall, this data suggests that trypanosomes are capable of penetrating the PM at

its point of production near to the type I and II cells of the PV annular pads. At this point PM2 is not fully formed and exists as an amorphous structure, so it's possible for trypanosomes to become engulfed in it rather than actively penetrating as previously suggested. This may partly explain why only a small proportion of flies become infected whilst the majority are able to clear infection by 5 d.p.i.

5.5.6 Rate of diuresis and volume of bloodmeal may be important in potentiating midgut infection

Initially, trypanosomes come into contact with the crop before being passed into the midgut via the PV. In the crop, trypanosomes may be able to differentiate from stumpy bloodstream forms into procyclics before entering the PV, or they may become stuck in here and are unable to go any further in their development. Initial examination of infected flies from 1 d.p.i. through to 5 d.p.i. showed a range of infection phenotypes (data not shown). At 1 d.p.i. GFP expressing trypanosomes were not seen in either the crop or PV and only in 10% of flies could very few parasites be seen in the midgut. This may be due to the initial clearance of trypanosomes in the bloodmeal being excreted when the fly undergoes diuresis. At 2 d.p.i., again, trypanosomes couldn't be seen in the crop or PV but 90% of flies had low numbers of parasites in their midguts. By 3 d.p.i., trypanosomes were visible in the crop and PV and only half of the flies examined had trypanosomes in the midgut. Some also had clearly dead trypanosomes in the midgut. At 4 and 5 d.p.i, only 30% had parasites in the midgut but all of these infected flies had trypanosomes in the PV and crop. This suggests that by 2 d.p.i. the fly is already clearing most of the parasites through excretion of the bloodmeal and the fact parasites were not visible in the PV prior to this could be explained by their presence in very low numbers with the majority having been excreted. When a tsetse feeds, it can take up to 30µl of blood in one go. This exerts a huge amount of pressure on the PM and the midgut and its possible that only a small proportion of trypanosomes are able to sink through the PM at the site of production when the fly is engorged. This creates a small window of opportunity in which the trypanosomes can establish themselves. It's possible that the longer the fly is engorged the better chance the trypanosomes have at establishing due to a prolonged stretching of the PM. In addition, it can be assumed the longer a bloodmeal stays in the PV, the better chance the parasites have first to differentiate from bloodstream into procyclics and then to establish the infection. Once the bloodmeal passes through and the engorgement subsides, trypanosomes will have either passed through to the ectoperitrophic space, stayed in the luminal side, or else have become stuck between the layers of the PM. At this point they are moribund and unable to escape. If the fly feeds soon after, it is possible that the trypanosomes are able to escape as the PM becomes taut once more near to its secretion,

however because the PM is constantly secreted, it is more likely that these trypanosomes are carried along the gut in the PM and excreted in the hindgut. In addition, those trypanosomes that have been unable to cross the PM by the time the next bloodmeal is ingested will most likely be excreted with the force of the blood coming into the PV. However, it may be possible that these trypanosomes have a chance to penetrate the PM during the second bloodmeal when the PM is stretched tight once again if they are able to stay in the region of PM secretion. The rate of diuresis may also be a factor in successful trypanosome establishment in the fly. Slower rates of diuresis would mean that the rate of blood passage into the gut would be slower, providing ample opportunity for trypanosomes to become established.

5.5.7 PM maturity and production rate is important for determining infection

This study also provides insights into the teneral phenomenon whereby younger, unfed flies are more susceptible to trypanosome infection. The length of the PM in young flies is less than the length of the gut. It is a closed sac until it reaches the end of the hindgut after approximately 3-4 days. The rate of crop emptying is determined by the length of the PM and so large meals when there is a short PM means the blood stays in the crop longer and only a small amount initially passes through to the gut (Fig. 5.232). Older flies with a more mature (and longer) PM will have more of the blood meal passed through to the gut. This suggests that the longer the trypanosomes are contained in the first infected bloodmeal within the crop and PV, the more chance they have of firstly being able to differentiate into procyclics and also coming into contact with the freshly secreted, fluid like PM. Also, the lower rate of infection in older flies may result from a more mature PM in regards to protein components. Trypanosomes may be only to establish an infection in younger flies before critical PM components are laid down. These components may include peritrophins (chapter 4).

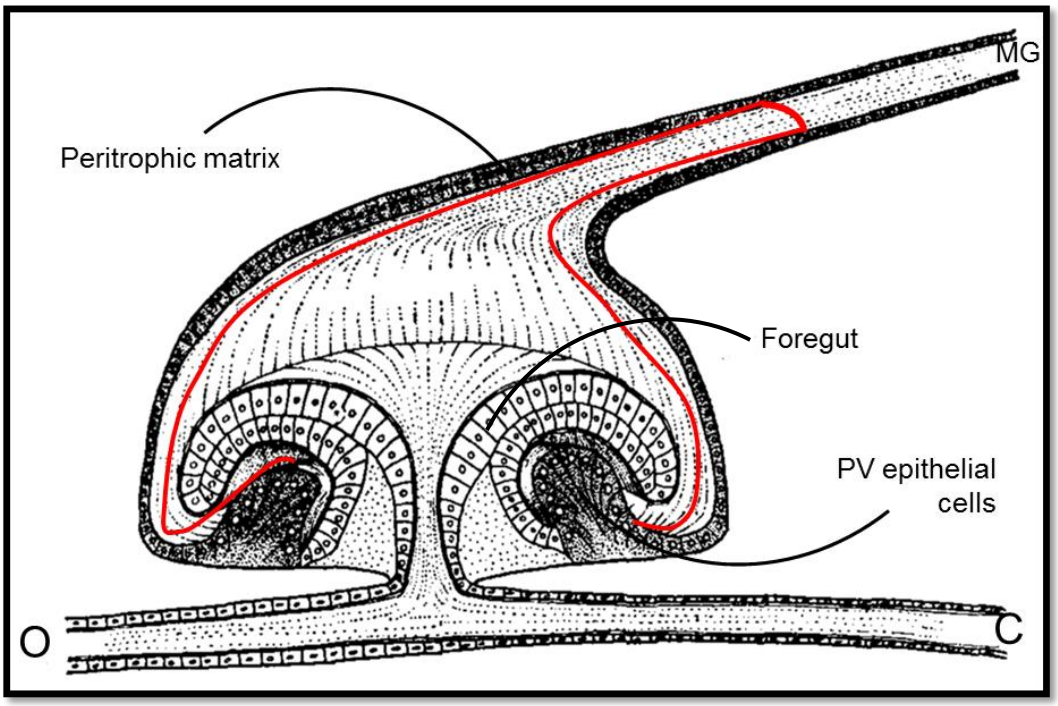


Figure 5.23. A schematic showing a sagittal section of the tsetse PV. *In young flies, the PM (red lines) forms a closed sac until it reaches the hindgut. Any blood (and parasites) ingested before the PM reaches the hindgut will be tightly contained in the crop and/or PV, only passing through the gut as the secreted PM extends along. Figure adapted from [206].*

5.5.8 New proposed mechanism for initial establishment of midgut infections by *T. brucei*

The collective evidence from these studies allows for a new proposed migratory route by trypanosomes in tsetse. It would seem that the initial midgut establishment by trypanosomes occurs in the proventriculus of the fly where they are able to penetrate the newly secreted, amorphous PM. After initial transformation and colonisation, it is predicted that the trypanosomes stay in the proventriculus throughout the course of the infection. This would allow them to stay in a relatively safe location away from the harsh environment of the midgut. At 11 d.p.i., trypanosomes are tightly contained between the PM and the midgut epithelial cells of the PV (i.e. the ectoperitrophic space). If they are able to stay in this location and to develop here then they are in a perfect position to be able to penetrate back through the soft PM region, into the proventricular sinus and back up the oesophagus *en route* to the salivary glands where they complete their transformation. Many studies have shown trypanosomes packed into the ES from as little as 4 d.p.i., however this may be explained by the 'overflow' of trypanosomes coming from the PV. It is presumed that trypanosomes cross the PM in the anterior region of the midgut and migrate anteriorly towards the PV, but until now it has never been suggested that those ES trypanosomes are coming directly from the ES in the PV (i.e. anteriorly to posteriorly). This may also explain

the many different lifecycle forms of parasite that can be found in the midgut at different time points during infection. It may be that they are not developing in the midgut but rather developing in the PV and those that are found further down the gut have either become trapped in the PM and secreted as it is pushed along or find themselves trapped in the midgut ES region due to the pressure exerted on them from the PV. It could be argued that trypanosomes undergoing proliferation in the midgut must be doing so because it's beneficial to the parasite otherwise the trait would have been lost during the course of evolution. Indeed, flies can present having a high midgut parasite load but no PV infection throughout the longevity of the fly. The inability of such an infection to progress to PV or salivary gland has previously been suggested as one reason for the low infection rates observed in tsetse. However, the evidence from this study suggests a different reason for the fly having a midgut but not a PV infection. It is possible that upon entering the PV, trypanosomes are triggered to immediately differentiate from stumpy forms to early procyclics. Then one of three things may happen; if the fly clears these parasites quickly then they become refractory, if the fly is unable to quickly clear infection then the fly will become infected, or the fly may clear the trypanosomes from the PV but parasites that have penetrated the immature PM and have become trapped in the ES may continue to develop within the ES of the midgut. In the latter case, the proliferating trypanosomes are able to continue the midgut infection during the flies life span but are unable to migrate anteriorly back to the PV and salivary glands and are therefore unable to progress to an mammalian infectious stage. Thus, only those trypanosomes that are able to initially colonise the PV immediately after bloodmeal ingestion can potentially complete the lifecycle in the tsetse. In addition, it is presumed that the ES is a safe environment for trypanosomes to develop in. However, this anterior part will still be flooded with epithelial cell secretions such as enzymes and immune components, which pass through the PM into the lumen, whereas the epithelial cells of the PV have no enzymatic secretory function and is probably a safer environment for the trypanosomes. Moreover, if the trypanosomes do develop in the ES of the anterior midgut, to find their way back towards the PV from that distance would take a lot of invested energy on their part. Due to the highly convoluted nature of the PM, the parasites would need to navigate their way back through many folds and may only be able to reach the PV if the fly takes a bloodmeal and the PM becomes stretched once more. It would, therefore be more beneficial to the trypanosomes to stay in the PV near to the region of PM secretion.

5.6 Summary/conclusions

The data presented in this chapter provides significant evidence that contradicts the previously proposed migration of *Trypanosoma brucei* through the midgut of *Glossina morsitans morsitans.* Here it was shown that trypanosomes have no penetrative power that allow them to degrade the mature tsetse peritrophic matrix, instead they have only a small window of opportunity in which to invade the PM and this occurs at the site of PM synthesis in the proventriculus. This partly explains why tsetse are poor disease vectors and why such a small number of wild caught flies present midgut infections. In addition, it is proposed that trypanosomes are able to complete the midgut life stage in the proventriculus of the fly and that only these are potentially able to mature to salivary gland infections. These findings present an opportunity to further investigate the role the PV and PM proteins that are synthesised only in mature PMs have in mediating trypanosome infection in *Glossina*.

Chapter 6. Overall discussion, conclusions and future perspectives

Ode to a supervisor…….

Chickpea Soup

Serves 1-2 hungry supervisors

Costs per person: 1 favour

500g dried chickpeas, soaked overnight in cold water (at least 12 hrs)

Drain and simmer in fresh water for 1 hour

If the water remains frothy keep on changing it until it is clear

Add 5 large peeled and chopped carrots

Add 1-2 large cloves of garlic, chopped

Add a bunch of spring onions, chopped

Simmer for a further 20 minutes

Add 3 vegetable stock cubes pre-dissolved in boiling water – adjust water to suit your consistency liking

Mix thoroughly and then blend until large lumps are no longer visible

Add a small drizzle of olive oil

Add pepper to season

Add chilli salt to your taste

If using canned chickpeas, there is no need to soak overnight or to simmer for one hour – just lob everything in together and simmer until carrots are cooked.

6.1 Concluding remarks on the tsetse PM and general summary.

Although the presence of a peritrophic matrix in insects has been known for over 200 years, it wasn't until the late 1990's that detailed studies of the components of the PM began to emerge with understanding of their role in PM function. Moreover, few studies have focussed on the molecular interactions between trypanosomes and their vector, *Glossina.* This is in part due to the lack of established tsetse colonies and the difficulties in maintaining them. *Trypanosoma brucei brucei* is the model kinetoplastid organism due to its easy genetic manipulation, lack of human infectivity and simplicity of using it in animal models, and such studies have greatly increased our understanding of trypanosome biology and how that theoretically relays in a natural environment. Despite this, however, relatively few studies have translated that into working tsetse models to see if what appears to be happening *in vitro* actually happens *in vivo*. Additionally, it has been only relatively recently that genetic manipulation of tsetse through RNAi was developed and so it opened a gateway for studying *Glossina* gene function and tsetse-trypanosome interactions. This current study has combined the use of trypanosomes, wild type and genetically manipulated tsetse in order to vastly increase our understanding of the *Glossina morsitans morsitans* PM morphology, structure and function, and how that relates to trypanosome migration through the fly. In addition, combined proteomics and bioinformatics have provided the first insights into related PM components in other *Glossina* vectors and a comparative non blood feeding species *Musca domestica.*

The tsetse PM is comprised of 3 layers, each differing in thickness and composition, and although known to comprise of a combination of chitin, glycoproteins and glycosaminoglycans (GAGs), the identity of these molecules was previously unknown. Here, a proteomics approach was carried out on teneral PMs using liquid chromatography tandem mass spectrometry (LC-MS/MS), which identified nearly 300 proteins from both in-gel and in-solution PM digestion.Most of these proteins could be classified according to GO terms, including enzymes involved in chitin metabolism and digestion, proteins involved in immunity such as the C-type lectins and tsetseEP, and those proteins containing chitin binding domains (CBDs), including peritrophins. This is the first report to identify 3 novel peritrophins as definitive constitutes of the PM; GMOY002708 (GmmPer66), GMOY009171 (GmmPer108) and GMOY011810 (GmmPer12). In addition to these, 27 proteins from *Sodalis glossinidius*, a commensal bacterium linked to increased susceptibility to trypanosome infection in tsetse, were identified suggesting a close relationship between the bacterium and the PM. The results of this study suggest that the *Glossina morsitans* PM is more complex than previously thought. Although many of the proteins originate from the

epithelial cells and are only transiently linked to the PM, they appear to have critical roles in the function of the PM, (discussed in chapter 2).

