The identification and partial characterization of molecules found within the midgut of the tsetse fly

by

Lee Rafuse Haines
B.Sc., Trinity Western University 1994

A Dissertation Submitted in Partial Fulfilment of the Requirements for the Degree of

MASTER OF SCIENCE

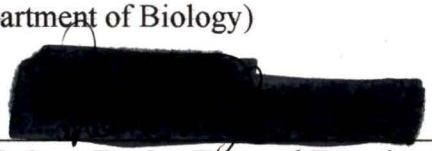
in the Department of Biochemistry and Microbiology

We accept this dissertation as conforming to the required standard

Dr. Terry Pearson, Supervisor
(Department of Biochemistry and Microbiology)

Dr. Robert Olafson, Department Member
(Department of Biochemistry and Microbiology)

Dr. Richard Ring, Outside Member
(Department of Biology)



Dr. Robert Burke, External Examiner
(Department of Biology)

University of Victoria

Supervisor: Dr. Terry W. Pearson

Abstract

Molecules in the midgut of the tsetse fly (Diptera: Glossinidae) are thought to play an important role in the life cycle of African trypanosomes by influencing their initial establishment in the midgut and subsequent differentiation events that ultimately affect parasite transmission. It is thus important to determine the molecular composition of the tsetse midgut to aid in understanding disease transmission by these medically important insect vectors. To my surprise, I found that the most abundant protein in midguts of teneral (unfed) *Glossina morsitans morsitans* was a 60 kDa molecular chaperone of bacterial origin. Since two species of symbiotic bacteria reside in the tsetse midgut, *Sodalis glossinidius* and *Wigglesworthia glossinidia,* 2-D gel electrophoresis and mass spectrometry were used to determine which of these organisms was the source of the 60 kDa molecule. To do this, peptide mass maps were compared to virtual peptide maps predicted for *S. glossinidius* and *W. glossinidia* chaperone sequences. Four signature peptides were identified, revealing that the source of the chaperone was *W. glossinidia.* Comparative two-dimensional gel electrophoresis and immunoblotting further revealed that this protein was localised to the anterior midgut containing the bacteriome and not the distal portion of the tsetse midgut. The possible function of this highly abundant endosymbiont chaperone in the tsetse midgut is discussed.

Both *Sodalis glossinidius* and *Wigglesworthia glossinidia* are either required for tsetse viability or fecundity, and may influence the life cycle of African trypanosomes. *S. glossinidius* is thought to modulate the transmission of trypanosomes and thus is of considerable interest. To obtain probes for analysis of this symbiont, pure cultures were established and used for immunisation and derivation of monoclonal antibodies (mAbs). One mAb bound to an abundant 60 kDa protein that was highly expressed on the surface of *Sodalis* and was secreted as a soluble protein into the culture medium. This mAb recognised molecules only in *S. glossinidius* and *W. glossinidia* and not in other members of the Enterobacteriaceae family. Using the mAb in microimmunoadsorbent columns, the 60 kDa protein was isolated from *Sodalis* culture supernatants and identified by mass

spectrometry as a GroEL-like chaperone. Three mAbs were specific for *Sodalis* and did not bind the primary midgut symbiont, *W. glossinidia,* and thus are useful probes for detection of *Sodalis* in tsetse.

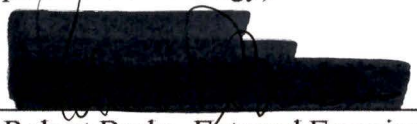
A monoclonal antibody, which recognises a repetitive EP polypeptide epitope and which was originally derived against procyclic form *Trypanosoma brucei*, cross-reacted with midgut lysates from both *Glossina palpalis palpalis* and *Glossina morsitans morsitans.* The DNA sequences of the genes encoding these EP proteins were determined and the translated protein sequences were obtained. Protein structure and function analysis using a variety of predictive algorithms showed that the EP repeat proteins contained distinct domains and antigenic regions. A putative signal peptide and post-translational modifications were predicted, supporting the idea that these molecules are expressed in the lumen of the midgut, and thus are good candidates for participation in vector-parasite interactions.

Examiners:

Dr. Terry Pearson, Supervisor
(Department of Biochemistry and Microbiology)

Dr. Robert Olafson, Department Member
(Department of Biochemistry and Microbiology)

Dr. Richard Ring, Outside Member
(Department of Biology)



Dr. Robert Burke, External Examiner
(Department of Biology)

Insects do not sting out of malice but because they also want to live: likewise our critics—they want our blood, not our pain.
(Friedrich Nietzsche)

Table of Contents

List of Figures

List of Tables

Acknowledgements

I had the fortune of growing up with a reverence for the natural world infused into my veins. From this merger arose a fascination with the intricate and the infinitesimal, and a curiosity to seek for the truths explaning each mystifying observation. However, this innate scientific curiosity alone did not bring me to the place where now I stand. I liken my educational journey to rafting a whitewater river filled with difficult, long and often violent rapids, big drops, hidden currents and steep gradients; all of which evoking both excitement and fear. The raft is what symbolizes my desire to search for the truths hidden within this river of knowledge. A successful expedition requires three main components: a durable raft equipped with reliable gear, a proficient guide possessing the foresight necessary to steer the craft through whitewater and concealed obstructions, and a determined team of paddlers to generate the strength required to punch through standing waves and hydraulics without capsizing.

I am indebted to many people in the past who have served as both mentors and trip leaders. Ed Day, Nevio Rossi, Ed Wong, Ary Bovenkerk, Karen Steensma, and Richard Paulton were among those who taught me how to intelligently scout the water for unobstructed chutes and instilled in me a self-confidence that allowed me to effectively navigate my way through the river of discovery. To those who equipped me with techniques, materials and advice along the way, Dr. Ron Gooding, Dr. Serap Aksoy, Dr. Caren Helbing, Dr. Ed Ishiguro, Dr. William Hintz, Dr. Chaman Singla, Dr. Lee Mizzen and to those on my supervisory committee, Dr. Robert Burke, Dr. Robert Olafson and Dr. Richard Ring, thank you for your invaluable contributions to my research. I am indebted to the tireless lab crew, Jen Chase, Mike Bridge, Neeloffer Mookherjee, Jody Haddow and Morag Booy who were always there paddling in cadence, sharing equally in the triumphs and failures over the years. Albert Labossiere, Scott Scholz, Steve Horak, Derek Smith, Darryl Hardie, Melinda Powell and Claire Tugwell- I cannot imagine how many times you have made sense out of the chaos and trouble-shooted the seemingly impossible. Thank you.

It is difficult to express how grateful I am to those, my family and friends, who have provided the necessary strength behind the paddles to keep me focused on reaching the take-out. In particular I want to acknowledge the roomates who tolerated, loved and

generously gave me pieces of themselves whenever I found life difficult to cope with. Your lended determination, passion, humour, enthusiasm and desire to see me defeat the high water, holes and seemingly impassable Class VI rapids were invaluable gifts that I hope in the future to echo. And to the remaining unnamed, if you know that I have, at last, crawled out of the raft to stand on the bank, then it is you whom I deeply thank for being a part of my journey. Without your assistance, camaraderie, and devotion, my rafting trip would have been impeded by random drifting, aimless circling in back eddies or the lethal entrapment in a logjam.

Come to the edge, he said.
They said: we are afraid.
Come to the edge, he said.
They came.
He pushed them....

And they flew.

(Guillaume Apollinaire)

Terry, not only has your guidance and friendship been exemplary, but your fusion of optimism with scientific astuteness inspired me to pursue this research with abounding enthusiasm. Thank you for unveiling my wings for now I know, one day, I will soar.

And at last, to my parents, Barry and Louise Rafuse. You drafted the blueprints for the raft, you are the fabric it is woven of, you believe in the power of the river and you are the reason I am. It is to you I dedicate this thesis.

Chapter 1. Introduction

Case Study Number 1:

A 50 year old male in reasonably good physical condition walked into a clinic complaining of a persistent fever, abdominal discomfort, headache and general myalgia. Upon questioning, it was found that he recently was on safari in Uganda for two weeks and suffered numerous insect bites. Based on his symptoms, the initial conclusion was that the patient had malaria. Microscopic examination of blood smears, however, showed high numbers of trypanosomes. As the subspecies of *Trypanosoma brucei* are morphologically indistinguishable, a more exact diagnosis was based on epidemiology and the rapidity of disease progression. An infection with *Trypanosoma brucei rhodesiense*, the agent of East African sleeping sickness, was confirmed. The patient was immediately hospitalised. Further examination revealed severe anaemia, thrombocytopenia and tachycardia (135 bpm) but his cerebro spinal fluid (CSF) was negative for trypomastigotes. The only available treatment was Pentamidine, used for the treatment of early stage Gambian sleeping sickness but essentially ineffective against the generally resistant *T. b. rhodesiense*. Suramin, the toxic anti-trypanosomal drug of choice was ordered from the only known depository at the Centers for Disease Control (CDC) in Atlanta, GA. The patient was given pentamidine intravenously for 24 hours until the Suramin arrived. A few hours post Suramin treatment, the parasites had lysed and patient prognosis improved until death was no longer imminent.

Case Study Number 2:

J. H., male, aged 30, a strong young man, was travelling through Sudan in the fall of 1999. He was admitted to the clinic in June 2000 with classical symptoms of an infection: irregular fever, erythema and swollen lymphatic glands. A blood film showed no evidence of either malarial merozoites or African trypanosomes. An oral regime (10 d) of penicillin was prescribed and the patient was released. Four months later, the patient was re-admitted complaining of insomnia, intense headaches and a loss of coordination. The patient displayed extreme emaciation, mental confusion, high fever and slurred speech. Microscopic examination of a CSF sample confirmed that the patient had an advanced neuralgic *T. b. gambiense* trypanosomiasis. The implementation of a rigorous treatment regime (3 injections per day) with the arsenical therapeutic, Melarsoprol, did not prevent further deterioration. The patient developed Melarsoprol-induced haemorrhagic encephalopathy and died two days later (liberally adapted from Ross and Thomson, 1910; Moore *et al.*, 1999).

Summary:

Case study number 1 was based on two recent reports of travellers, American (Dunavan, 2000) and British (Moore *et al.*, 2002) that contracted the usually fatal blood borne parasitic disease, African sleeping sickness. Fortunately, in both of these cases, the CDC (one of the only remaining international repositories of anti-trypanosomal compounds left in the world since manufacturers deemed the continued production of these drugs as "commercially unattractive") was able to supply the appropriate drug for treatment.

It is a sad state of affairs that living in regions with epidemic or endemic African trypanosomiasis, where access to CDC supplies and drug distribution is limited, significantly reduces one's chance of survival. Even if the slower, less virulent form of the disease, West African (Gambian) sleeping sickness, is acquired (as described in Case 2), inadequate diagnostic techniques (Cattand, 1987), diminishing supplies of anti-parasitic therapeutics (White, 2000), drug toxicity and developing trypanosome resistance are severe impediments to the treatment of this disease.

Human African trypanosomiasis, or sleeping sickness, is one of several uncontrolled plagues that ravage sub-Saharan Africa. It is transmitted by tsetse flies infected with either *Trypanosoma brucei gambiense* or *Trypanosoma brucei rhodesiense*. In Africa, 15% of the sub-Saharan population (60 million people) risk contracting this disease in the endemic regions that span 36 countries (Smith *et al.*, 1998). Five of these countries, Sudan, Central African Republic, Uganda, the Democratic Republic of Congo and Angola are also confronting a serious epidemic where an underestimated 500,000 people are currently infected (Barrett, 1999; Moore *et al.*, 1999). Along with the rise of sleeping sickness in Africa, the number of cases of this disease surfacing in non-endemic regions of the world is proportionally increasing. To quote Dr. David A. J. Moore, a specialist in tropical medicine at the Hospital for Tropical Disease (London, UK) in reference to an increased number of British travellers infected with sleeping sickness, "…..with increases in tsetse fly activity reported in many regions of sub-Saharan Africa, the potential for further (international) cases is substantial. Neither (of) the locations

from which our patients acquired their infections is novel, but the sudden appearance of two cases in such a short time is noteworthy" (Moore et al., 2002). As the recent human trypanosomiasis epidemic continues to expand and contemporary control methods (vector control, chemotherapy, and chemoprophylaxis) are slowly rendered either ineffective or obsolete, attention is turning to the development of new approaches for disease management and ultimately, for long-sought eradication.

The African trypanosome has two distinct proliferative life cycle stages that alternate between the insect vector (tsetse fly) and the mammalian host. In both stages, the trypanosomes have developed unique strategies to survive in either the alimentary tract of the fly or the bloodstream of the mammal. A greater understanding of the vector-host-parasite interactions that control disease transmission requires first, a better understanding of the basic biology of the tsetse and the trypanosome and second, an investigation into the molecules of both the vector and the parasite.

Tsetse distribution

The tsetse fly is perhaps the single, most significant factor that has influenced the development of the African continent. With a range of 9 million km^2 (Budd, 1999), tsetse effectively render much of the land unsuitable for livestock (Figure 1.1, Panel A). The reason for this is that tsetse transmit animal trypanosomiasis (known as Nagana, meaning "poorly" in Zulu), a serious threat to wild game but more so, to livestock. A study released in 1999 estimated that 3 billion pounds (6.8 billion Canadian dollars) are lost annually due to cattle mortality and inaccessibility to agriculturally rich land (Budd, 1999). Several wildlife species have evolved resistance mechanisms to animal trypanosomiasis but remain carriers of the parasite, acting as the living reservoir

responsible for the persistence of both human sleeping sickness and Nagana (Vickerman *et al.*, 1993). Nagana is primarily a bovine disease characterized by reduced milk production, retarded growth, weight loss, sterility and, inevitably, loss of the animal. All of these manifestations indirectly contribute to the overall impoverishment and economic devastation in large parts of sub-Saharan Africa. Human sleeping sickness is also spread by tsetse vectors, putting at risk millions of people in endemic foci (Figure 1.1, Panel B), and periodically spreading disease further, in great epidemics.

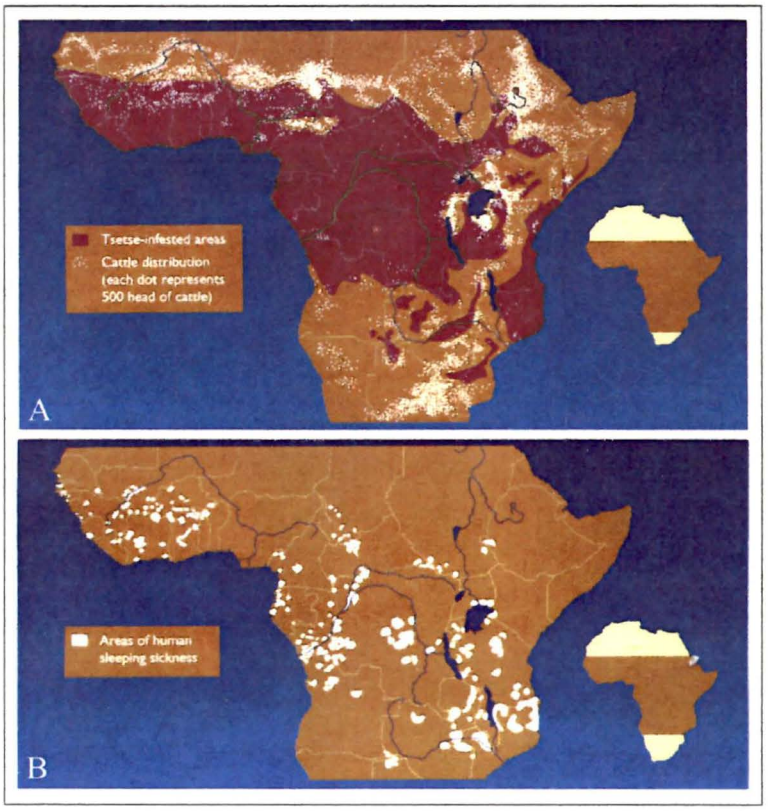


Figure 1.1. ***Geographic distribution of tsetse flies, cattle, and human sleeping sickness in Africa.***

Panel A: tsetse and cattle distribution. Panel B: endemic foci of human sleeping sickness (modified from Hursey, 2001).

Tsetse biology

The name tsetse (pronounced tsee-tsee) is derived from the noise that the fly makes during flight. However, it is interesting to note that it means "fly" in the Tswana language, and in Sechuana, it is interpreted as "fly destructive to cattle". Tsetse are classified under the single genus, *Glossina* and under the order of the two-winged flies, *Diptera.* In addition to two localities in the Arabian peninsula, tsetse fly territory covers a third of Africa south of the Sahara (Figure 1.1, Panel A). There are 23 species and 8 subspecies thus far identified and these are divided into three distinct clades: *morsitans, palpalis* and *fusca,* which refer to the most common species in each subgenus. Each member is classified according to their preferential habitats: river, savannah and forest and to morphological differences in their genitalia (Leak, 1999). In central and west Africa, the riverine species (*Palpalis* group) tend to feed predominantly on reptiles and ungulates. However, as human settlement patterns are typically in close proximity to a water supply, the flies are also vectors of human sleeping sickness. The savannah-woodlands species (*Morsitans* group) are the most economically important as they preferentially feed on livestock and wildlife found in the open grasslands, often found in reserves and national parks. Both the *palpalis* and *morsitans* groups are vectors of *Trypanosoma brucei* spp. Tsetse from the third and most primitive clade (*Fusca* group), inhabit the damp, evergreen forests and are not considered to be medically or agriculturally important as none of the species in this group are vectors of trypanosomiasis.

Despite the diversification in the species, all tsetse are easy to distinguish from other insects. The flies are light brown to black in colour and are roughly twice the size

of a housefly, depending on the species. There are two anatomical features that make visual identification from other insects relatively easy. The predominant characteristic is the "hatchet-shaped" cell found in the center of each wing between the 4th and 5th veins (Figure 1.2, Panel B). The other less obvious feature is the presence of long, fine hairs (setae) on the third antennal segment situated between the eyes (ellipse, Panel A).