The major components of the PM, the peritrophins, are critical in determining the structure and formation of the PM and may be involved in potentiating trypanosome infection. By comparing peritrophins of *Glossina morsitans* to other *Glossina spp* and two non bloodfeeding Dipterans provided insights into their putative roles. Peritrophins can be classed into one of four groups: simple, binary, complex and repetitive according to their structural organisations, and each species analysed contained at least one from each group. Simple peritrophins are by far the most widespread throughout Dipterans and unlike many other insect species contain peritrophins containing all three types of chitin binding domain; PADs, PBDs and PCDs. Structural analysis revealed that both PADs and PCDs are likely to be involved in the majority of crosslinking within the chitin component of the PM as seen in type I PM models [37, 104]. Presence of peritrophins in the PM may already be regulated prior to full maturity but the degree of cross linking between chitin and other PM components is most likely low in immature PMs. This would present an opportunity for trypanosomes to breach the PM at this particular point before the PM matures and other PM constitutes are secreted and interlinked to the PM. This can be seen with GmmPer66 where protein expression seems to decrease after bloodfeeding and may indicate that whilst the 2nd layer of the PM is at an amorphous stage, as seen by TEM analysis, the peritrophin has not yet been extensively crosslinked with chitin and other proteins and so is more accessible.

With the identification of orthologues to *G. morsitans* peritrophins complete, it would be fascinating to see if the peritrophins in each species have the same role as proposed here. This may provide insights as to how certain species are more refractory to certain trypanosome species and may allow for the development of better control methods.

6.2 The PM as a barrier to infection

The migration of trypanosomes through the tsetse midgut has been well documented for over a century, with various theories for how they are able to cross the PM for development (Chapter 5). The hypothesis of direct penetration through a mature PM has been taken at face value for over 40 years, mainly due to TEM evidence. Despite this, there has been no direct evidence of PM crossing by trypanosomes. The study presented here provides profound evidence that trypanosomes in fact do not cross the mature PM in the anterior gut, but rather breach the immature, freshly secreted PM near to its point of production in the proventriculus. Extensive confocal and TEM analysis showed no evidence of a complete breach of the PM, nor could any damage be seen to all three PM layers at the same time.

In fact, very rarely was there a complete break of any layer and in the few times this was seen has most probably been related to poor fixation in initial trials. Additionally, there was no evidence of trypanosomes crossing as parasites were never observed to be emerging from any PM layers. Instead, in almost all instances, trypanosomes could be seen inside the first electron dense layer and the thicker 2nd layer. At first it was presumed that they were replicating here before penetration of the 3rd layer and escaping into the ectoperitrophic space. However, 3D reconstructions of a part of the PM containing multiple trypanosomes between PM1 and PM2 showed that trypanosomes were completely engulfed by the PM. Moreover there was no evidence to suggest they were replicating or that they were able to breach the PM and these cyst-like structures could also be observed in the hindgut. Subsequent analysis of the tsetse PV showed that, in some cases, the PV of tsetse could be seen to be highly infected as little as 5 days post infection whereas refractory tsetse showed no sign of infection. Moreover, by 11 days post infection, trypanosomes could be seen in the PV to be neatly contained between the PM and epithelial cells (i.e. in the ectoperitrophic space) as well as between the PM layers, and trypanosomes further down the midgut could be seen in the ES, the lumen and between the PM layers.

Although the PM has been suggested as a barrier to pathogen invasion, there has been no evidence to explain how it functions as a barrier, other than that it is a physical obstacle that the parasites must overcome. However, recent evidence suggests that the PM does not provide a physical protective barrier, as flies with a compromised PM were able to successfully clear normally lethal bacterial infections and succumb more readily to trypanosome infections. The reasons given were due to earlier immune components acting on both pathogens and determining the outcome of infection. The evidence presented here partially agrees with this. Although the more mature PM does offer protection, the immature PM in the PV is susceptible to breach by trypanosomes in flies at certain physiological states and previous studies have shown the PV to be an important organ in the immunity of insects [218]. There is a small window of opportunity which enables trypanosomes to overcome the tsetse PM in the PV. If they are successful in doing so it may be that they are able to regulate or manipulate the tsetse immune system by way of the PV. It has previously been shown that in early midgut infections, the transcription of two antimicrobial peptides, *attacin* and *defensin*, do not change significantly in either the gut or the surrounding fat bodies. However, as the infection progresses, they are significantly up-regulated but do not affect the viability of parasites. On the other hand the antimicrobial peptide *diptericin* is constitutively expressed in both the fat body and midgut of normal and immune-stimulated flies, which inhibits trypanosome establishment if previously immune-stimulated [219]. However, only the infection in the midgut was taken into consideration and

it may be that immune responses in the PV play a part in trypanosome clearance. In addition, it may be worth noting that the PV is surrounded by fat bodies both in teneral and mature flies, whereas midgut fat bodies are more noticeably found in older flies that have had several bloodmeals. This suggests that there may be PV-specific fat bodies, which have an immune function related only to the PV. Therefore, future research should focus on early infections to monitor the expression of immune related genes in response to early PV infections rather than midgut infections as the critical time may be when the trypanosomes first establish in the PV.

Collectively, evidences from these studies suggest a new proposed function for the tsetse PM and a new alternative for *Trypanosoma brucei* midgut development;

1. The tsetse PM not only acts as a partial physical barrier to trypanosome infection but also acts as a "sticky trap" to capture and remove trypanosomes. As mucins have been implicated in pathogen adhesion and elimination, it is proposed that the mucin domains of binary and complex peritrophins help to encapsulate the trypanosomes within the PM and keep them bound until the natural progression of the PM along the gut ensures elimination.
2. The lack of evidence of direct penetration through the tsetse PM by trypanosomes in the anterior midgut and the presence of trypanosomes in the ectoperitrophic space in the PV as early as 5 d.p.i. suggest an alternative explanation to the midgut cycle. It is proposed that the developmental cycle of trypanosomes in the tsetse midgut occurs entirely (or in parallel to the developmental cycle occurring in the ectoperitrophic space) within the PV before progression to the salivary glands (Fig. 6.1).

6.3 PM status and infection

The fact that older, starved flies have a similar chance of infection as teneral flies, may reflect the PM status at the time of infection. The rate of PM production slows down considerably in older flies and may be comparable to that of younger flies with an immature PM. It is proposed that a shorter PM in young flies may be a reason for increased susceptibility to trypanosome infection by prolonging the chance of trypanosomes being able to successfully breach the PM. However, the PM of an older fly extends throughout the entire gut and so the length has no bearing on infection status. Instead it is suggested that because the rate of production slows down, so too does the secretion and integration of crucial, structural PM components such as the peritrophins.

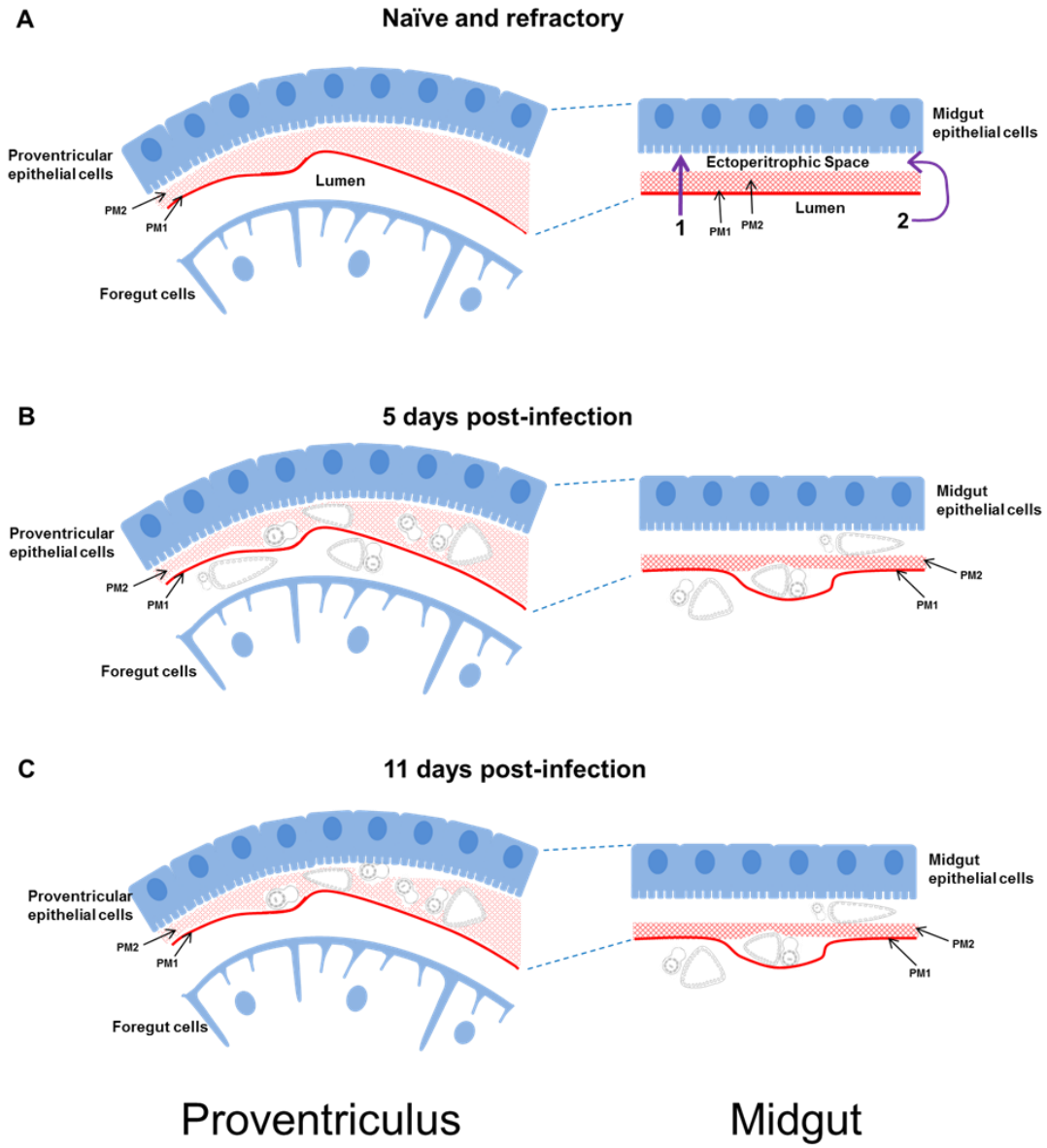


Figure 6.1. Model depicting an alternative development of trypanosoma brucei in the tsetse gut. A *represents the proventriculus (PV), (left hand side) and the midgut (right hand side) of a gut from a naïve or refractory fly. Specialised cells of the PV are responsible for the secretion of the peritrophic matrix (PM), where up to 3 layers are formed (PM1, PM2 and PM3). When initially secreted, PM2 is a fluid-like structure before maturation into a more rigid structure as seen in the midgut. Trypanosomes have been hypothesised to take one of two routes in the midgut in order to continue their lifecycle;* ***1****, direct penetration of the PM to reach the ectoperitrophic space (ES) or* ***2****, circumnavigation around the end of the PM in the hindgut.* ***B*** *depicts an alternative route whereby trypanosomes are able to complete their lifecycle in the PV. At 5 days post infection, parasites are able to penetrate the more fluid-like PM2 in the PV and as the PM matures the parasites are either trapped between PM1 and PM2, have successfully migrated to the ES or have not been successful in PM penetration and remain in the PV lumen. Those that have become trapped between PM1 and PM2 are carried through to the midgut during PM growth and maturation. They*

are probably unlikely to escape and are carried along the entire length of the gut until excreted with the PM. ***C*** *represents the alternative route at 11 days post infection. Trypanosomes are again either trapped between layers or are neatly contained in the ES, very rarely are they seen in the PV lumen. It is suggested that during a successful infection, trypanosomes colonise the PV and can undergo their entire gut lifecycle in this tissue, rather than or in parallel to that in the midgut.*

6.4 Proposed model of the type II PM in tsetse

With the identification of new PM proteins and a new proposed midgut development by *T. brucei*, this study has provided a greater understanding of the role of the peritrophic matrix throughout midgut infection. This warrants a new proposed model of the tsetse PM in regards to its structural status when immature or starved and also following maturity, and how this may affect the successful establishment of trypanosomes (Fig. 6.2). In similarities to previous studies, repetitive peritrophins are proposed to firmly hold the chitin bundles rigid seen as these possess multiple CBDs. As trypsin and chymotrypsin cleavage sites are common in such structures which form disulphide bridges, access to the sites by proteases are limited and so make a solid underlying network. An important feature of this model relates to the glycosylation of the peritrophins. The high degree of *O*-linked glycosylation in the IIMs, such as GmmPer66 and GmmPer108, serve to protect the PM from damage from physical abrasions and proteolysis [56]. In addition, due to their high likelihood of hydrophobicity, they are most likely situated on the 1st and 3rd layers of the PM where they aid water retention and may help to adhere trypanosomes for elimination. Those peritrophins containing *N*-glycosylation sites may be important in the formation of protein complexes. The core of an *N*-linked glycan contains two GlcNAc residues linked by β-(1, 4) bonds like those found in chitin, suggesting the GlcNAc disaccharide could be recognised by the CBD of other peritrophins [37]. The simple peritrophins serve to fill in the gaps after chitin assembly. It has previously been shown that Peritrophin-15, a simple peritrophin containing a single PAD, binds specifically and strongly to chitin [159]. Because binding occurs at a low stoichiometry and the protein is small, it is predicted that rather than have a role in cross-linking chitin fibrils Peritrophin-15 instead caps the end of individual chitin polymer chains. Similarly, peritrophins Pro1, Pro2 and GmmPer12 may be involved in protecting the chitin chains from exochitinases originating from either the tsetse or the bacterial symbionts. When immature or starved, the components of the PM are not fully assembled and so accessibility to the ectoperitrophic space via the PM in the PV is probably greatest at these time points.