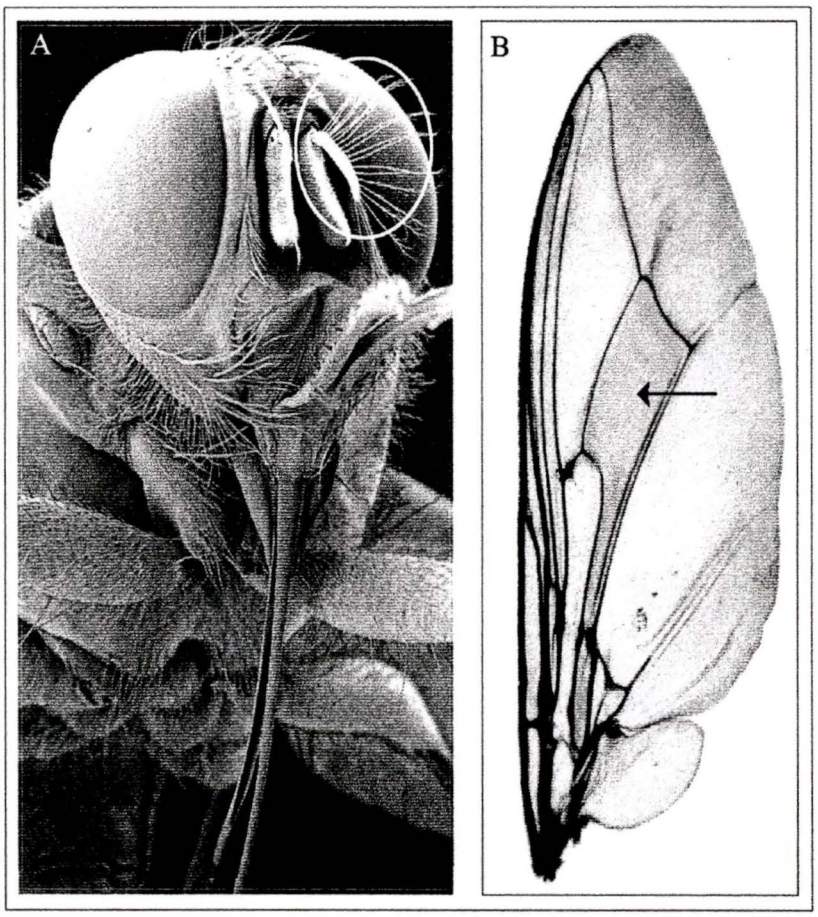


Figure 1.2. ***Phenotypic identification of tsetse using two features that distinguish them from other species of flies.***

Panel A: Secondary branching hairs on antennae (ellipse). Panel B: Characteristic hatchet-shaped cell within each wing membrane (arrow). Panel A: Scanning electron micrograph taken by Oliver Meckes (Dunavan, 2000). Panel B: Photograph taken by L. R. Haines.

Life cycle of the tsetse

The life cycle of the tsetse is both unusual and fascinating as the female fly does not lay eggs but instead gives birth to 6-12 mature larvae in a normal lifespan of 4 months (Hoffman, 1954). In most *Glossina* spp., the female is sexually mature 48 -72 hours after emergence while males become fertile several days following eclosion. Mating may take several hours, as a uterine spermatophore (temporary sperm capsule) must form prior to sperm deposition. The female needs to mate only once as the sperm remain viable (within a specialized sperm storage structure called the spermatheca) during the reproductive life of the female. Males, however, can mate up to 15 times during their lifetime (Leak, 1999). Female tsetse reproduce by an unusual process called adenotrophic viviparity, which, in colloquial terms, is like giving birth to teenagers. This process involves the hatching of a single egg and subsequent larval development (to third instar) entirely *in utero*. The larva is nourished by the secretions from a pair of uterine glands commonly referred to as the milk glands. At the time of larviposition (birth), the mature third instar larvae weighs more than the adult female fly. The entire process, from fertilization to parturition takes 9-10 days (Denlinger and Ma, 1974). Upon deposition in suitable soil, the larva quickly burrows below the surface to pupate. After only a few hours, the puparium (pupa case) becomes fully sclerotised and remains in this protective casing until eclosion (Zdarek and Denlinger, 1993). The puparial period is long, typically taking 30 –40 days, although environmental variations such as temperature, light intensity and soil type can positively or negatively influence the length (and success) of metamorphosis (Leak, 1999). As well, differences in puparial duration have been observed between the species and sexes (Saunders, 1962). The young teneral

(condition after eclosion before the insect cuticle is completely sclerotised) adult exits the puparium and emerges from the soil in mid-afternoon (Dean *et al.*, 1968) whereupon the wings slowly inflate, the exoskeleton hardens and the cycle begins again. Figure 1.3 shows the physical characteristics of each of the stages described. With energy reserves depleted during metamorphosis, teneral tsetse immediately seek a vertebrate host to obtain the first bloodmeal. In both west Africa (Weitz, 1963) and east Africa (Weitz and Glasgow, 1956), tsetse flies, in particular *G. palpalis*, will initially feed on anything from mammals and reptiles to birds. In fact, tsetse are ranked the most promiscuous hematophagous insects in the world with respect to their complete lack of host specificity (Okedi, 1995). This lack of host specificity for feeding, together with the extensive intrauterine larval maturation, drastically increases tsetse survivorship and more than compensates for the abnormally low rate of reproduction.

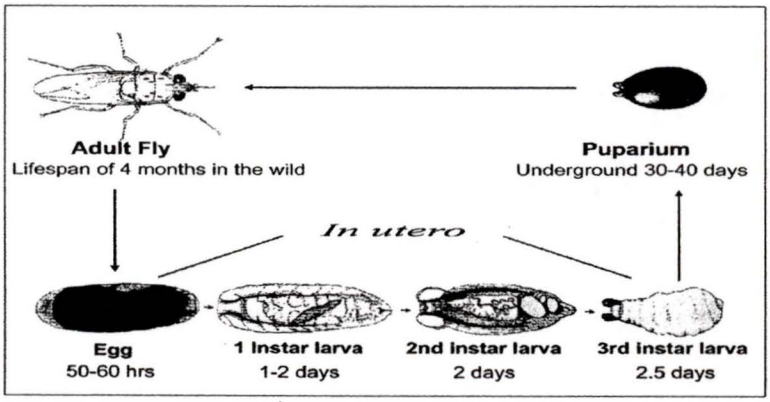


Figure 1.3. ***Schematic representation of the tsetse life cycle.***
Female tsetse produce a single egg every 10 days. Four days after the egg is fertilized it hatches within the uterus as a first instar larva. The larva feeds on highly specialized intrauterine milk glands and parturition of the third instar larva occurs on day 9. After deposition, the larva burrows into the soil to pupate. The adult emerges after a puparial period of at least one month, depending on environmental conditions (adapted from Leak, 1999 p. 16).

Tsetse are the only insect known to cyclically transmit African trypanosomes. The interactions between the trypanosome and its invertebrate and vertebrate hosts are thus of utmost importance. As realised by Robert Koch when commenting on the "tsetse-fly sickness":

> *"Before passing, however, to the principal subject of my discourse, it seems to me important for your enlightenment to give you a short description of the trypanosoma... "(Koch, 1904).*

Trypanosome physiology

The African trypanosomes are eukaryotic salivarian parasites belonging to the Genus Trypanosoma. Members of this genus alternate between two life cycle stages. In the bloodstream form (Figure 1.4, Panel A), the typical sickle shape is created by the presence of microtubules underlying the plasma membrane. The flagellum (light blue) originates at the flagellar pocket, the site of endocytosis, and extends the entire length of the cell to which an undulating membrane is attached. Other major features, shared by most protists, are a predominant nucleus (maroon), complex endoplasmic reticulum, and a single branched mitochondrion (pink). The kinetoplast (dark blue) is a specialised section of the mitochondrion that contains a large mass of mitochondrial DNA (K-DNA). It is the presence of this morphological structure that taxonomically places the trypanosome under the order of Kinetoplastida (Vickerman, 1969). In addition to morphological changes in relative size and linearity of the trypanosomes, the organelles also reposition in response to cellular changes induced by differentiation (Figure 1.4, Panel B). The nucleus, kinetoplast and flagellum persist throughout all of the various life cycle stages and move to different positions within the cell that correlate to the stage of trypomastigote development (Vickerman, 1985).