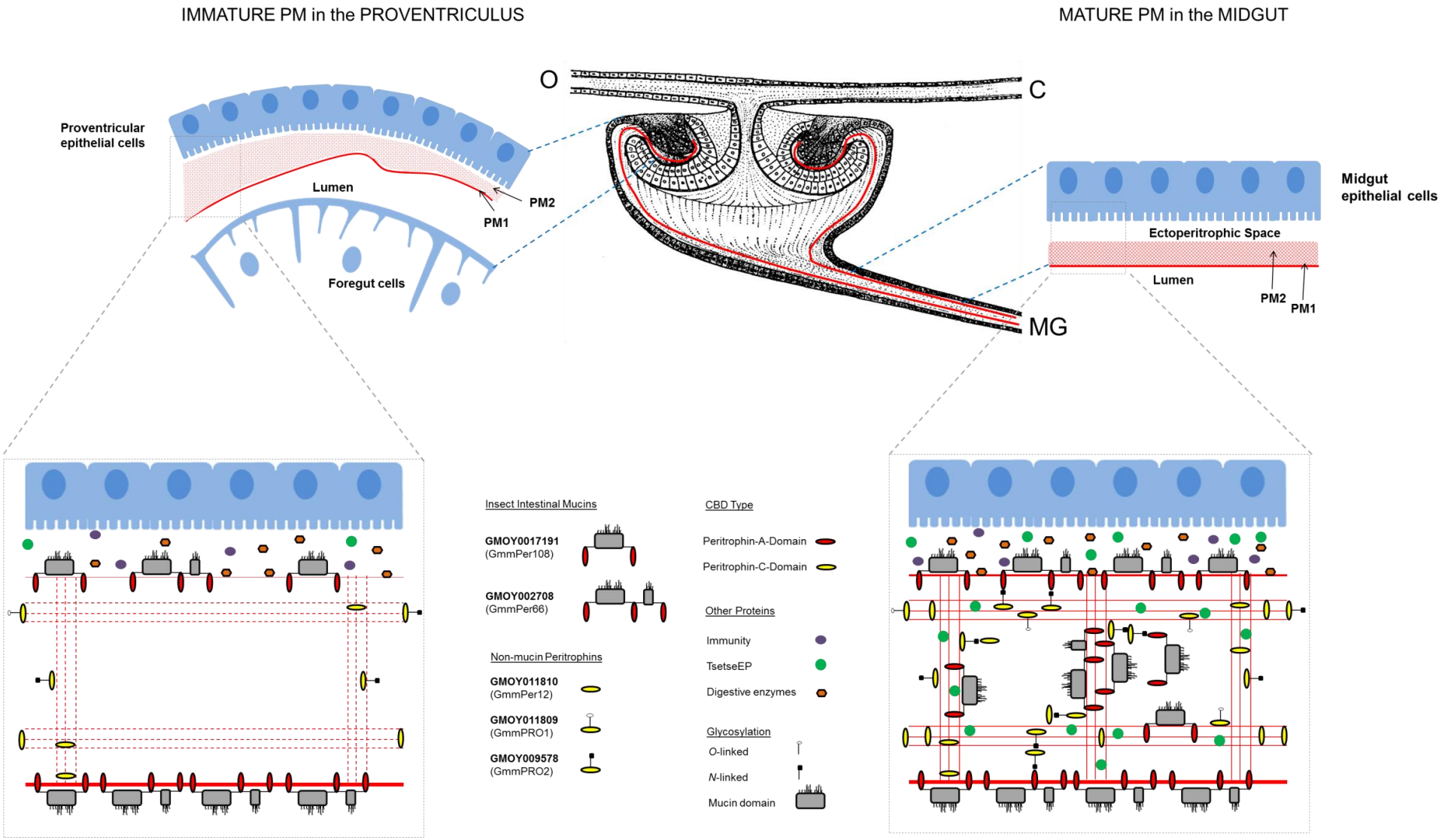

IMMATURE PM in the PROVENTRICULUS
MATURE PM in the MIDGUT
O
C
MG
Proventricular epithelial cells
Lumen
PM2
PM1
Foregut cells
Midgut epithelial cells
Ectoperitrophic Space
Lumen
PM2
PM1
Insect Intestinal Mucins
GMOY0017191 (GmmPer108)
GMOY002708 (GmmPer66)
Non-mucin Peritrophins
GMOY011810 (GmmPer12)
GMOY011809 (GmmPRO1)
GMOY009578 (GmmPRO2)
CBD Type
Peritrophin-A-Domain
Peritrophin-C-Domain
Other Proteins
Immunity
TsetseEP
Digestive enzymes
Glycosylation
O-linked
N-linked
Mucin domain

Figure 6.2. Model of the Glossina morsitans morsitans peritrophic matrix molecular architecture when (A) immature or starved or (B) mature *Chitin bundles consisting of multiple chitin chains assemble to form a scaffold. When freshly secreted in the PV, PM2 is fluid-like and few proteins are integrated at this point including peritrophins and tsetseEP. In a mature PM, PM2 is fully formed and compact. The mucin domains of the IIMs are portrayed with the O-linked glycans facing both the lumen and the ectoperitrophic space and also projecting into the interstitial space. Peritrophins containing N-linked glycans are shown interacting with the CBDs of other peritrophins via the GlcNAc dimer at the core of the glycan. TsetseEP is up-regulated in older, fed flies which correlates with a mature PM. TEM analysis shows PM1 is uniform in both immature and mature PMs and is depicted as such in the above model. Model based on type II PMs of Lepidopteran larvae and information on tsetse PM components and the type I PMs of adult Diptera, [35, 37, 49, 59, 61, 77, 78, 84, 118, 159, 192]*

6.5 Future Perspectives

Future investigations should include using RNAi to determine the roles of other peritrophins such as GMOY007191 and GMOY011810 in 1) the structural integrity of the PM and 2) the development of trypanosomes in the midgut. Furthermore, tissue expression of these peritrophins should be determined to see if they are expressed specifically in midgut tissues (including the PM, PV and crop) or if they are also expressed elsewhere in the fly. This would tell us if they are specifically peritrophic matrix proteins or if they belong to the cuticular proteins analogous to peritrophins (CPAP) family and would allow us to make a more precise model of PM structure. Localisation of the peritrophins in the immature/starved and mature PM in both the proventriculus and midgut would further our understanding of how the type II PM is synthesised and what bearing this has on trypanosome development. For example, the use of fluorescent in situ hybridisation (FISH) would allow for the specific localisation of peritrophins in the PM both within the PV and also the midgut in situ. Although the related techniques such as Northern blotting, subtractive cDNA library screening and RT-PCR are more sensitive in comparison, they are much better suited for analysing transcript levels. Immunogold and immunofluorescence studies may also be used to determine the distribution of the peritrophins within the PM using specific antibodies.

Analysis of transcript levels of peritrophins at time points throughout trypanosome infection may additionally provide insight into the differential expression of peritrophins in response to infection. Additional analyses should be made in flies collected from the field as these may elicit a stronger peritrophin expression response in comparison to laboratory strains which

have been maintained for several years and may have lost the capacity to strongly react to gut pathogens as seen in previous studies [220].

Annotation of the *Stomoxys calcitrans* genome is almost complete and so it would be worthwhile to compare the number and type of peritrophins to those identified in this study (chaoter 3). This would provide more information about the role of the PM in regards to the peritrophin function. In addition, phylogenetic analysis should be carried out to compare between insect species that secrete a type 1 PM, such as mosquito and sandfly species, to those insects that were analysed in this study, all of which secrete a type 2 PM. Differences in the peritrophins between larval and adult stages of each insect should also be taken into account; this comparison would be particularly useful when looking at insects who have a larval type 2 PM but then switch to producing a type 1 PM as an adult. Insects that secrete a type 1 PM as an adult are usually far more permissive disease vectors. Differences in peritrophin structure, spatial distribution in the PM and time frame of secretion into the PM should be investigated for both type 1 and 2 PMs in both larvae and adult.Through publication of the *Glossina* genome, it was determined that *G. morsitans* contains more aquaporins than any other insect so far analysed [221]. As these genes are involved in diuresis in the fly and this study suggests diuresis may be important in potentiating infection in the fly, it would be worth knocking these genes down using RNAi to assess the role, if any, impaired diuresis has on the progression of trypanosome development.

Further TEM and 3D reconstruction analyses should also be carried out on the PV to try and pin-point the breaching event by trypanosomes in this tissue. Moreover, the time point of this analysis should start earlier than 5 days to determine the exact window of opportunity (time wise) of trypanosome clearance or establishment in the fly.

To help determine if the hypothesis of an alternative lifecycle in the fly is correct, developmental stages of the trypanosomes should be compared between the PV and midgut at different time points during infection. Trypanosomes ingested with a bloodmeal quickly express procyclin molecules during transformation into procyclics. Initially they express both EP and GPEET procyclins before predominantly expressing GPEET during an early infection. As infection progresses, the glycosylated EP isoforms EP1 and EP3 are expressed and GPEET is supressed. Antibodies to GPEET, unglycosylated EP and glycosylated EP have been used previously to distinguish between early and late procyclics [114, 222]. Flies should be infected with 100% stumpy forms (as these do not express procyclins) and therefore anti-GPEET antibodys will detect only those parasites that have undergone transformation. To demonstrate they do not express procyclin, stumpy forms should be stained with anti-procyclin antibodies. PVs and midguts should be dissected and immediately separated from 4+ hrs following an infected bloodmeal and trypanosomes pooled before staining with each antibody. Dissections should be carried out every 2 days for a period of 2

weeks and in parallel, an in vitro staining assay should be carried out to compare before and after transformation with cis-aconitate.

In summary, this work has provided useful insights into the tsetse PM, both at a structural and functional level and may contribute to novel control strategies which target the vector and the development of salivarian trypanosomes within the midgut, including those responsible for HAT. The insect PM is an attractive target as it not only aids digestion and nutrient uptake in the fly but because of its position is also the first point of contact to ingested compounds so is suitable for *per os* control applications. Although multiple PM components have been characterised here there are still many potential targets that warrant further investigation, namely the peritrophins.

References

1. Bentivoglio, M., et al., *Overview of Neglected Tropical Diseases and Conditions of the Nervous System: Past, Present and Perspectives*, in *Neglected Tropical Diseases and Conditions of the Nervous System*, M. Bentivoglio, et al., Editors. 2014, Springer New York. p. 3-19.
2. W.H.O. *Trypanosomiasis, human African (sleeping sickness)*. 2014 March 2014 [cited April 2014]; Fact Sheet No. 259:[Available from: http://www.who.int/mediacentre/factsheets/fs259/en/.
3. Barrett, M.P., et al., *The trypanosomiases.* Lancet, 2003. **362**(9394): p. 1469-1480.
4. Simarro, P.P., et al., *Estimating and Mapping the Population at Risk of Sleeping Sickness.* Plos Neglected Tropical Diseases, 2012. **6**(10).
5. Lozano, R., et al., *Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010.* The Lancet, 2012. **380**(9859): p. 2095-2128.
6. Malvy, D. and F. Chappuis, *Sleeping sickness.* Clin Microbiol Infect, 2011. **17**(7): p. 986-95.
7. Brun, R., et al., *Human African trypanosomiasis.* Lancet, 2010. **375**(9709): p. 148-159.
8. Pentreath, V.W., *Royal Society of Tropical Medicine and Hygiene Meeting at Manson House, London, 19 May 1994. Trypanosomiasis and the nervous system. Pathology and immunology.* Trans R Soc Trop Med Hyg, 1995. **89**(1): p. 9-15.
9. MacLean, L.M., et al., *Focus-specific clinical profiles in human African Trypanosomiasis caused by Trypanosoma brucei rhodesiense.* PLoS Negl Trop Dis, 2010. **4**(12): p. e906.
10. Hackett, F., et al., *Incorporating scale dependence in disease burden estimates: the case of human African trypanosomiasis in Uganda.* PLoS Negl Trop Dis, 2014. **8**(2): p. e2704.
11. Chamond, N., et al., *<italic>Trypanosoma vivax</italic> Infections: Pushing Ahead with Mouse Models for the Study of <italic>Nagana</italic>. I. Parasitological, Hematological and Pathological Parameters.* PLoS Negl Trop Dis, 2010. **4**(8): p. e792.
12. Matovu, E., et al., *Drug resistance in Trypanosoma brucei spp., the causative agents of sleeping sickness in man and nagana in cattle.* Microbes and Infection, 2001. **3**(9): p. 763-770.
13. Benoit, J.B., et al., *Sphingomyelinase activity in mother's milk is essential for juvenile development: a case from lactating tsetse flies.* Biol Reprod, 2012. **87**(1): p. 17, 1-10.
14. Barry, J.D. and R. McCulloch, *Antigenic variation in trypanosomes: Enhanced phenotypic variation in a eukaryotic parasite*, in *Advances in Parasitology, Vol 49*. 2001. p. 1-70.
15. Cross, G.A., H.S. Kim, and B. Wickstead, *Capturing the variant surface glycoprotein repertoire (the VSGnome) of Trypanosoma brucei Lister 427.* Mol Biochem Parasitol, 2014. **195**(1): p. 59-73.
16. Alsford, S., et al., *Genetic dissection of drug resistance in trypanosomes.* Parasitology, 2013. **140**(12): p. 1478-91.
17. Baker, N., et al., *Drug resistance in African trypanosomiasis: the melarsoprol and pentamidine story.* Trends Parasitol, 2013. **29**(3): p. 110-8.
18. Vickerman, K., *Developmental cycles and biology of pathogenic trypanosomes. .* British Medical Bulletin 1985. **41**: p. 105-114.
19. Verberg, M.F., et al., *Hyperemesis gravidarum, a literature review.* Hum Reprod Update, 2005. **11**(5): p. 527-39.
20. Knüsel, S. and I. Roditi, *Insights into the regulation of GPEET procyclin during differentiation from early to late procyclic forms of Trypanosoma brucei.* Molecular and biochemical parasitology, 2013. **191**(2): p. 66-74.
21. Ferrero, S., et al., *Haptoglobin beta chain isoforms in the plasma and peritoneal fluid of women with endometriosis.* Fertil Steril, 2005. **83**(5): p. 1536-43.
22. Ferrero, S., et al., *Vitamin D binding protein in endometriosis.* J Soc Gynecol Investig, 2005. **12**(4): p. 272-7.

23. Caljon, G., L. Vooght, and J. Abbeele, *The Biology of Tsetse–Trypanosome Interactions*, in *Trypanosomes and Trypanosomiasis*, S. Magez and M. Radwanska, Editors. 2014, Springer Vienna: Vienna. p. 41-59.
24. Oberholzer, M., et al., *Social motility in African trypanosomes.* PLoS Pathog, 2010. **6**(1): p. e1000739.
25. Imhof, S., et al., *Social motility of African trypanosomes is a property of a distinct life-cycle stage that occurs early in tsetse fly transmission.* PLoS Pathog, 2014. **10**(10): p. e1004493.
26. Imhof, S., et al., *A Glycosylation Mutant of Trypanosoma brucei Links Social Motility Defects In Vitro to Impaired Colonization of Tsetse Flies In Vivo.* Eukaryotic cell, 2015. **14**(6): p. 588-592.
27. *Edited by*, in *Advances in Insect Physiology*, P.D. Evans and V.B. Wigglesworth, Editors. 1987, Academic Press. p. iii.
28. Dow, J.A.T., *Insect Midgut Function*, in *Advances in Insect Physiology*, P.D. Evans and V.B. Wigglesworth, Editors. 1987, Academic Press. p. 187-328.
29. Chapman, R.F., *The insects: structure and function*. 1998: Cambridge university press.
30. Capinera, J.L., *Encyclopedia of entomology*. Vol. 4. 2008: Springer Science & Business Media.
31. Lyonet, P., *Traite anatomique de la chenille, qui ronge le bois de saul, augmente d'une explication des planches*. 1762: Gosse et Pinet.
32. Brandt, C.R., M.J. Adang, and K.D. Spence, *The peritrophic membrane: ultrastructural analysis and function as a mechanical barrier to microbial infection in Orgyia pseudotsugata.* Journal of Invertebrate Pathology, 1978. **32**(1): p. 12-24.
33. Peters, W., *Peritrophic membranes*. 1992, Berlin ; London: Springer-Verlag. xi, 238 p.
34. Engel, P. and N.A. Moran, *The gut microbiota of insects - diversity in structure and function.* FEMS Microbiol Rev, 2013. **37**(5): p. 699-735.
35. Lehane, M.J., *Peritrophic matrix structure and function.* Annual Review of Entomology, 1997. **42**: p. 525-550.
36. Tellam, R., *The peritrophic matrix*, in *The insect midgut*, M.J. Lehane and P.F. Billingsley, Editors. 1996, Chapman and Hall: London.
37. Shi, X., et al., *Modeling the structure of the type I peritrophic matrix: characterization of a Mamestra configurata intestinal mucin and a novel peritrophin containing 19 chitin binding domains.* Insect Biochem Mol Biol, 2004. **34**(10): p. 1101-15.
38. Lehane, M.J. and P.A. Billingsley, eds. *The Biology of the Insect Midgut.* 1996, Chapman and Hall: London. 480.
39. Lehane, M.J., P.G. Allingham, and P. Weglicki, *Composition of the peritrophic matrix of the tsetse fly, Glossina morsitans morsitans.* Cell and Tissue Research, 1996. **283**(3): p. 375-384.
40. Peters, W., *Peritrophic membranes*. 1st ed. 1992, Berlin: Springer.
41. Hegedus, D., et al., *New insights into peritrophic matrix synthesis, architecture, and function.* Annu Rev Entomol, 2009. **54**: p. 285-302.
42. Becker, B., *Effects of polyoxin D on in vitro synthesis of peritrophic membranes in Calliphora erythrocephala.* Insect Biochem., 1980. **10**: p. 101-106.
43. Toprak, U., M. Erlandson, and D. Hegedus, *Peritrophic matrix proteins.* Trends in Entomology, 2010. **6**: p. 23-51.
44. Tellam, R.L. and C. Eisemann, *Chitin is only a minor component of the peritrophic matrix from larvae of Lucilia cuprina.* Insect Biochemistry And Molecular Biology, 2000. **30**(12): p. 1189-1201.
45. Huber, H., E. Cabib, and L.H. Miller, *Malaria parasite chitinase and penetration of the mosquito peritrophic membrane.* Proceedings of National Academy of Science,USA, 1991. **88**: p. 2807-2810.
46. Rogers, M.E., et al., *Leishmania chitinase facilitates colonization of sand fly vectors and enhances transmission to mice.* Cell Microbiol, 2008. **10**(6): p. 1363-72.
47. Berriman, M., et al., *The Genome of the African Trypanosome Trypanosoma brucei.* Science, 2005. **309**(5733): p. 416-422.

48. Jackson, A.P., et al., *The Genome Sequence of <italic>Trypanosoma brucei gambiense</italic>, Causative Agent of Chronic Human African Trypanosomiasis.* PLoS Negl Trop Dis, 2010. **4**(4): p. e658.
49. Tellam, R.L., G. Wijffels, and P. Willadsen, *Peritrophic matrix proteins.* Insect Biochem Mol Biol, 1999. **29**(2): p. 87-101.
50. Shen, Z. and M. Jacobs-Lorena, *Evolution of chitin-binding proteins in invertebrates.* J Mol Evol, 1999. **48**(3): p. 341-7.
51. Raikhel, N.V., H.I. Lee, and W.F. Broekaert, *Structure and function of chitin-binding proteins.* Annual review of plant biology, 1993. **44**(1): p. 591-615.
52. Ross, J., et al., *Serine proteases and their homologs in the Drosophila melanogaster genome: an initial analysis of sequence conservation and phylogenetic relationships.* Gene, 2003. **304**: p. 117-131.
53. Willis, J.H., *Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era.* Insect biochemistry and molecular biology, 2010. **40**(3): p. 189-204.
54. Shao, L., M. Devenport, and M. Jacobs-Lorena, *The peritrophic matrix of hematophagous insects.* Archives of insect biochemistry and physiology, 2001. **47**(2): p. 119-125.
55. Marquardt, W.H., *Biology of disease vectors*. 2004: Academic Press.
56. Wang, P. and R.R. Granados, *An intestinal mucin is the target substrate for a baculovirus enhancin.* Proceedings of the National Academy of Sciences, 1997. **94**(13): p. 6977-6982.
57. Zhang, X. and W. Guo, *Isolation and identification of insect intestinal mucin HaIIM86-the new target for Helicoverpa armigera biocontrol.* Int J Biol Sci, 2011. **7**(3): p. 286-296.
58. Van Klinken, B.J., et al., *Mucin gene structure and expression: protection vs. adhesion.* American Journal of Physiology-Gastrointestinal and Liver Physiology, 1995. **269**(5): p. G613-G627.
59. Rose, C., et al., *An investigation into the protein composition of the teneral Glossina morsitans morsitans peritrophic matrix.* PLoS Negl Trop Dis, 2014. **8**(4): p. e2691.
60. Waterhouse, D.F., *The occurrence and significance of the pertirophic membrane, with special reference to adult lepidoptera and diptera.* Australian Journal of Zoology, 1953. **1**(3): p. 299-318.
61. Miller, N., Lehane, M., J. , *In vitro perfusion studies on the peritrophic membrane of the tsetse fly Glossina morsitans morsitans (Diptera: Glossinidae).* J. Insect Physiol., 1990. **36**: p. 813-18.
62. Dow, J.A.T. and M.J. O'Donnell, *Reversible Alkalinization by Manduca Sexta Midgut.* Journal of Experimental Biology, 1990. **150**(1): p. 247-256.
63. Spence, K.D. and M.Y. Kawata, *Permeability characteristics of the peritrophic membranes of Manduca sexta larvae.* Journal of insect physiology, 1993. **39**(9): p. 785-790.
64. Pruzinova, K., et al., *Comparison of Bloodmeal Digestion and the Peritrophic Matrix in Four Sand Fly Species Differing in Susceptibility to <italic>Leishmania donovani</italic>.* PLoS One, 2015. **10**(6): p. e0128203.
65. Agrawal, S., et al., *Two essential peritrophic matrix proteins mediate matrix barrier functions in the insect midgut.* Insect Biochem Mol Biol, 2014. **49**: p. 24-34.
66. Barbehenn, R.V., *Roles of peritrophic membranes in protecting herbivorous insects from ingested plant allelochemicals.* Arch Insect Biochem Physiol, 2001. **47**(2): p. 86-99.
67. Rayms-Keller, A., et al., *Molecular cloning and characterization of a metal responsive Aedes aegypti intestinal mucin cDNA.* Insect Mol Biol, 2000. **9**(4): p. 419-26.
68. Pascoa, V., et al., *Aedes aegypti peritrophic matrix and its interaction with heme during blood digestion.* Insect Biochem Mol Biol, 2002. **32**(5): p. 517-23.
69. Devenport, M., et al., *Identification of the Aedes aegypti peritrophic matrix protein AeIMUCI as a heme-binding protein.* Biochemistry, 2006. **45**(31): p. 9540-9549.
70. Kuraishi, T., et al., *Genetic evidence for a protective role of the peritrophic matrix against intestinal bacterial infection in Drosophila melanogaster.* Proc Natl Acad Sci U S A, 2011. **108**(38): p. 15966-71.

71. Weiss, B.L., et al., *The peritrophic matrix mediates differential infection outcomes in the tsetse fly gut following challenge with commensal, pathogenic, and parasitic microbes.* J Immunol, 2014. **193**(2): p. 773-82.
72. Welburn, S.C., et al., *Sleeping sickness: a tale of two diseases.* Trends Parasitol, 2001. **17**(1): p. 19-24.
73. Baker, N., et al., *Drug resistance in African trypanosomiasis: the melarsoprol and pentamidine story.* Trends in Parasitology, 2013. **29**(3): p. 110-118.
74. MacLeod, E.T., et al., *Antioxidants promote establishment of trypanosome infections in tsetse.* Parasitology, 2007. **134**: p. 827-831.
75. Harrington, J.M., *Antimicrobial peptide killing of African trypanosomes.* Parasite Immunology, 2011. **33**(8): p. 461-469.
76. Bullard, W., et al., *Haptoglobin-hemoglobin receptor independent killing of African trypanosomes by human serum and trypanosome lytic factors.* Virulence, 2012. **3**(1): p. 72-6.
77. Hegedus, D., et al., *New Insights into Peritrophic Matrix Synthesis, Architecture, and Function.* Annual Review of Entomology, 2009. **54**: p. 285-302.
78. Lehane, M.J., P.G. Allingham, and P. Weglicki, *Composition of the peritrophic matrix of the tsetse fly, Glossina morsitans morsitans.* Cell Tissue Res, 1996. **283**(3): p. 375-84.
79. Coutinho-Abreu, I.V., et al., *Characterization of Phlebotomus papatasi peritrophins, and the role of PpPer1 in Leishmania major survival in its natural vector.* PLoS Negl Trop Dis, 2013. **7**(3): p. e2132.
80. Sieber, K.P., et al., *The peritrophic membrane as a barrier. Its penetration by Plasmodium gallinaceum and the effect of a monoclonal antibody to ookinetes.* Experimental Parasitology, 1991. **72**: p. 145-156.
81. Sutherland, D.R., B.M. Christensen, and B.A. Lasee, *Midgut barrier as a possible factor in filarial worm vector competency in Aedes trivittatus.* J. Invert. Path., 1986. **47**(4): p. 1-7.
82. Kato, N., et al., *Evaluation of the function of a type I peritrophic matrix as a physical barrier for midgut epithelium invasion by mosquito-borne pathogens in Aedes aegypti.* Vector Borne Zoonotic Dis, 2008. **8**(5): p. 701-12.
83. Weiss, B.L., et al., *Trypanosome infection establishment in the tsetse fly gut is influenced by microbiome-regulated host immune barriers.* PLoS Pathog, 2013. **9**(4): p. e1003318.
84. Lehane, M.J. and A.R. Msangi, *Lectin and peritrophic membrane development in the gut of Glossina m.morsitans and a discussion of their role in protecting the fly against trypanosome infection.* Med Vet Entomol, 1991. **5**(4): p. 495-501.
85. Hao, Z. and S. Aksoy, *Proventriculus-specific cDNAs characterized from the tsetse, Glossina morsitans morsitans.* Insect Biochem Mol Biol, 2002. **32**(12): p. 1663-71.
86. Haddow, J.D., et al., *Identification of midgut proteins that are differentially expressed in trypanosome-susceptible and normal tsetse flies (Glossina morsitans morsitans).* Insect Biochemistry And Molecular Biology, 2005. **35**(5): p. 425-433.
87. Terra, W.R., *The origin and functions of the insect peritrophic membrane and peritrophic gel.* Arch Insect Biochem Physiol, 2001. **47**(2): p. 47-61.
88. Walshe, D.P., M.J. Lehane, and L.R. Haines, *Post eclosion age predicts the prevalence of midgut trypanosome infections in Glossina.* PLoS One, 2011. **6**(11): p. e26984.
89. Ellis, D.S. and D.A. Evans, *Passage of Trypanosoma brucei rhodesiense through the peritrophic membrane of Glossina morsitans morsitans.* Nature, 1977. **267**(5614): p. 834-5.
90. Evans, D.A. and D.S. Ellis, *Recent observations on the behaviour of certain trypanosomes within their insect hosts.* Adv Parasitol, 1983. **22**: p. 1-42.
91. Aksoy, S., W.C. Gibson, and M.J. Lehane, *Interactions between tsetse and trypanosomes with implications for the control of trypanosomiasis.* Adv Parasitol, 2003. **53**: p. 1-83.
92. Ghosh, A.K., et al., *Plasmodium ookinetes coopt mammalian plasminogen to invade the mosquito midgut.* Proc Natl Acad Sci U S A, 2011. **108**(41): p. 17153-8.
93. Link, A.J. and J. LaBaer, *Trichloroacetic acid (TCA) precipitation of proteins.* Cold Spring Harb Protoc, 2011. **2011**(8): p. 993-4.

94. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-5.
95. Neuhoff, V., et al., *Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250.* Electrophoresis, 1988. **9**(6): p. 255-62.
96. Belda, E., et al., *Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of Sodalis glossinidius, the secondary endosymbiont of tsetse flies.* BMC Genomics, 2010. **11**: p. 449.
97. Acosta-Serrano, A., et al., *The procyclin repertoire of Trypanosoma brucei - Identification and structural characterization of the Glu-Pro-rich polypeptides.* Journal Of Biological Chemistry, 1999. **274**(42): p. 29763-29771.
98. Campbell, P.M., et al., *Proteomic analysis of the peritrophic matrix from the gut of the caterpillar, Helicoverpa armigera.* Insect Biochem Mol Biol, 2008. **38**(10): p. 950-8.
99. Goldberg, H.A. and K.J. Warner, *The staining of acidic proteins on polyacrylamide gels: enhanced sensitivity and stability of "Stains-all" staining in combination with silver nitrate.* Anal Biochem, 1997. **251**(2): p. 227-33.
100. Haines, L.R., et al., *Tsetse EP protein protects the fly midgut from trypanosome establishment.* PLoS Pathog, 2010. **6**(3): p. e1000793.
101. Haines, L.R., et al., *Increased expression of unusual EP repeat-containing proteins in the midgut of the tsetse fly (Glossina) after bacterial challenge.* Insect Biochemistry And Molecular Biology, 2005. **35**(5): p. 413-423.
102. Haines, L.R., et al., *Increased expression of unusual EP repeat-containing proteins in the midgut of the tsetse fly (Glossina) after bacterial challenge.* Insect Biochem Mol Biol, 2005. **35**(5): p. 413-23.
103. Toprak, U., Erlandson, M., Hegedus, D., D., *Peritrophic matrix proteins.* Trends in Entomology, 2010. **6**: p. 23-51.
104. Wang, P., G. Li, and R.R. Granados, *Identification of two new peritrophic membrane proteins from larval Trichoplusia ni: structural characteristics and their functions in the protease rich insect gut.* Insect Biochem Mol Biol, 2004. **34**(3): p. 215-27.
105. Eisemann, C., G. Wijffels, and R.L. Tellam, *Secretion of the type 2 peritrophic matrix protein, peritrophin-15, from the cardia.* Archives of Insect Biochemistry and Physiology, 2001. **47**(2): p. 76-85.
106. Mullens, B.A. and M.J. Lehane, *Fluorescence as a tool for age determination in Culicoides variipennis sonorensis (Diptera: Ceratopogonidae).* J Med Entomol, 1995. **32**(4): p. 569-71.
107. Shen, Z. and M. Jacobs-Lorena, *A type I peritrophic matrix protein from the malaria vector Anopheles gambiae binds to chitin. Cloning, expression, and characterization.* J Biol Chem, 1998. **273**(28): p. 17665-70.
108. Dimopoulos, G., et al., *Malaria infection of the mosquito Anopheles gambiae activates immune-responsive genes during critical transition stages of the parasite life cycle.* EMBO Journal, 1998. **17**(21): p. 6115-6123.
109. Waterhouse, D.F., *The rate of production of the peritrophic membrane in some insects.* Aust J Biol Sci., 1954. **7**: p. 59-72.
110. Zhang, D., et al., *Functional analysis of two chitinase genes during the pupation and eclosion stages of the beet armyworm Spodoptera exigua by RNA interference.* Arch Insect Biochem Physiol, 2012. **79**(4-5): p. 220-34.
111. Zhang, J., et al., *Comparative genomic analysis of chitinase and chitinase-like genes in the African malaria mosquito (Anopheles gambiae).* PLoS One, 2011. **6**(5): p. e19899.
112. Merzendorfer, H. and L. Zimoch, *Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases.* J Exp Biol, 2003. **206**(Pt 24): p. 4393-412.
113. Bolognesi, R., W.R. Terra, and C. Ferreira, *Peritrophic membrane role in enhancing digestive efficiency. Theoretical and experimental models.* J Insect Physiol, 2008. **54**(10-11): p. 1413-22.

114. Acosta-Serrano, A., et al., *The surface coat of procyclic Trypanosoma brucei: programmed expression and proteolytic cleavage of procyclin in the tsetse fly.* Proc Natl Acad Sci U S A, 2001. **98**(4): p. 1513-8.
115. Natori, S., *Insect lectins and innate immunity.* Adv Exp Med Biol, 2001. **484**: p. 223-8.
116. Welburn, S.C., I. Maudlin, and D.S. Ellis, *Rate of trypanosome killing by lectins in midguts of different species and strains of Glossina.* Medical and Veterinary Entomology, 1989. **3**(1): p. 77-82.
117. Iozzo, R.V., *Perlecan: a gem of a proteoglycan.* Matrix Biol., 1994. **14**(3): p. 203-208.
118. Dinglasan, R.R., et al., *The Anopheles gambiae adult midgut peritrophic matrix proteome.* Insect Biochem Mol Biol, 2009. **39**(2): p. 125-34.
119. Hu, X., et al., *Proteomic analysis of peritrophic membrane (PM) from the midgut of fifth-instar larvae, Bombyx mori.* Mol Biol Rep, 2012. **39**(4): p. 3427-34.
120. Gibbs, G.M., K. Roelants, and M.K. O'Bryan, *The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins--roles in reproduction, cancer, and immune defense.* Endocr Rev, 2008. **29**(7): p. 865-97.
121. Fang, K.S., et al., *cDNA cloning and primary structure of a white-face hornet venom allergen, antigen 5.* Proc Natl Acad Sci U S A, 1988. **85**(3): p. 895-9.
122. Hoffman, D.R., *Fire ant venom allergy.* Allergy, 1995. **50**(7): p. 535-44.
123. Li, S., J. Kwon, and S. Aksoy, *Characterization of genes expressed in the salivary glands of the tsetse fly, Glossina morsitans morsitans.* Insect Molecular Biology, 2001. **10**(1): p. 69-76.
124. Kovalick, G.E., et al., *Structure and expression of the antigen 5-related gene of Drosophila melanogaster.* Insect Biochem Mol Biol, 1998. **28**(7): p. 491-500.
125. Megraw, T., T.C. Kaufman, and G.E. Kovalick, *Sequence and expression of Drosophila Antigen 5-related 2, a new member of the CAP gene family.* Gene, 1998. **222**(2): p. 297-304.
126. Assumpcao, T.C., et al., *Salivary antigen-5/CAP family members are Cu2+-dependent antioxidant enzymes that scavenge O(2)(-). and inhibit collagen-induced platelet aggregation and neutrophil oxidative burst.* J Biol Chem, 2013. **288**(20): p. 14341-61.
127. Janssen, K.P., et al., *Characterization of CD36/LIMPII homologues in Dictyostelium discoideum.* J Biol Chem, 2001. **276**(42): p. 38899-910.
128. Silverstein, R.L. and M. Febbraio, *CD36, a scavenger receptor involved in immunity, metabolism, angiogenesis, and behavior.* Sci Signal, 2009. **2**(72): p. re3.
129. Means, T.K., et al., *Evolutionarily conserved recognition and innate immunity to fungal pathogens by the scavenger receptors SCARF1 and CD36.* J Exp Med, 2009. **206**(3): p. 637-53.
130. Jouvenaz, D.P., M.S. Blum, and C.J. Mac, *Antibacterial activity of venom alkaloids from the imported fire ant, Solenopsis invicta Buren.* Antimicrob Agents Chemother, 1972. **2**(4): p. 291-3.
131. Theopold, U., et al., *Helix pomatia lectin, an inducer of Drosophila immune response, binds to hemomucin, a novel surface mucin.* Journal of Biological Chemistry, 1996. **271**(22): p. 12708-12715.
132. Graca-Souza, A.V., et al., *Adaptations against heme toxicity in blood-feeding arthropods.* Insect Biochem Mol Biol, 2006. **36**(4): p. 322-35.
133. Aksoy, S. and R.V.M. Rio, *Interactions among multiple genomes: Tsetse, its symbionts and trypanosomes.* Insect Biochemistry And Molecular Biology, 2005. **35**(7): p. 691-698.
134. Akman, L., et al., *Genome size determination and coding capacity of Sodalis glossinidius, an enteric symbiont of tsetse flies, as revealed by hybridization to Escherichia coli gene arrays.* J Bacteriol, 2001. **183**(15): p. 4517-25.
135. Welburn, S.C., et al., *Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse-flies.* Parasitology, 1993. **107**(Pt2): p. 141-145.
136. Welburn, S.C. and I. Maudlin, *Rickettsia-Like Organisms, Puparial Temperature and Susceptibility to Trypanosome Infection in Glossina-Morsitans.* Parasitology, 1991. **102**: p. 201-206.

137. Welburn, S.C. and I. Maudlin, *Tsetse-typanosome interactions: Rites of passage.* Parasitology Today, 1999. **15**(10): p. 399-403.
138. Dale, C. and S.C. Welburn, *The endosymbionts of tsetse flies: manipulating host-parasite interactions.* International Journal For Parasitology, 2001. **31**(5-6): p. 628-631.
139. Maudlin, I. and D.S. Ellis, *Association between intracellular rickettsial-like infections of midgut cells and susceptibility to trypanosome infection in Glossina spp.* Z Parasitenkd, 1985. **71**(5): p. 683-7.
140. Dyer, N.A., et al., *Flying tryps: survival and maturation of trypanosomes in tsetse flies.* Trends Parasitol, 2013. **29**(4): p. 188-96.
141. Jariyapan, N., et al., *Peritrophic matrix formation and Brugia malayi microfilaria invasion of the midgut of a susceptible vector, Ochlerotatus togoi (Diptera: Culicidae).* Parasitol Res, 2013. **112**(7): p. 2431-40.
142. Huber, M., E. Cabib, and L.H. Miller, *Malaria parasite chitinase and penetration of the mosquito peritrophic membrane.* Proc Natl Acad Sci U S A, 1991. **88**(7): p. 2807-10.
143. Tellam, R.L. and C. Eisemann, *Chitin is only a minor component of the peritrophic matrix from larvae of Lucilia cuprina.* Insect Biochem Mol Biol, 2000. **30**(12): p. 1189-201.
144. International Glossina Genome, I., *Genome sequence of the tsetse fly (Glossina morsitans): vector of African trypanosomiasis.* Science, 2014. **344**(6182): p. 380-6.
145. Vizcaino, J.A., et al., *The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013.* Nucleic Acids Res, 2013. **41**(Database issue): p. D1063-9.
146. Budd, L.T. and G.B.D.f.I. Development, *DFID-funded Tsetse and Trypanosome Research and Development Since 1980: Economic analysis*. 1999: DFID.
147. Horn, D., *Antigenic variation in African trypanosomes.* Mol Biochem Parasitol, 2014. **195**(2): p. 123-9.
148. Unciti-Broceta, J.D., et al., *Specific Cell Targeting Therapy Bypasses Drug Resistance Mechanisms in African Trypanosomiasis.* PLoS Pathog, 2015. **11**(6): p. e1004942.
149. PATTEC, A. Available from; http://rea.au.int/en/RO/PATTEC 2014.
150. Obiero, G.F., et al., *Odorant and gustatory receptors in the tsetse fly Glossina morsitans morsitans.* PLoS Negl Trop Dis, 2014. **8**(4): p. e2663.
151. Benoit, J.B., et al., *A novel highly divergent protein family identified from a viviparous insect by RNA-seq analysis: a potential target for tsetse fly-specific abortifacients.* PLoS Genet, 2014. **10**(4): p. e1003874.
152. Attardo, G.M., et al., *The homeodomain protein ladybird late regulates synthesis of milk proteins during pregnancy in the tsetse fly (Glossina morsitans).* PLoS Negl Trop Dis, 2014. **8**(4): p. e2645.
153. Telleria, E.L., et al., *Insights into the trypanosome-host interactions revealed through transcriptomic analysis of parasitized tsetse fly salivary glands.* PLoS Negl Trop Dis, 2014. **8**(4): p. e2649.
154. Benoit, J.B., et al., *Aquaporins are critical for provision of water during lactation and intrauterine progeny hydration to maintain tsetse fly reproductive success.* PLoS Negl Trop Dis, 2014. **8**(4): p. e2517.
155. Dereeper, A., et al., *Phylogeny.fr: robust phylogenetic analysis for the non-specialist.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W465-9.
156. Dereeper, A., et al., *BLAST-EXPLORER helps you building datasets for phylogenetic analysis.* BMC Evol Biol, 2010. **10**: p. 8.
157. Chevenet, F., et al., *TreeDyn: towards dynamic graphics and annotations for analyses of trees.* BMC Bioinformatics, 2006. **7**: p. 439.
158. Tellam, R.L., et al., *Identification of an immuno-protective mucin-like protein, peritrophin-55, from the peritrophic matrix of Lucilia cuprina larvae.* Insect Biochem Mol Biol, 2003. **33**(2): p. 239-52.