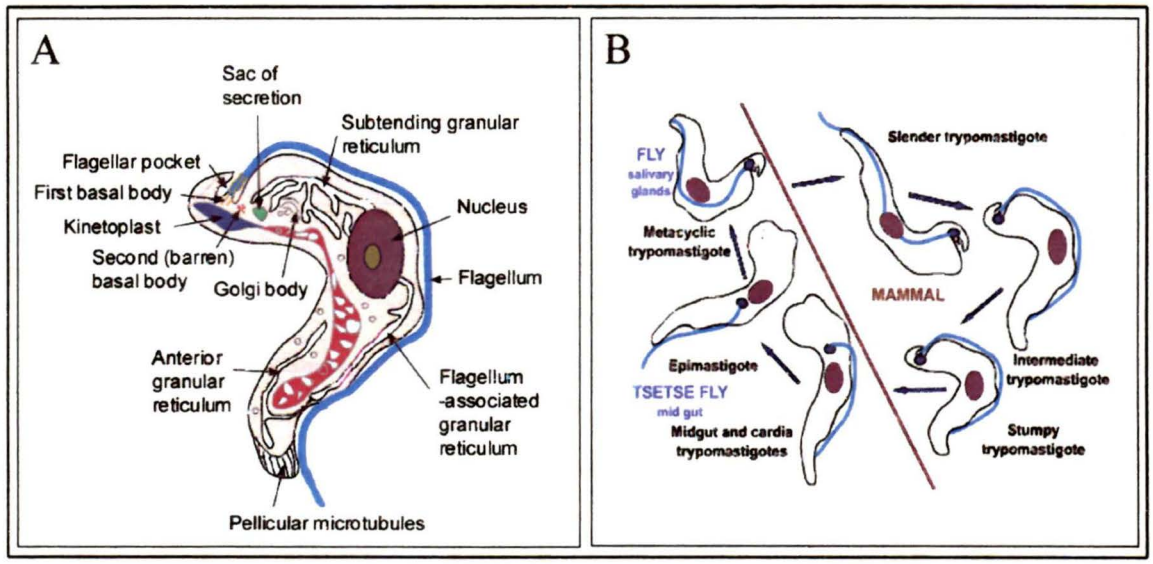


Figure 1.4. ***Schematic representation of a bloodstream form of an African trypanosome and its differentiation in the mammalian host and insect vector.***

Panel A: Principal structures in a bloodstream trypomastigote form shared between all African trypanosomes. Panel B: Changes to the flagellum, kinetoplast and nucleus as the trypanosome differentiates into forms characteristic of the host environment. Note: for clarity, the VSG and procyclin coats have been omitted from the appropriate stages. Panel A: adapted from Vickerman, 1969. Panel B: adapted from Vickerman, 1985.

Trypanosome lifecycle

Mammalian blood infected with African trypanosomes contains extracellular parasites that are covered with an immunogenic surface calyx of approximately 10^7 identical variant surface glycoprotein (VSG) molecules (Vickerman, 1969). This dense, continuous coat physically shields underlying membrane proteins from host immune responses and is central to antigenic variation. This survival strategy allows some parasites in the population to evade immune destruction, the survivors replicating to form the next wave of parasitaemia. The consecutive, but unpredictable expression of a large repertoire of VSG genes permits expansion of antigenically distinct trypanosome

populations within the host. After host immune assault (in response to high parasitaemia) the majority of the parasite population is destroyed. Despite this attack, minor sub-populations in the bloodstream survive by expressing VSG molecules antigenically distinct from the last (variant antigenic types) and consequently remain undetected by the anti-VSG antibodies. The continuous cycles of trypomastigote replication and destruction result in waves of fluctuating parasitaemia clinically observed as irregular fevers and other transient symptoms.

When tsetse consume infective bloodmeals (Figure 1.5), the VSG-covered bloodstream trypomastigotes, along with the blood, pass into the crop for temporary storage. Entrance into the midgut is preceded by the encasement of the bolus by the type II peritrophic membrane (Lehane and Msangi, 1991). Only the nonproliferating stumpy bloodstream form is pre-adapted to survive the now cooler (27°C), proline-rich insect environment. However, only 24 hours after ingestion, even the dual nature of this bloodstream form is lost (Bruce, 1915) as it irreversibly differentiates into the procyclic insect midgut form. This phenomenon is characterized by the complete replacement of the VSG coat by a new, structurally distinct surface of acidic glycoproteins, the procyclins. Procyclic trypanosomes are covered with approximately 5 million procyclin molecules per cell (Pays and Nolan, 1998) and have been proposed to protect the parasite from proteolytic digestion (Acosta-Serrano *et al.*, 2001), and to serve in parasite development and possible ligand-associated parasite-vector signaling (Roditi and Pearson, 1990; Ruepp *et al.*, 1997).

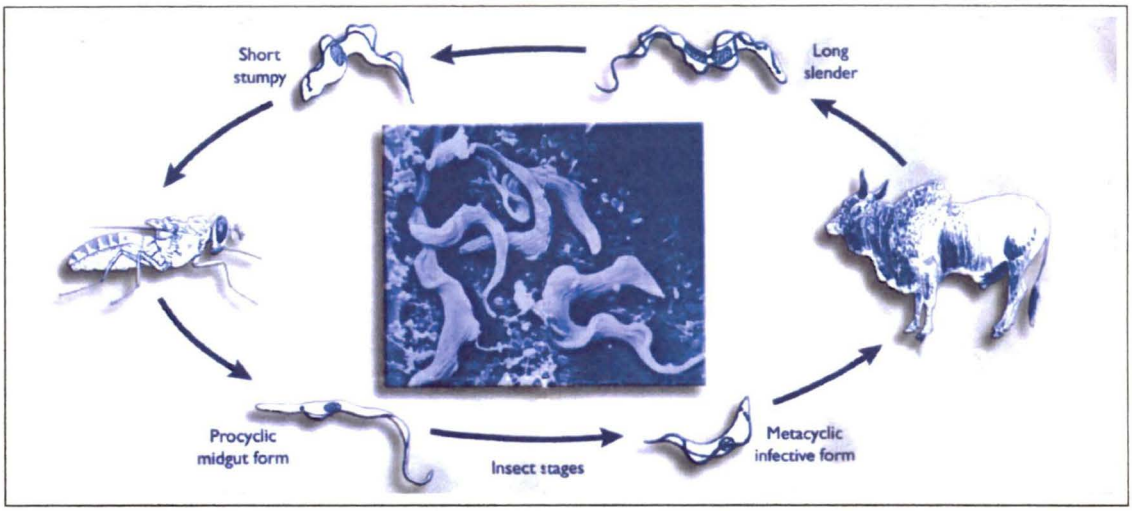


Figure 1.5. ***Schematic representation of the life cycle of African trypanosomes.*** *In the fly, the ingested bloodstream forms transform to procyclic forms in the midgut. The procyclic midgut forms migrate to the salivary glands (or mouthparts) to mature and differentiate into the mammal-infective metacyclic trypomastigotes. The central panel is a scanning electron micrograph of bloodstream form T. brucei (adapted from Hursey, 2001).*

Following successful survival within the tsetse midgut, the differentiated procyclic form must undergo establishment and asexual proliferation. Only a minority of infections proceed beyond this stage yet, in order for successful transmission, the parasites must further differentiate into the infective metacyclic form within the salivary glands. The migration route of the procyclic trypomastigotes from the midgut to the salivary glands is still speculative. Two theories exist. The classical theory (Figure 1.6) describes a route through the peritrophic membrane via the hindgut followed by ectoperitrophic transit through the alimentary canal and into the salivary ducts (Vickerman *et al.*, 1988; Maudlin, 1991). The shorter route involves a breach of the peritrophic membrane and gut epithelium (physical and/or enzyme-assisted) into the haemocoel followed by direct penetration into the salivary glands (Evans and Ellis, 1983). Upon reaching the "holy ground", the procyclic trypomastigotes attach their

flagella to the salivary epithelium by interdigitation of their membranes and transform into epimastigotes. The non-infective epimastigotes complete several rounds of replication and differentiation. This differentiation includes the appearance of a surface coat and mitochondrial changes to accompany the posterior migration of the kinetoplast (Figure 1.4, Panel B) before they detach into the lumen as mature, free-form, infective metacyclic trypomastigotes. At this point, each mature metacyclic parasite has acquired a unique VSG coat (forming an antigenically heterogeneous population) and the relevant metabolic transformations necessary for survival after transmission to a mammalian host.

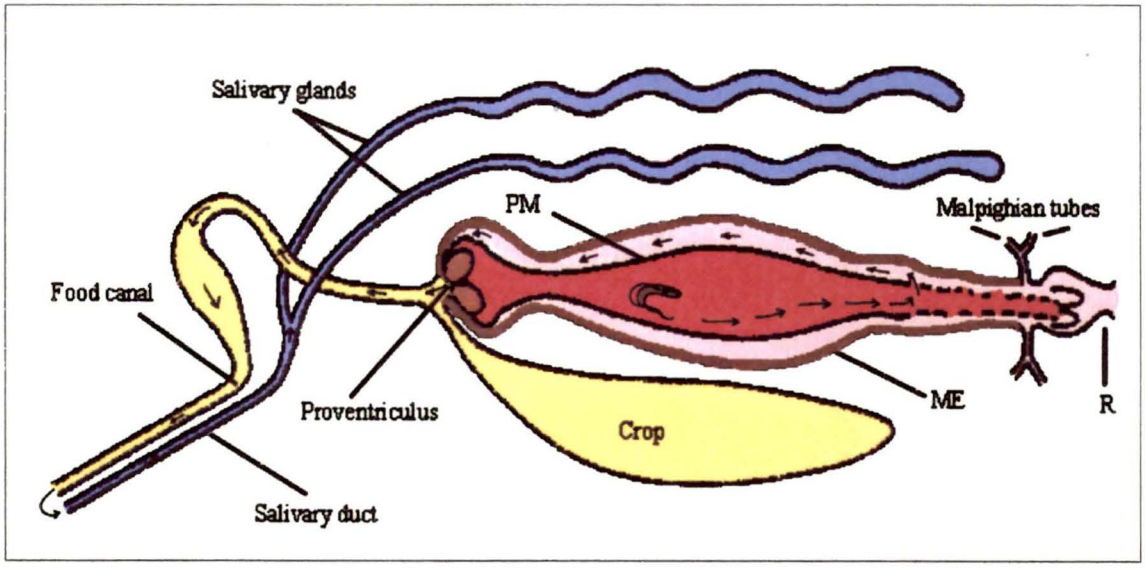


Figure 1.6. ***Schematic representation of the "classical route" of trypanosome migration from the midgut to the salivary glands of the tsetse fly.***
The midgut-localized procyclic trypomastigotes migrate to the hindgut towards the rectum (R) and penetrate the underdeveloped peritrophic membrane (PM) into the ectoperitrophic space bordered by the PM and the midgut epithelium (ME). As the trypanosomes re-traverse the midgut, multiplication ceases and they exit through the proventriculus into the food canal. The journey continues as the trypomastigotes ascend the food canal and enter the salivary glands via the salivary duct. The arrows outline this generally accepted 'classical' parasite route from the midgut lumen to tsetse saliva (Liberally modified from Vickerman et al., 1988).