159. Wijffels, G., et al., *A novel family of chitin-binding proteins from insect type 2 peritrophic matrix. cDNA sequences, chitin binding activity, and cellular localization.* J Biol Chem, 2001. **276**(18): p. 15527-36.
160. Baldacchino, F., et al., *Transmission of pathogens by Stomoxys flies (Diptera, Muscidae): a review.* Parasite, 2013. **20**: p. 26.
161. Watanabe, J., et al., *Genome Sequence of the Tsetse Fly (Glossina morsitans): Vector of African Trypanosomiasis.* Science, 2014. **344**(6182): p. 380-386.
162. Petersen, F.T., et al., *The phylogeny and evolution of host choice in the Hippoboscoidea (Diptera) as reconstructed using four molecular markers.* Mol Phylogenet Evol, 2007. **45**(1): p. 111-22.
163. Mohamed-Ahmed, M. and A. Odulaja, *Diel activity patterns and host preferences of Glossina fuscipes fuscipes (Diptera: Glossinidae) along the shores of Lake Victoria, Kenya.* Bulletin of Entomological Research, 1997. **87**(02): p. 179-186.
164. Ariel, E., *Viruses in reptiles.* Vet Res, 2011. **42**(1): p. 100.
165. Greenberg, B., *Flies and disease*. Vol. 2. 1973: Princeton University Press Princeton.
166. Tabachnick, W.J., *Nature, nurture and evolution of intra-species variation in mosquito arbovirus transmission competence.* Int J Environ Res Public Health, 2013. **10**(1): p. 249-77.
167. MÜLler, G.C., et al., *Attraction of Stomoxys sp. to various fruits and flowers in Mali.* Medical and Veterinary Entomology, 2012. **26**(2): p. 178-187.
168. Foil, L.D., et al., *Mechanical Transmission of Equine Infectious-Anemia Virus by Deer Flies (Chrysops-Flavidus) and Stable Flies (Stomoxys-Calcitrans).* American Journal of Veterinary Research, 1983. **44**(1): p. 155-156.
169. Hawkins, J.A., et al., *Role of Horse Fly (Tabanus-Fuscicostatus Hine) and Stable Fly (Stomoxys-Calcitrans L) in Transmission of Equine Infectious Anemia to Ponies in Louisiana.* American Journal of Veterinary Research, 1973. **34**(12): p. 1583-1586.
170. Mellor, P.S., R.P. Kitching, and P.J. Wilkinson, *Mechanical Transmission of Capripox Virus and African Swine Fever Virus by Stomoxys-Calcitrans.* Research in Veterinary Science, 1987. **43**(1): p. 109-112.
171. Doyle, M.S., et al., *Vector Competence of the Stable Fly (Diptera: Muscidae) for West Nile Virus.* Journal of Medical Entomology, 2011. **48**(3): p. 656-668.
172. Johnson, G., et al., *Detection of West Nile Virus in Stable Flies (Diptera: Muscidae) Parasitizing Juvenile American White Pelicans.* Journal of Medical Entomology, 2010. **47**(6): p. 1205-1211.
173. Mihok, S., et al., *Mechanical Transmission of Trypanosoma Spp by African Stomoxyinae (Diptera, Muscidae).* Tropical Medicine and Parasitology, 1995. **46**(2): p. 103-105.
174. Sumba, A.L., S. Mihok, and F.A. Oyieke, *Mechanical transmission of Trypanosoma evansi and T-congolense by Stomoxys niger and S-taeniatus in a laboratory mouse model.* Medical and Veterinary Entomology, 1998. **12**(4): p. 417-422.
175. Mohammed, Y.O., et al., *Detection of Trypanosoma brucei gambiense and T.b. rhodesiense in Glossina fuscipes fuscipes (Diptera: Glossinidae) and Stomoxys flies using the polymerase chain reaction (PCR) technique in southern Sudan.* African Journal of Biotechnology, 2010. **9**(38): p. 6408-6412.
176. Berberian, D.A., *Successful transmission of cutaneous leishmaniasis by the bites of Stomoxys calcitrans.* Proceedings of the Society for Experimental Biology and Medicine, 1938. **38**(2): p. 254-256.
177. Traversa, D., et al., *Identification of the intermediate hosts of Habronema microstoma and Habronema muscae under field conditions.* Medical and Veterinary Entomology, 2008. **22**(3): p. 283-287.
178. Yarmut, Y., et al., *Ophthalmic and Cutaneous Habronemiasis in a Horse: Case Report and Review of the Literature.* Israel Journal of Veterinary Medicine, 2008. **63**(3): p. 87-90.

179. Lehane, M.J., et al., *Adult midgut expressed sequence tags from the tsetse fly Glossina morsitans morsitans and expression analysis of putative immune response genes.* Genome Biol, 2003. **4**(10): p. R63.
180. Barry, M.K., A.A. Triplett, and A.C. Christensen, *A peritrophin-like protein expressed in the embryonic tracheae of Drosophila melanogaster.* Insect Biochemistry and Molecular Biology, 1999. **29**(4): p. 319-327.
181. Suetake, T., et al., *Production and characterization of recombinant tachycitin, the Cys-rich chitin-binding protein.* Protein Eng, 2002. **15**(9): p. 763-9.
182. Kawabata, S., et al., *Tachycitin, a small granular component in horseshoe crab hemocytes, is an antimicrobial protein with chitin-binding activity.* J Biochem, 1996. **120**(6): p. 1253-60.
183. Asensio, J.L., et al., *NMR investigations of protein-carbohydrate interactions: studies on the relevance of Trp/Tyr variations in lectin binding sites as deduced from titration microcalorimetry and NMR studies on hevein domains. Determination of the NMR structure of the complex between pseudohevein and N,N',N''-triacetylchitotriose.* Proteins, 2000. **40**(2): p. 218-36.
184. Boraston, A.B., et al., *Carbohydrate-binding modules: fine-tuning polysaccharide recognition.* Biochem J, 2004. **382**(Pt 3): p. 769-81.
185. Gibson, W. and M. Bailey, *The development of Trypanosoma brucei within the tsetse fly midgut observed using green fluorescent trypanosomes.* Kinetoplastid Biol Dis, 2003. **2**(1): p. 1.
186. Roditi, I. and M.J. Lehane, *Interactions between trypanosomes and tsetse flies.* Curr Opin Microbiol, 2008. **11**(4): p. 345-51.
187. Kubi, C., et al., *The effect of starvation on the susceptibility of teneral and non-teneral tsetse flies to trypanosome infection.* Med Vet Entomol, 2006. **20**(4): p. 388-92.
188. Willett, K.C., *Development of Peritrophic Membrane in Glossina (Tsetse Flies) and Its Relation to Infection with Trypanosomes.* Experimental Parasitology, 1966. **18**(2): p. 290-&.
189. Vickerman, K., *Developmental cycles and biology of pathogenic trypanosomes.* Br Med Bull, 1985. **41**(2): p. 105-14.
190. Tetley, L. and K. Vickerman, *Differentiation in Trypanosoma brucei: host-parasite cell junctions and their persistence during acquisition of the variable antigen coat.* J Cell Sci, 1985. **74**: p. 1-19.
191. Vickerman, K., et al., *Biology of African trypanosomes in the tsetse fly.* Biol Cell, 1988. **64**(2): p. 109-19.
192. Eisemann, C., G. Wijffels, and R.L. Tellam, *Secretion of the type 2 peritrophic matrix protein, peritrophin-15, from the cardia.* Arch Insect Biochem Physiol, 2001. **47**(2): p. 76-85.
193. Haddow, J.D., et al., *Identification of midgut proteins that are differentially expressed in trypanosome-susceptible and normal tsetse flies (Glossina morsitans morsitans).* Insect Biochem Mol Biol, 2005. **35**(5): p. 425-33.
194. Bruce, D., *Preliminary report on the tsetse fly disease or nagana, in Zululand*. 1895, Durban: Bennett and Davis.
195. Bruce, D., et al., *The development of trypanosomes in tsetse flies. .* Report of the Sleeping Sickness Commission of the Royal Society . 1910. **11** p. 12-33.
196. Aksoy, S., W. Gibson, and M.J. Lehane, *Interactions between tsetse and trypanosomes with implications for the control of trypanosomiasis.* Adv Parasitol, 2003. **53**: p. 1 - 83.
197. Dyer, N.A., et al., *Flying tryps: survival and maturation of trypanosomes in tsetse flies.* Trends in Parasitology, 2013. **29**(4): p. 188-196.
198. Ridgley, E.L., Z.H. Xiong, and L. Ruben, *Reactive oxygen species activate a Ca2+-dependent cell death pathway in the unicellular organism Trypanosoma brucei brucei.* Biochemical Journal, 1999. **340**: p. 33-40.
199. Vanhollebeke, B. and E. Pays, *The trypanolytic factor of human serum: many ways to enter the parasite, a single way to kill.* Mol Microbiol, 2010. **76**(4): p. 806-14.

200. Moloo, S.K. and S.B. Kutuza, *Feeding and crop-emptying in Glossina brevipalpis Newstead.* Acta Trop., 1970. **27**(4): p. 356-377.
201. Taylor, A.W., *The Development of West African Strains of Trypanosoma gambiense in Glossina tachinoides under Normal Laboratory Conditions, and at Raised Temperatures.* Parasitology, 1932. **24**(03): p. 401-418.
202. Yorke, W., *The relation of polymorphic trypanosomes, developing in the gut of Glossina, to the peritrophic membrane.* Annals of Tropical Medicine and Parasitology, 1933. **27**: p. 347-350.
203. Gordon, R.M., *Trypanosoma congolense in its passage through the peritrophic membrane of Glossina morsitans. .* Transactions of the Royal Society of Tropical Medicine and Hygiene, 1957. **51**: p. 296.
204. Freeman, J.C., *The penetration of the peritrophic membrane of the tsetse flies by trypanosomes. .* Acta Tropica 1973. **30**: p. 347-355.
205. Berridge, M.J., *A structural analysis of intestinal absorption.* Symp R Entomol Soc Lond, 1970. **5**: p. 135-150.
206. Wigglesworth, V.B., *Digestion in the tsetse-fly: a study of structure and function.* Parasitology, 1929. **21**: p. 288-321.
207. Wijers, D.J., *Factors that may influence the infection rate of Glossina palpalis with Trypanosoma gambiense. I. The age of the fly at the time of the infected feed.* Ann Trop Med Parasitol, 1958. **52**(4): p. 385-90.
208. Freeman, J.C., *The penetration of the peritrophic membrane of the tsetse flies by trypanosomes.* Acta Trop, 1973. **30**(4): p. 347-55.
209. Evans, D.A. and D.S. Ellis, *The penetrative ability of sleeping-sickness trypanosomes.* Trans R Soc Trop Med Hyg, 1978. **72**(6): p. 653-5.
210. Langer, R.C. and J.M. Vinetz, *Plasmodium ookinete-secreted chitinase and parasite penetration of the mosquito peritrophic matrix.* Trends Parasitol, 2001. **17**(6): p. 269-72.
211. Deerinck, T.J., et al., *NCMIR methods for 3D EM: a new protocol for preparation of biological specimens for serial block face scanning electron microscopy.* Microscopy, 2010: p. 6-8.
212. Welburn, S.C., I. Maudlin, and D.H. Molyneux, *Midgut lectin activity and sugar specificity in teneral and fed tsetse.* Med Vet Entomol, 1994. **8**(1): p. 81-7.
213. Van den Abbeele, J., et al., *Trypanosoma brucei spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis.* Parasitology, 1999. **118**: p. 469-478.
214. Lorenz, P., et al., *Elongation and clustering of glycosomes in Trypanosoma brucei overexpressing the glycosomal Pex11p.* EMBO J, 1998. **17**(13): p. 3542-55.
215. Voncken, F., et al., *Depletion of GIM5 Causes Cellular Fragility, a Decreased Glycosome Number, and Reduced Levels of Ether-linked Phospholipids in Trypanosomes.* Journal of Biological Chemistry, 2003. **278**(37): p. 35299-35310.
216. Li, F.-J. and C.Y. He, *Acidocalcisome is required for autophagy in Trypanosoma brucei.* Autophagy, 2014. **10**(11): p. 1978-1988.
217. Welburn, S.C. and I. Maudlin, *Control of Trypanosoma brucei brucei infections in tsetse, Glossina morsitans.* Medical and Veterinary Entomology, 1997. **11**(3): p. 286-289.
218. Hao, Z.G., I. Kasumba, and S. Aksoy, *Proventriculus (cardia) plays a crucial role in immunity in tsetse fly (Diptera : Glossinidiae).* Insect Biochemistry and Molecular Biology, 2003. **33**(11): p. 1155-1164.
219. Hao, Z., et al., *Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis.* Proceedings of the National Academy of Sciences U.S.A. , 2001. **98**(22): p. 12648-53.
220. Boulanger, N., et al., *Immune response of Drosophila melanogaster to infection with the flagellate parasite Crithidia spp.* Insect Biochemistry and Molecular Biology, 2001. **31**(2): p. 129-137.

221. Benoit, J.B., et al., *Aquaporins Are Critical for Provision of Water during Lactation and Intrauterine Progeny Hydration to Maintain Tsetse Fly Reproductive Success.* Plos Neglected Tropical Diseases, 2014. **8**(4).
222. Ruepp, S., et al., *Survival of Trypanosoma brucei in the tsetse fly is enhanced by the expression of specific forms of procyclin.* J Cell Biol, 1997. **137**(6): p. 1369-79.

Supporting Information

Ch2_SF1 In-gel single hit validation list of *Glossina morsitans morsitans* peritrophic matrix proteins from in-gel digestion analysis (ion score cut off of 30).

Ch2_SF2 In-solution single peptide hit validation

Ch2_ST1 Preliminary mass spectrometry results of *Glossina* and *Sodalis* peptide hits

Ch2_SD1 *Glossina morsitans* Mascot data.

Ch2_SD2 *Sodalis glossinidius* Mascot data

Ch3_SI1 Tables of fly habitats and host preference

Ch3_SF1 High resolution version of phylogenetic analysis of Dipteran PADs

Ch3_ST1 Metadata analysis form for VectorBase submission

Ch4_SI1 Protocol for dsRNA injection of tsetse

Ch4_SF1 Immunoblots of pre-immune rabbit serum against tsetse midguts

Ch5_S1 Serial sections of a “cyst” of TSW196 trypanosomes in the PM of an 11 d.p.i fly

Ch5_S2 3D reconstruction of a “cyst” of TSW196 trypanosomes in the PM of an 11 d.p.i fly

Ch5_S3 Serial sections of Region of Interest 1 of a PV from a 5 d.p.i infected fly

Ch5_S4 Serial sections of Region of Interest 2 of a PV from a 5 d.p.i infected fly

Ch5_S5 3D reconstruction of Region of Interest 2 of a PV from a 5 d.p.i infected fly

Ch5_S6 Selection of micrographs from infected flies which differ from the norm

Ch5_S7 Selection of micrographs from the hindgut of naïve and refractory flies

Appendix 1

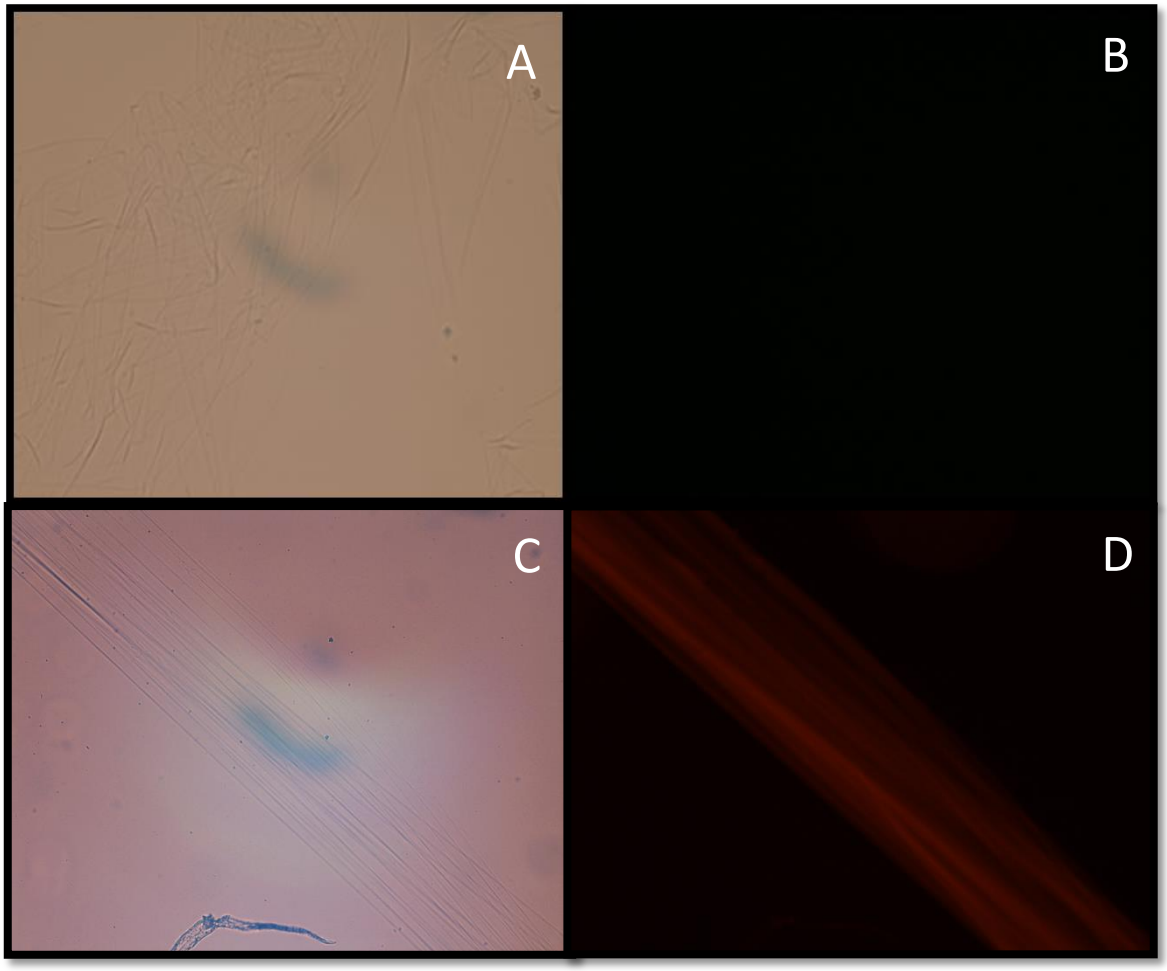


Figure A1. PMs stained with WGA after chitin hydrolysate treatment. *(A) Brightfield image of a PM incubated in fluorescein succinylated WGA after the latter was incubated in chitin hydrolysate (1:4 dilution). (B) Fluorescent image of the same area as in A, showing no labelling as the succinylated WGA was unable to bind. (C) Brightfield image of a PM incubated in rhodamine WGA after the latter was incubated in chitin hydrolysate (1:40 dilution). (D) Fluorescent image of the same area as in C, showing reduced labelling of the PM compared to no incubation of WGA with chitin hydrolysate.*

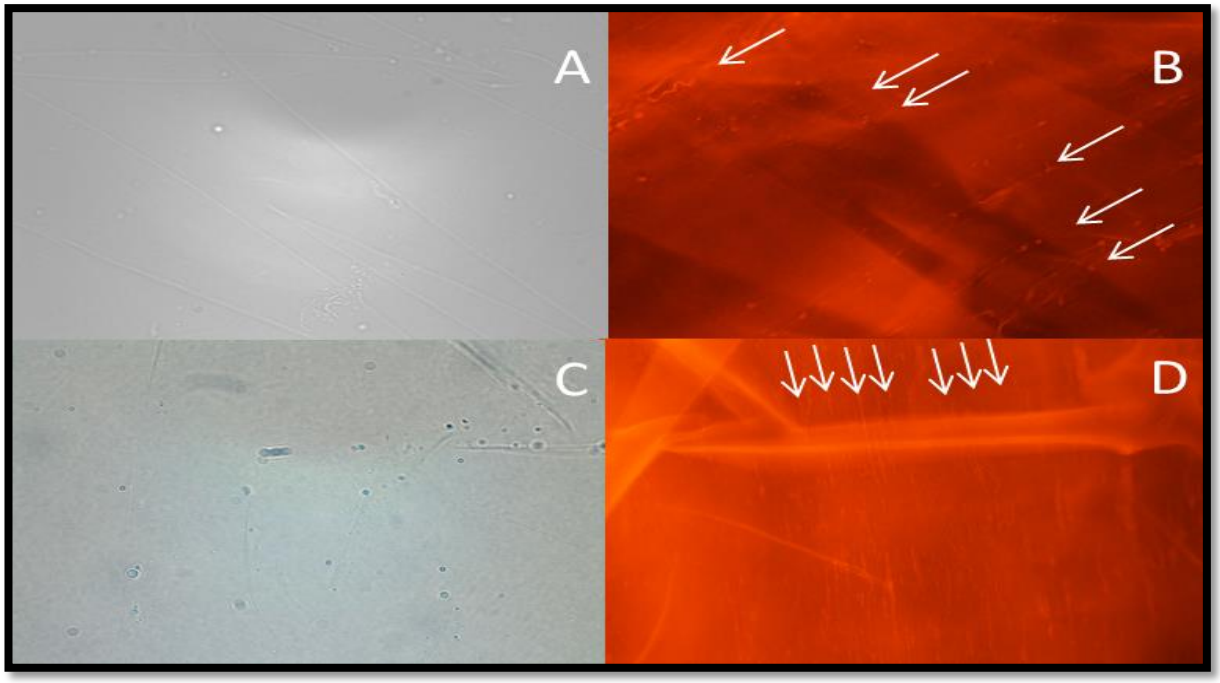


Figure A2. Variations in WGA staining of the tsetse PM. *(A) Brightfield image of an isolated PM from an 11 day old fly infected with TSW 196* T. brucei*. (B) Fluorescence image from the same area of the PM as A showing increase rhodamine signal on parts of the PM, resulting in parallel tracts (arrows). (C) Brightfield image of an isolated PM from a 5 day old naïve fly. (D) Fluorescence image of the same area of PM as C, showing parallel tracts (arrows).*

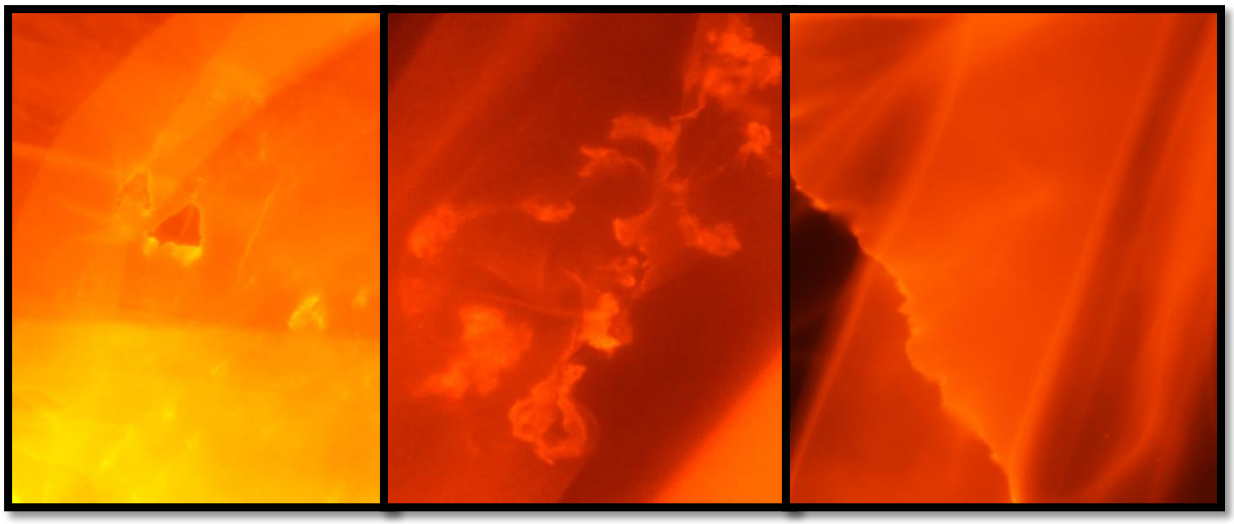


Figure A3. PMs showing mechanical damage from three different flies and ages; *(A) 8 day control PM, (B) 8 day infected PM, (C) 11 day infected PM*

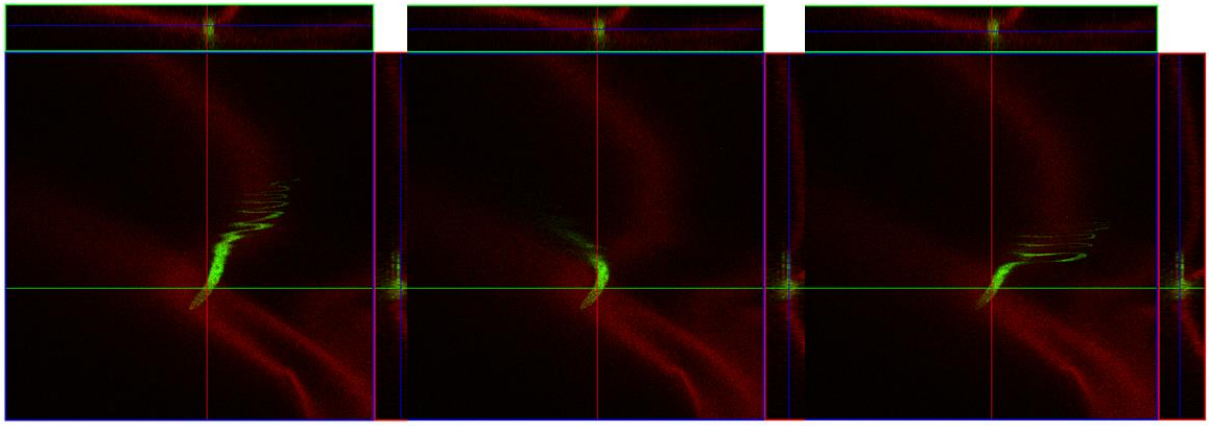


Figure A4. Stills taken from a Z-stack series. *The posterior of a GFP J10 trypanosome can be seen adhered to the PM from a fly 4 d.p.i. The flagellum is free and is flapping about. 100x*

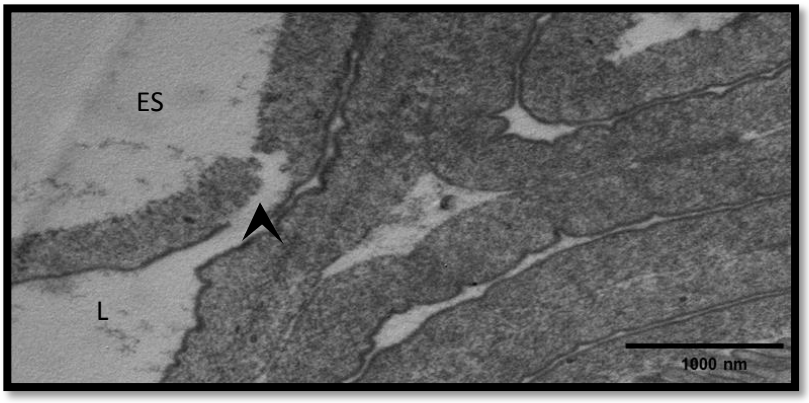


Figure A5. An electron micrograph of a section of PM from a naïve fly at 11 days old. *ES: Ectoperitrophic space, L: Lumen, Arrowhead: Shows a transverse tear in the PM most probably caused by mechanical damage from dissections. 26,500x*