Midgut complexity

Given that trypanosome differentiation and establishment of an infection within the tsetse midgut are the first steps in the transmission cycle, determination of the molecules involved in triggering or controlling these events is of immense interest. Unfortunately, both genomic and proteomic approaches to identifying and characterizing these molecules is fraught with complications. First, insect, trypanosome and bloodmeal components are all present in the same midgut sea. To further complicate matters, the

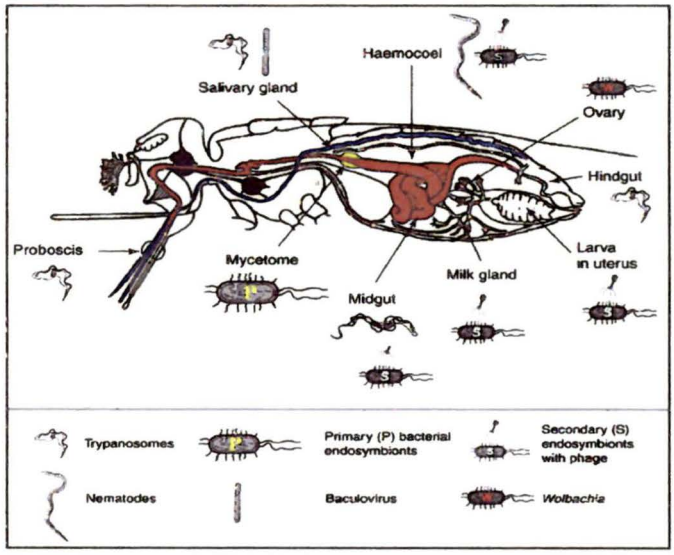


Figure 1.7. Diagrammatic representation of the potential DNA and protein complexity within tsetse.

The primary endosymbiont, Wigglesworthia glossinidia (designated by the letter P) is restricted to the bacteriome (yellow). Sodalis glossinidius, a secondary symbiont (designated by the letter S), is found in the hemolymph, fat body (not shown), milk gland and the midgut. A third bacterium, Wolbachia spp., is localized to the reproductive tissues. African trypanosomes, nematodes and bacteriophages can also be found in tsetse midgut, thus further complicating genomic and proteomic analysis (adapted from Maudlin and Welburn, 2001).

tsetse midgut may potentially contain additional DNA and protein contamination from multiple sources: bacterial endosymbionts and symbionts, phages, baculoviruses and parasitic nematodes (Figure 1.7). Some of these organisms, in particular the bacterial

endosymbionts, are thought to influence disease transmission (Maudlin *et al.*, 1990; Welburn *et al.*, 1993).

Research hypothesis and thesis outline

In my thesis research, I explored the tsetse midgut environment for novel molecules that could directly or indirectly affect the vectorial capacity of the fly. This focus resulted in the identification of several novel midgut proteins from *G. m. morsitans* and *G. p. palpalis,* and two symbiont proteins of bacterial origin. Chapter 1 is a general introduction to African sleeping sickness, the tsetse and the parasite to outline the importance of this vector-based investigation. Chapter 2 describes the biochemical and immunological characterization of the major soluble protein found in the midgut of *G. m. morsitans* (Westwood). This predominant 60 kDa heat shock protein was identified as a bacterial molecule, not a tsetse molecule as expected. The third chapter identifies molecules isolated from an axenic, *in vitro* culture of the bacterial symbiont, *Sodalis glossinidius*. The final chapter investigates a midgut-associated EP repeat sequence derived from the riverine species, *G. p. palpalis.*

"If we knew what it was we were doing, it would not be called research, would it?"
Albert Einstein

Chapter 2. Identification of major proteins in the midgut of the tsetse, *Glossina morsitans morsitans*

1. Introduction

Although Dr. David Bruce first identified tsetse as the vector of African sleeping sickness in 1895 (Bruce *et al.*, 1909), little progress has been made to eradicate the disease from Africa. Measures for controlling tsetse populations are used less than they were several decades ago (Allsopp, 2001) and vaccine work targeting the causative parasite, the African trypanosome, although showing renewed promise (Pearson, 2001), has yet to come to fruition. There is an urgent need to adopt a new strategy for disease control, one that considers the role of the tsetse during trypanosome development and maturation. The identification of molecules expressed by both parasite and vector during the transit of the trypanosome through the insect stages of the life cycle may allow the design of new control strategies. As a first step in this process, one of my goals was to identify tsetse midgut molecules involved in either tsetse survival or the establishment and differentiation processes involved in the transmission of pathogenic African trypanosomes.

Upon ingestion by a non-refractory fly, bloodstream forms (BSF) of trypanosomes travel to the midgut and undergo transformation to procyclic forms, which then rapidly proliferate. Ensuing migration pathways (*en route* to the salivary glands) through or around the peritrophic membrane barrier and into the ectoperitrophic space in the midgut lumen have been proposed, though the processes controlling either active penetration or passive uptake have yet to be defined (Leak, 1999). In the *Trypanosoma*

brucei group, mature infections are established when the salivary glands are populated by epimastigote forms which become anchored to the epithelium and subsequently transform into metacyclic trypomastigotes that express variant surface glycoprotein coats (Vickerman, 1985). It is this infective salivarian stage that is transported with the tsetse saliva into a vertebrate host during feeding.

The establishment of midgut-adapted trypanosomes involves midgut factors that are thought to mediate parasite transformation, growth and survival. A study by Nguu *et al.*, (1996) demonstrated that BSF trypanosomes transformed into procyclic forms *in vitro* when incubated with midgut homogenates from non-teneral tsetse. In addition, several midgut proteins are thought to influence parasite survival. These include lectins (Welburn *et al.*, 1989; Lehane and Msangi, 1991), trypanolysins and proteolytic enzymes (Endege *et al.*, 1989; Imbuga *et al.*, 1992). Six proteolytic enzymes were partially purified (Cheeseman and Gooding, 1985) but these were not described in biochemical detail. However, two genes encoding midgut serine proteases (Yan *et al.*, 2001b) and genes encoding a cathepsin and two zinc-metalloproteases were recently sequenced (Yan *et al.*, 2001a).

The idea that midgut molecules function during the developmental stages of vector-borne parasites is supported by a recent publication describing anti-*Anopheles* midgut monoclonal antibodies (mAbs) that, in an artificial membrane feeding system, markedly reduced both vector competence and survival rates (Lal *et al.*, 2001). However, the tsetse digestive tract differs from that of the mosquito in that it hosts two species of symbiotic bacteria that have been reported to affect refractoriness for trypanosome

infections in both teneral and fed flies (Maudlin and Ellis, 1985; Maudlin *et al.*, 1990). Thus tsetse midgut proteins may be of both eukaryotic and prokaryotic origin.

Although molecular biology techniques have been useful for determination of gene expression and thus indirectly in identification of proteins present in a cell, the often poor relationship between messenger RNA levels and levels of protein (Anderson and Seilhamer, 1997; Aebersold and Goodlett, 2001) can lead to false assumptions about the amounts of expressed proteins. Unfortunately, direct identification and characterization of proteins present in only small amounts of materials, such as the tsetse midgut, or those with highly similar chemical characteristics are not easily performed by standard protein chemical techniques. For this reason I have used protein microchemical techniques involving mass spectrometry to identify major proteins in midguts of teneral *Glossina morsitans morsitans.* Using this approach, I unexpectedly identified the predominant midgut protein in teneral *G. m. morsitans* as a microbial chaperonin.

2. Experimental procedures

2.1. Tsetse

The *G. m. morsitans* Westwood, from which midguts were obtained, were from several subcultures maintained at the University of Alberta. All of these colonies descended from material originating near Kariba, Zimbabwe. For a brief history of the colonies, see Gooding and Jordan (1986). Throughout my experiments, the colonies were maintained at 24.5YC and approximately 60% relative humidity. Tsetse flies were fed on rabbits every other day using a protocol that conformed to guidelines of the Canadian Council on Animal Care.