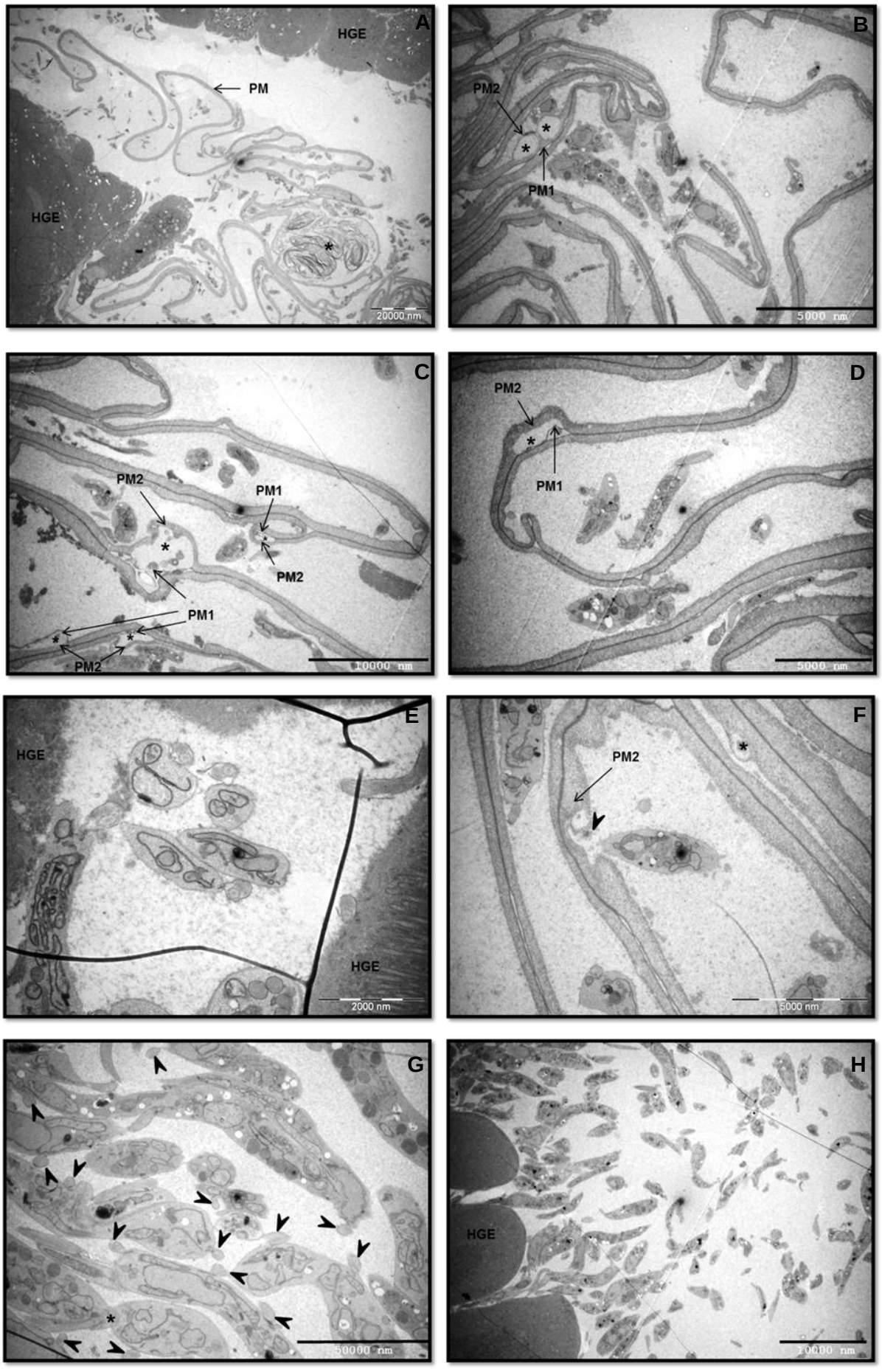

A
HGE
PM
HGE
20000 nm
B
PM2
PM1
5000 nm
C
PM2
PM1
PM2
PM1
PM2
10000 nm
D
PM2
PM1
5000 nm
E
HGE
HGE
2000 nm
F
PM2
5000 nm
G
50000 nm
H
HGE
10000 nm

Figure A6. A selection of TEM micrographs taken from the hindgut of a fly at 11 d.p.i. *Details of each can be found in the text below.*

Fig. A6a. An overview of the hindgut of an infected fly. The epithelial cells of the hindgut are distinctive and the PM can be seen. In this region, the PM appears thinner, and material presumably from degraded and discarded PM can be seen (*). 790 x

Fig. A6b-d. Holes can be seen between PM1 and PM2 although in all holes examined there appeared to be no visible parasites.

Fig. A6e. In some instances, there appeared to be parasites that looked typical in shape to what is normally seen in the anterior midgut.

Fig. A6f. Similar to the rare instances that are observed in the anterior gut, there is sometimes damage to PM2 rather than a separation between PM1 and PM2. A trypanosome can be seen near to an atypical PM2, possibly breaking free into the ectoperitrophic space. Damage and a break can be seen (arrowhead). Damage can also be seen to PM2 (*) in another part of the PM.

Fig. A6g. Abnormal looking trypanosomes can be seen in the ectoperitrophic space of the hindgut. Arrowheads indicate the presence of multiple flagella per trypanosome body suggesting a problem in cytokinesis.

Fig. A6h. Parasites can be seen that appear to be uncharacteristic in shape to what is normally seen. There is a distinct lack of flagella present for many cells.

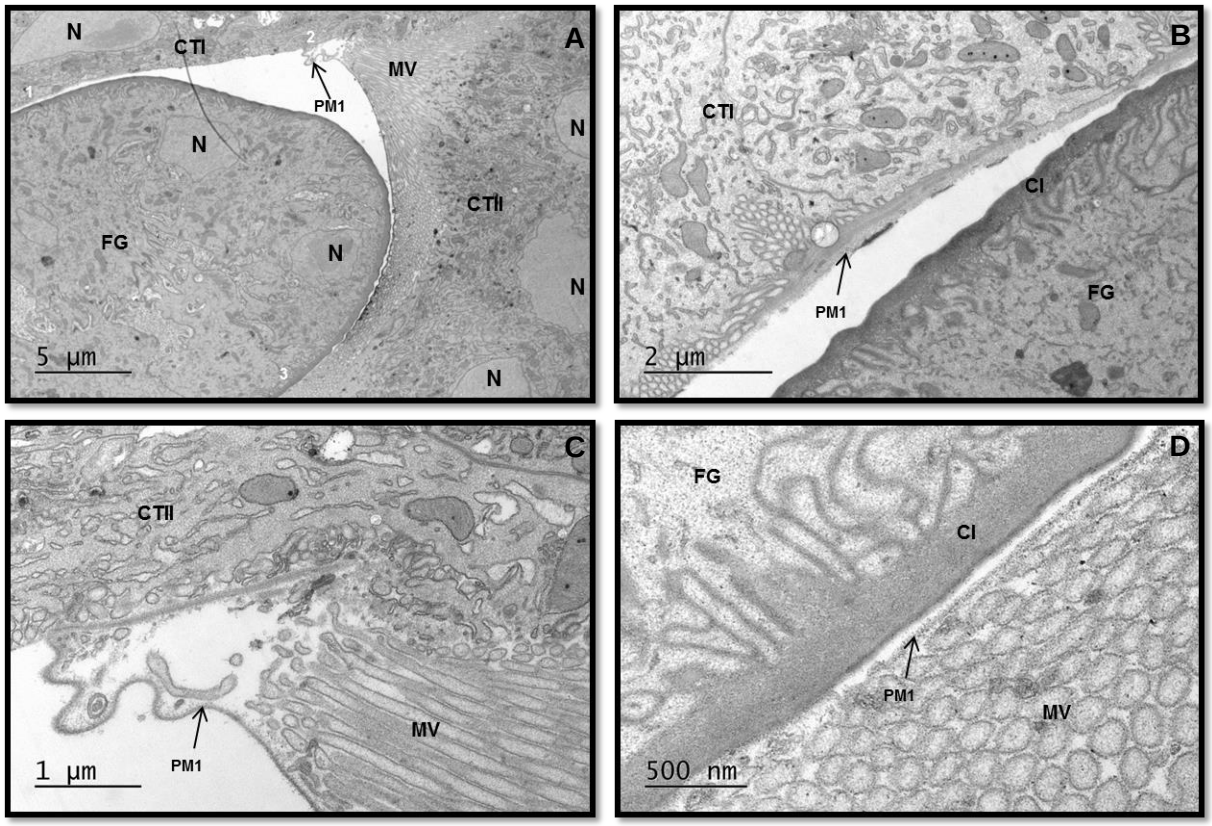


Figure A7. Electron micrographs of the tsetse PV showing the formation of the PM. *(A) An overview image showing the invagination of the foregut in context to the annular pad of the PV. (B) Area 1 as seen in A (white text). PM1 is being secreted in an unstructured fashion at its origin. (C) Area 2 as seen in A (white text). PM1 is a clearly defined electron-dense structure and secretions can be seen originating from the microvilli and merging with PM1. (D) Area 3 as seen in A (white text). PM1 can be seen in close proximity to the microvilli of the annular pad cell type II. Few secretions can be seen. Images taken from a naïve fly at 5 days old. Abb. CI: Cuticular intima, CTI: Cell type I of the PV, CTII: Cell type II of the PV, FG: Foregut, MV: Microvilli, N: Nucleus, PM1: PM layer 1.*

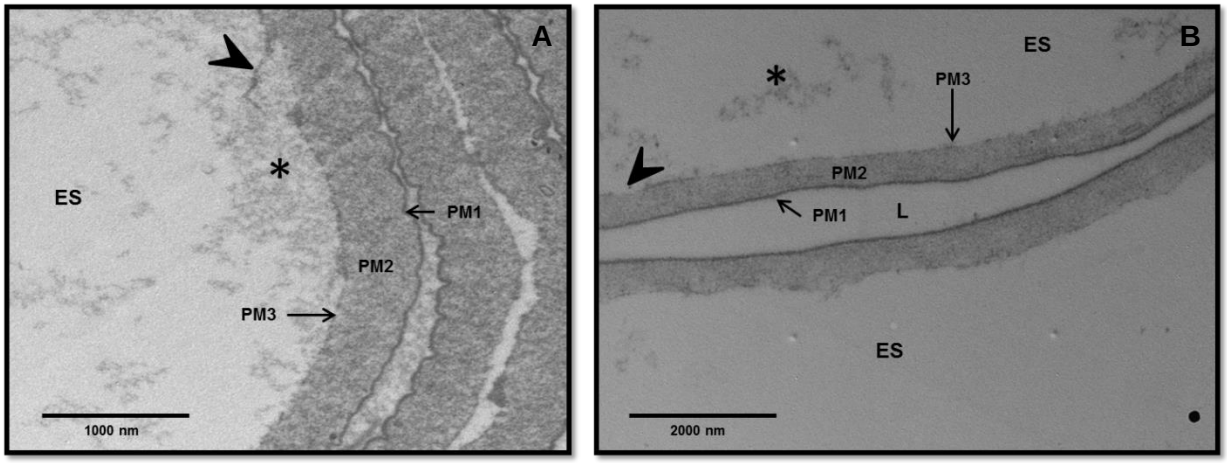


Figure A8. Transmission electron micrographs showing the 3rd layer of the tsetse PM being secreted and laid down onto a mature PM2. *(A) A micrograph showing the 3 layers of the PM and secretions presumably coming from the epithelial cells (*) are being laid down onto PM2 (arrowhead). Taken from a fly at 11 d.p.i. (B) A micrograph showing all 3 PM layers and few secretions (*) that stick onto the second layer (arrowhead). Taken from a naïve fly at 5 days old.*

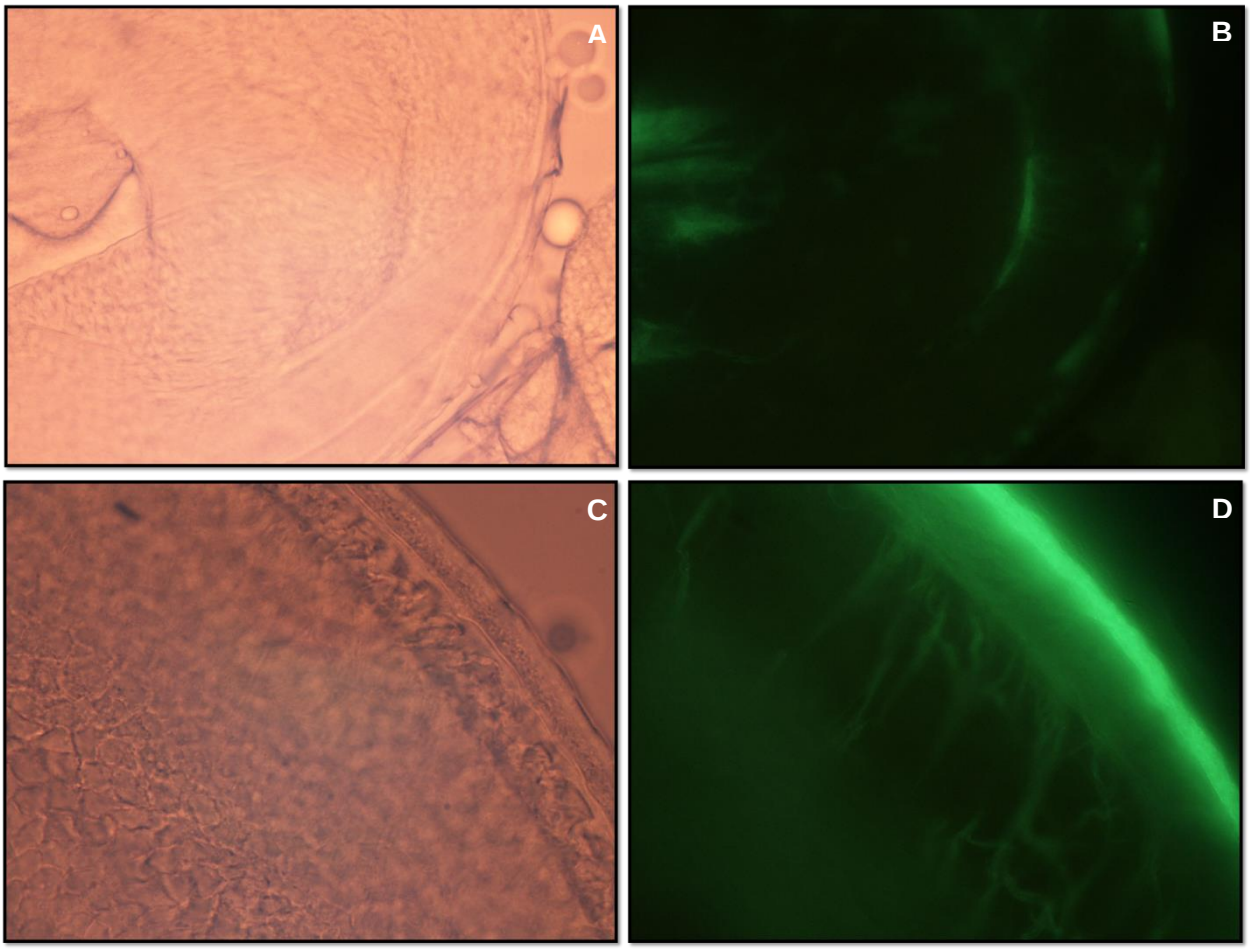


Figure A9. Fluorescent images of GFP expressing trypanosomes in a PV taken from a fly at 5 d.p.i. *(A) A brightfield image of a PV. The beginning of the midgut can be seen. (B) Fluoresent image of the same area as shown in A. GFP trypansomes can be seen lining up against the edge of the foregut and PV epithelial cells. (C) A brightfield image of the edge of a PV. (D) Fluorescent image of the same area as in C. Trypansomes can be seen in the space between the foregut and the annular pads of the PV. A and B 40x, C and D 100x.*