Intact midguts, bacteriomes (formerly called mycetomes) and midguts without bacteriomes were prepared from mixtures of female and male *G. m. morsitans*. Teneral (i.e. newly eclosed, unfed) adults aged 24 to 48 hours were placed in a glass shell vial and immobilised by cooling on ice for 30 minutes. After removing the wings and legs, the flies were submersed in sterile saline (0.85% NaCl) and positioned ventral side up under a dissecting microscope. Incisions were made along the sides of the abdomen, using fine tipped surgical scissors, to expose the digestive tract. To isolate intact midguts, the entire organ was gently teased free of adhering fat body and severed immediately posterior to the crop and anterior to the proctodaeum. In separate preparations, the midgut region containing the bacteriome was severed from the remainder of the midgut and both fractions saved in separate tubes. With all preparations, excised tissues were removed from the dissecting tray and excess saline blotted away, taking care to minimise loss of midgut contents. After transferring the midgut tissues to sterile Eppendorf microcentrifuge tubes (pre-cooled on dry ice) and freezing, the tubes were stored at –80°C, shipped on dry ice to the University of Victoria and then stored at –80°C until used for protein fractionation.

2.2. Procurement and axenic culture of Sodalis glossinidius

S. glossinidius were isolated at the University of Alberta according to a protocol modified from that of Welburn *et al.* (1987). In brief, six teneral *G. m. morsitans* Westwood, line 231, were surface sterilised using a series of washes with bleach, ethanol and sterile water. The head (posterior to the brain) was punctured using a sterile stainless steel pin and after removing 2-4 µl of hemolymph from each fly, the pooled hemolymph

was added to a single 5 ml culture tube previously seeded with *Aedes albopictus* cells (cell line C6/36). After centrifuging the pooled hemolymph and cells for 3 min at 1800 *x g*, the supernatant was discarded, thus removing inhibitory melanization factors in the hemolymph. Pelleted cells were resuspended in 0.5 ml serum-free Mitsuhashi and Maramorosch Insect (MMI) medium (Sigma-Aldrich Canada, Oakville, ON) and transferred to a non-vented T25 flask (Corning, Cambridge, MA) seeded with a layer of *A. albopictus* cells and incubated in the dark at 27YC for 20 days. An axenic culture was established by adding 20% of the volume of the original culture to serum-free MMI media. Serum-free axenic cultures, in establishment since 02-2000, were maintained by passaging every 4-5 days into fresh MMI medium. Culture purity was monitored by either microscopic examination (oil immersion 1000X) and the identity of the *S. glossinidius* culture was confirmed by Coomassie Brilliant Blue staining of SDS-PAGE gels that give a characteristic protein banding pattern.

2.3. One-dimensional gel electrophoresis

Midgut fractions (15 µl) were resuspended in 2X Laemmli sample buffer (Laemmli, 1970) and 25 µl of each were loaded onto 10% acrylamide, 0.75 mm 1-dimensional sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) minigels and run on a Mini-Protean II apparatus (Bio-Rad Laboratories, Hercules, CA). Proteins were electrophoresed at 50V for 30 min and then at 90V for an additional 90 min. A 10 kDa molecular weight ladder (#10064-012; Gibco BRL, Burlington, ON) was run on each gel. Following electrophoresis, gels were either stained with GelCode® Blue (Pierce

Chemical, Rockford, IL), silver stained (Merril *et al.*, 1981) or transferred onto membranes for subsequent immunoblotting or for N-terminal protein microsequencing.

2.4. *Immunoblotting*

Immunoblotting using BioTrace ™ polyvinylidene (PVDF) membrane (Pall Corporation, Ann Arbor, MI) was performed as previously described (Beecroft *et al.*, 1993) except with SuperSignal Dura chemiluminescence substrate (Pierce Chemical Company, Rockford, IL) a 1:500 dilution of primary antibody (mouse anti-GroEL; SPA-807, Stressgen Biotechnologies, Victoria, BC) and a 1:50,000 dilution of secondary antibody (goat anti-mouse IgG/IgM-horseradish peroxidase conjugate; Caltag Laboratories, South San Francisco, CA). Kodak Biomax MR film (Eastman Kodak Company, Rochester, NY) was used to detect chemiluminescence. After development of the autoluminograms, proteins were stained on the PVDF membrane with GelCode® Blue. The exposed film was superimposed on the stained PVDF membrane to reveal the precise location of the immunoreactive protein bands in relationship to the entire protein profile.

2.5. N-terminal protein microsequencing

Amino acid sequencing was performed by Edman degradation using a Perkin-Elmer-Applied Biosystems 475 gas-phase sequencer (Foster City, CA). Samples electroblotted onto BioTrace ™ PVDF membrane were stained with GelCode® Blue Stain (Pierce Chemical, Rockford, IL) and dark-staining protein bands were excised and the membrane pieces placed directly into the microsequencer. N-terminal protein

sequence information was submitted to the online FASTA (version 3.3t) sequence analysis program (European Bioinformatics Institute, http://www2.ebi.ac.uk/) for searching of the Swiss-Prot database.

2.6. Gene cloning and sequencing

Organisms from 1.0 ml of a 5-day *S. glossinidius* axenic culture (A_{600} = 0.20) were washed twice with sterile phosphate buffered saline (PBS) and genomic DNA extracted using phenol-chloroform as previously described (Maloy, 1990). Two *Sodalis*-specific primer sets, GP01 F/R (plasmid DNA) and RLO1 F/R (genomic DNA) designed by O'Neill *et al.* (1993) and Welburn and Dale (1997) respectively, were used to confirm the identity of the isolated *Sodalis* DNA. To amplify any GroEL-like sequences from *Sodalis*, primer sequences were derived from the 5'-terminal GroEL-like nucleotide sequence of the *Sitophilus oryzae* primary endosymbiont (Charles *et al.*, 1997) and the 3'-terminal nucleotide sequence from multiple alignment information (Clustal W; http://www.ebi.ac.uk/clustalw/) on GroEL sequences from selected Enterobacteriaceae. The software program, Primer Premier (PREMIER Biosoft International, Palo Alto, CA) allowed optimisation of non-degenerate primer specificity.

Polymerase chain reactions (PCR) were performed with a thermal cycler (GeneAmp PCR System 2400; Perkin-Elmer, Norwalk, CN) for 35 cycles at 94YC for 30 sec, 65YC for 45 sec, and 72YC for 2 min, followed by a 10 min extension at 70YC. The amplified 1.6 kb PCR product was gel purified using a QIAEX II Gel Extraction Kit (Qiagen Inc., Mississauga, ON) and cloned into *E. coli* TOP 10 cells (TOPO cloning Kit; Invitrogen, Carlsbad, CA) for blue/white screening without IPTG induction. Ten

transformants were grown overnight in Luria-Bertani broth supplemented with ampicillin (100 μg/ml) and plasmid purification performed using a Qiaprep Spin MiniPrep Kit (Qiagen Inc.) according to the instructions of the manufacturer. Of the ten clones, four were submitted to the Centre for Environmental Health, University of Victoria, for automated dideoxynucleotide sequencing. *S. glossinidius* GroEL-like sequence data were submitted to the GenBank database as accession number AF404511.

2.7. Two-dimensional gel electrophoresis

High-resolution two-dimensional SDS-PAGE was performed using the ISO-DALT multiple 2-D system (Anderson and Anderson, 1978a; Anderson and Anderson, 1978b) as previously described (Anderson *et al.*, 1985). Five midguts, five bacteriomes and five midguts with bacteriomes removed were separately solubilized in 30 μl of urea mix (9 M urea, 4% NP-40 (v/v), 2% Pharmalyte 3-10 ampholines (v/v), 2% DTT (w/v)) and mechanically disrupted using a 1 cc syringe equipped with a 21G needle. Samples were centrifuged for 30 sec at 10,000 *x g* to remove insoluble material and subsequently loaded onto pre-focused tube gels containing pH range 3-10 ampholines (Pharmalyte 3-10, Amersham Pharmacia, Upsala, Sweden). First dimension isoelectric focusing was conducted at 800 V for 18h (14,400 Vh). Following electrophoresis the tube gels were equilibrated for 15 min at room temperature in equilibration buffer and immediately mounted onto 10-16.5% gradient SDS-PAGE slab gels with the acidic end positioned to the left. Electrophoresis was performed at 4YC at 1 Amp until the dye front was about 1 cm from the bottom of the gel (about 5 hours). After electrophoresis the gels were fixed and then stained with colloidal Coomassie Brilliant Blue G-250.

2.8. Staining of gels with Coomassie Brilliant Blue G-250

Gels were agitated gently in fixative (50% (v/v) ethanol, 3% (v/v) ortho phosphoric acid) for 1-4 days at room temperature, washed three x 30 min in distilled water and allowed to equilibrate in Neuhoff's solution (16% (w/v) ammonium sulphate, 25% (v/v) methanol, 5% (v/v) ortho phosphoric acid (Neuhoff *et al.*, 1988) for one hour with gentle agitation. One gram of CBB G-250 (EM Science, Gibbstown, NJ) was sprinkled into the Neuhoff's solution and staining continued for 3-5 days. Once well-stained protein spots were visible, gels were either scanned and protein spots cored or the intact gels were transferred into a 20% (w/v) ammonium sulphate solution for storage at 4°C (Neuhoff *et al.*, 1988).

Digital images of both 1-D and 2-D CBB stained gels were captured by scanning at 300 dpi using a colour scanner (UMAX Astra 3400, Fremont, CA) after briefly rinsing the gels in distilled water. The images were manipulated and stored as JPEG and/or TIFF files using Photoshop™ 5.5 graphic software (Adobe Systems Inc., San Jose, CA).

2.9. Reduction, alkylation and tryptic digestion of 2-D gel spots

Protein spots of interest were cored using a 4 mm plastic straw and either transferred to 1.5 ml Eppendorf microcentrifuge tubes (previously autoclaved and rinsed with 50% methanol to remove any contaminants) for digestion or to 96 well sterile tissue culture plates (one spot per well in 10 µl of 20% (w/v) ammonium sulphate for storage at –20°C. For analysis by mass spectrometry, 2-D protein spots were de-stained (50% (v/v) methanol/5% (v/v) acetic acid), reduced with 10 mM DTT and alkylated with 100 mM

iodoacetamide as described previously by (Kinter and Sherman, 2000). The carboxyamidomethylated protein spots were digested overnight at 37°C with 20 ng/μl modified, sequence grade, porcine trypsin according to the manufacturer's directions (Promega, Madison, WI). Peptides were extracted from the gel pieces using a series of elutions with 50% (v/v) acetonitrile and 5% (v/v) formic acid. The resulting pooled eluates were each reduced to a final volume of 20 μl in a Speed Vac Concentrator (Savant, Hicksville, NY) and processed for mass spectrometry.

2.10. Q-TOF mass spectrometry

A capillary liquid chromatography system (CapLC system; Waters Corp., Milford, MA) was interfaced to a quadrupole-time-of-flight (Q-TOF) mass spectrometer (Q-Tof, Micromass, Beverly, MA) operated using Masslynx 3.4 software. System calibration was performed daily using the MS/MS spectrum of $[Glu]^1$ -fibrinopeptide B. Trypsinized peptides were loaded onto the column and separated at 300 nl/min using a 10 min, 2-42% acetonitrile gradient. Analyses of liquid chromatography profiles and Q-TOF peptide mass spectra were executed using the computer algorithms SEQUEST (Finnigan MAT, San Jose, CA) and/or Mascot (Matrix Science Ltd, London, UK) on the non-redundant (NR) protein database supplied by the National Centre for Biotechnology Information (NCBI).

2.11. MALDI-TOF mass spectrometry

The digest supernatant was mixed with analyte solution (α-cyano-4-hydroxycinnamic acid; Aldrich, Milwaukee, WI) and spotted onto a Voyager 100

position, stainless steel MALDI plate (Applied Biosystems, Foster City, CA). After drying, the plate was inserted into an Applied Biosystems Voyager DE-STR and mass spectrometry data was acquired in delayed extraction, reflectron mode. The masses of the observed tryptic peptides were submitted to MS-Fit (Protein Prospector software package; http://prospector.ucsf.edu/) and Mascot (Matrix Science; http://www.matrixscience.com/) to perform the peptide mass fingerprint searches.

3. Results

3.1. Analysis of tsetse midgut proteins by gel electrophoresis

Separation of solubilized midgut proteins by 1-D gel electrophoresis resolved a large number of proteins ranging from less than 20 kDa to greater than 200 kDa. (Figure 2.1, Panel A, lane 1). These proteins likely include any that may have been secreted into the lumen of the midgut since care was taken not to remove any loosely bound molecules during midgut dissection, washing, freezing and solubilization. Degradation appeared to be minimal since a number of high molecular weight proteins are clearly visible and the stained protein bands throughout the entire molecular weight range are sharp. Major proteins of 26, 28, 32-34, 42, 48, 52, 58-60 and 78 kDa were predominant.

Analysis of midgut proteins by high-resolution 2-D gel electrophoresis (Figure 2.1, Panel B) also revealed major protein spots of 26, 28, 32-34 and 58-60 kDa, in line with the 1-D gel banding pattern. In addition, relatively abundant lower molecular mass spots of 15 and 18 kDa were present that were not resolved in the non-gradient 1-D gel. It is interesting that the major 42 and 48 kDa proteins seen as bands in the 1-D gel pattern were not clearly visible in the 2-D gel implying they were comprised of multiple proteins

with the same approximate molecular mass but with different isoelectric points. The preponderant proteins on the 2-D map appeared as a major set of slightly acidic spots at 60 kDa (arrow). Surprisingly, in the 1-D gel profile, the 60 kDa protein, although major, does not stand out from the other abundant protein bands. As the most prolific protein in the 2-D profile, I selected the 60 kDa protein for identification by mass spectrometry.

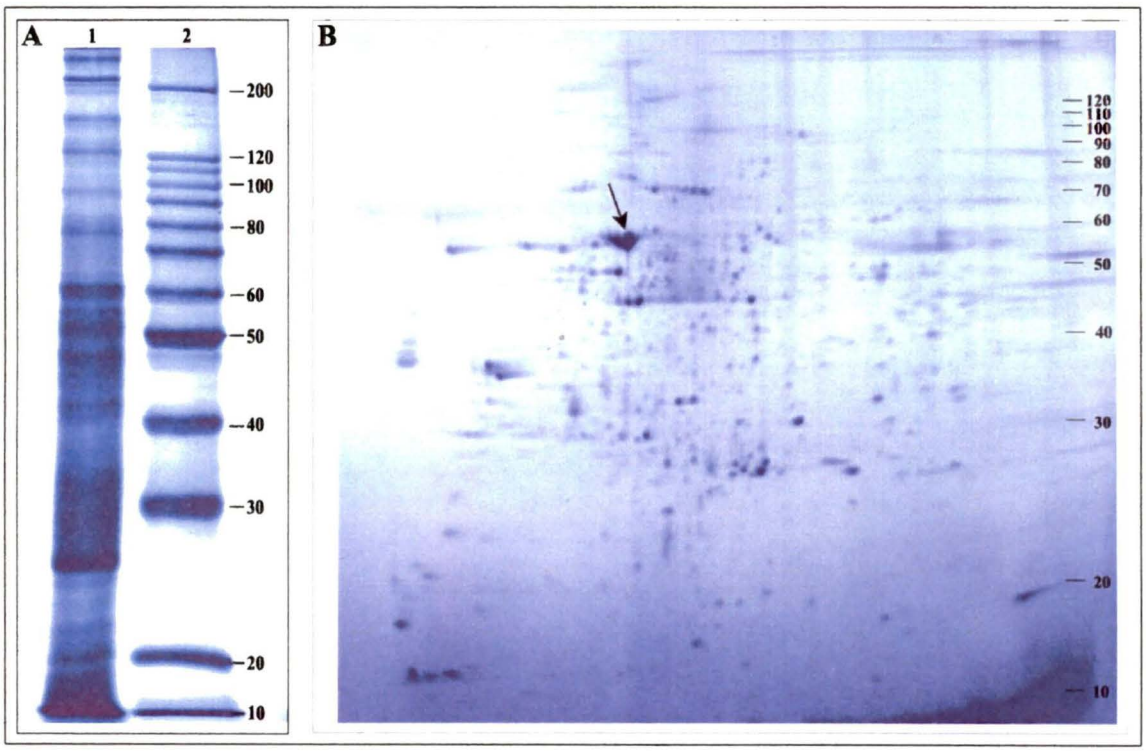


Figure 2.1. ***Protein profiles of whole midgut from teneral <u>Glossina morsitans morsitans</u>.***

Panel A: A single midgut was solubilized in 40 µl of Laemmli buffer and 20 µl of this mixture was separated in a 10% gel. Proteins were stained using colloidal Coomassie Brilliant Blue G-250. Lane 1, midgut lysate; Lane 2, molecular mass standards (10kDa ladder). Panel B: Two-dimensional polyacrylamide gel analysis of whole midgut from teneral <u>G. m. morsitans</u>. Proteins were stained using colloidal Coomassie Brilliant Blue G-250. The arrowhead points to the major 60 kDa protein excised for mass spectroscopy. Wide range ampholytes, pH 3-10, were used in the first dimension and a 10-16.5% gradient gel in the second dimension. The gel is shown with the acid end to the left.

3.2. Q-TOF mass spectrometry of the major midgut proteins

The major 60 kDa protein spots (Figure 2.1, Panel B, arrow) were destained, reduced, alkylated and digested with modified porcine trypsin to generate peptides for separation by capillary liquid chromatography and Q-TOF mass spectrometry. Peptide masses were used to probe protein databases and the search identified 5 peptide masses (Figure 2.2, blue highlight), which represented 8% coverage of the protein, a GroEL-like chaperonin of the weevil, *Sitophilus oryzae*, (NCBI accession number gi|7443844). The SymL protein (symbionin) from the aphid *Myzus persicae* (NCBI accession number gi|2754808) was also identified. These insect-derived sequences were expressed by the endosymbiotic bacteria *S. oryzae* principal endocytobiote (SOPE) and *Buchnera aphidicola,* respectively, and consequently led us to consider the symbionts of tsetse as the source of the 60 kDa tsetse midgut protein.

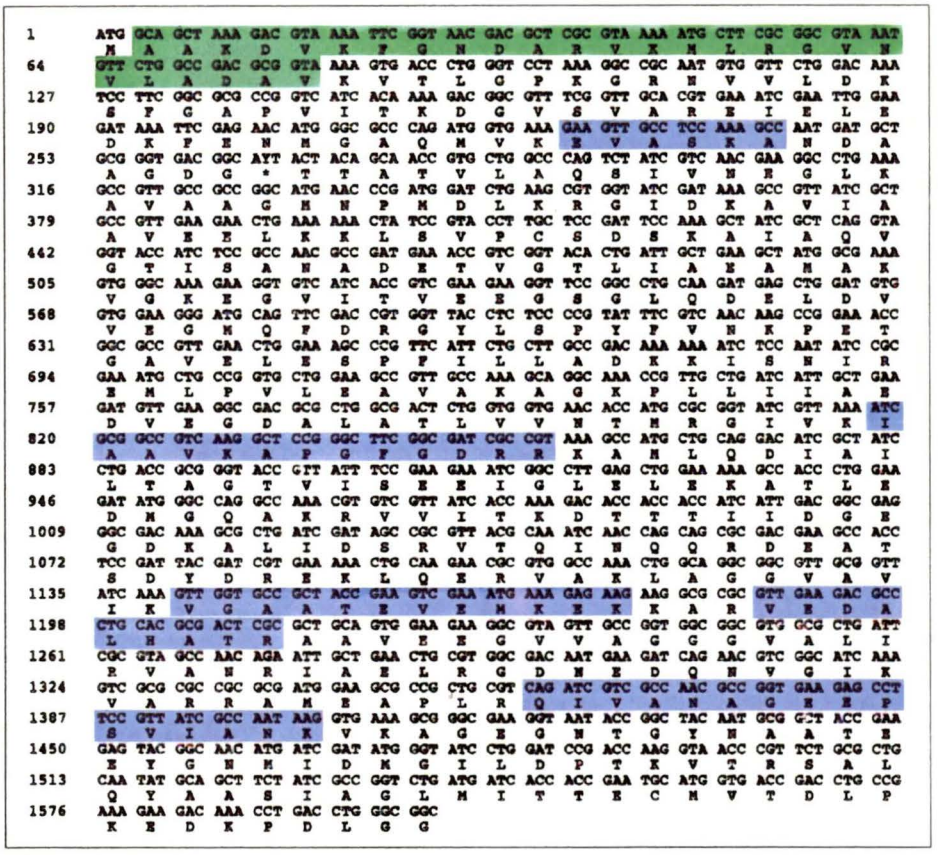


Figure 2.2. ***The DNA sequence and the translated protein sequence of the <u>Sodalis glossinidius</u> GroEL homologue.***

The DNA for sequencing was obtained by PCR-amplification of <u>S. glossinidius</u> genomic DNA. In addition, the N-terminal protein sequence (green highlight) was determined by gas-phase protein microsequencing and the five tryptic peptide sequences (highlighted in blue) were derived by Q-TOF mass spectrometry. The numbers on the left correspond to the nucleotide position.

3.3. Identification of major proteins of the tsetse symbiont Sodalis glossinidius

Of the three known symbiotic bacteria of tsetse only two inhabit the midgut, *Sodalis glossinidius* and *Wigglesworthia glossinidia* (Aksoy, 2000) and only the former can be cultured *in vitro*. This organism was therefore grown in serum-free medium in axenic culture and the proteins were separated by 1-D gel electrophoresis, transferred to PVDF membrane and stained with GelCode® Blue. The results are shown in Figure 2.3.

The most predominant protein bands were 35 kDa and 60 kDa (lower and upper arrows, respectively). Protein bands of 48-50 kDa were also present but did not appear to be abundant. N-terminal protein microsequencing of the 35, 50 and 60 kDa proteins after electroblotting to Immobilon-P^{TM} membrane was attempted.

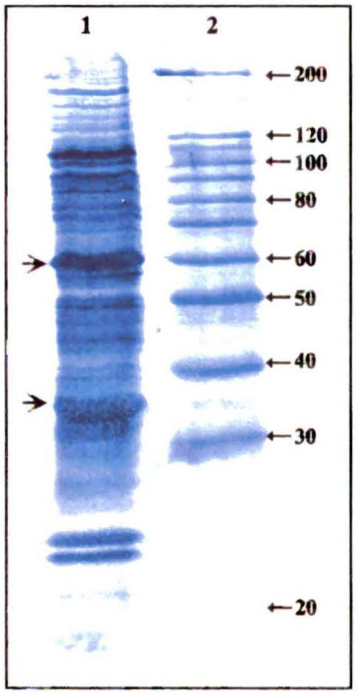


Figure 2.3. ***N-terminal sequencing of two <u>Sodalis glossinidius</u> proteins.***
The bacteria were grown in axenic culture in serum-free medium and washed twice in phosphate-buffered saline before solubilization. Proteins were stained using GelCode® Blue. Lane 1, <u>S. glossinidius</u> lysate (cells from 1.0 ml of culture at an OD_{600} of 0.2); Lane 2, molecular mass standards (10 kDa ladder). The arrows point to the 60 kDa and 35 kDa proteins used for N-terminal protein microsequencing.

No sequence was obtained for the 50 kDa protein as it was blocked at the N-terminus. Sequences for the 35 and 60 kDa proteins are shown in Table 2.1.

Table 2.1. *Amino-terminal amino acid sequences[a] of the major 60 kDa and 35 kDa proteins of Sodalis glossinidius.*

60 kDa:	A A K D V K F G N D_{10} A R V K M L R G N V_{20} L A D A V K
35 kDa:	A E V Y N K D G N K_{10} L D I Y G K A V G L_{20} H Y F S

[a]Gas-phase, N-terminal protein microsequencing was performed after blotting of the proteins shown in Figure 2.3 onto BioTrace™ membrane.

Searching (FASTA, European Bioinformatics Institute) of the protein database (SwissProt) showed that the 26 residue sequence of the 60 kDa protein matched the highly conserved N-termini of the 60 kDa chaperonins of many different bacteria. The 24 amino acid sequence of the 35 kDa protein matched the outer membrane porin sequences of several Enterobacteriaceae species. This sequence was submitted to the Swiss-Prot protein sequence database as Accession Number P83079. The DNA sequence of the 60 kDa *Sodalis* chaperonin was determined by PCR amplification using primers specific for the N-terminal sequence and two nested tryptic peptides sequenced by Q-Tof mass spectrometry (Figure 2.4). This sequence was conceptually translated into the amino acid sequence used for peptide mass mapping (see Figure 2.5).

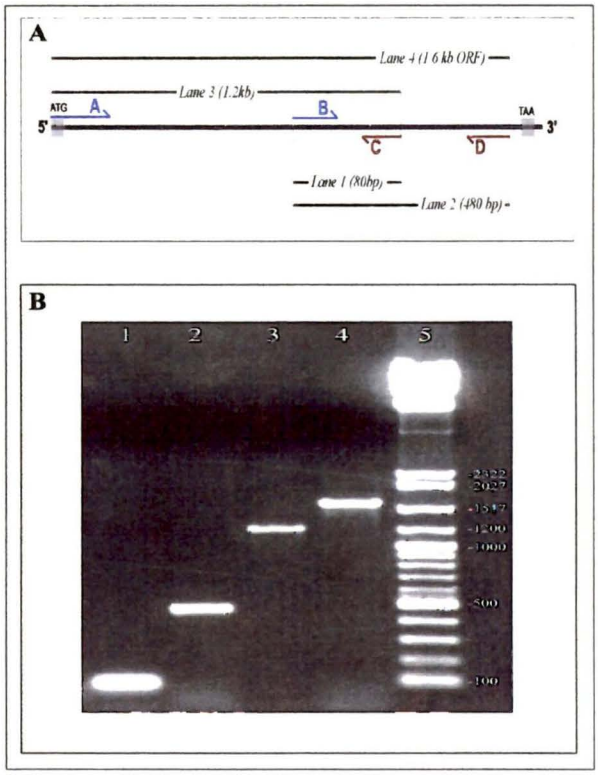


Figure 2.4. ***Polymerase chain reaction (PCR) analysis of the GroEL gene from S. glossinidius.***

Panel A: Schematic diagram of the strategy used for PCR amplification of genomic DNA encoding S. glossinidius GroEL. The predicted size of the PCR products (in base pairs) and the gel lanes used for their analysis are indicated for each primer set. Panel B: Agarose gel separation of the PCR products obtained using the primer sets shown schematically in panel A. Genomic DNA from S. glossinidius was used as the template for the primer pairs (lanes 1-4). Molecular size standards (Lambda Hind III digest\|100 bp ladder) are in lane 5.

3.4. MALDI-TOF mass spectrometry of the 60 kDa major tsetse midgut protein

To determine the exact origin of the 60 kDa GroEL-like midgut chaperonin, peptide mass fingerprinting was performed and the results compared to virtual (predicted) tryptic peptide sequences of the 60 kDa chaperonin sequences of both *S. glossinidius* and *W. glossinidia.* The results are shown in Figure 2.5 below.

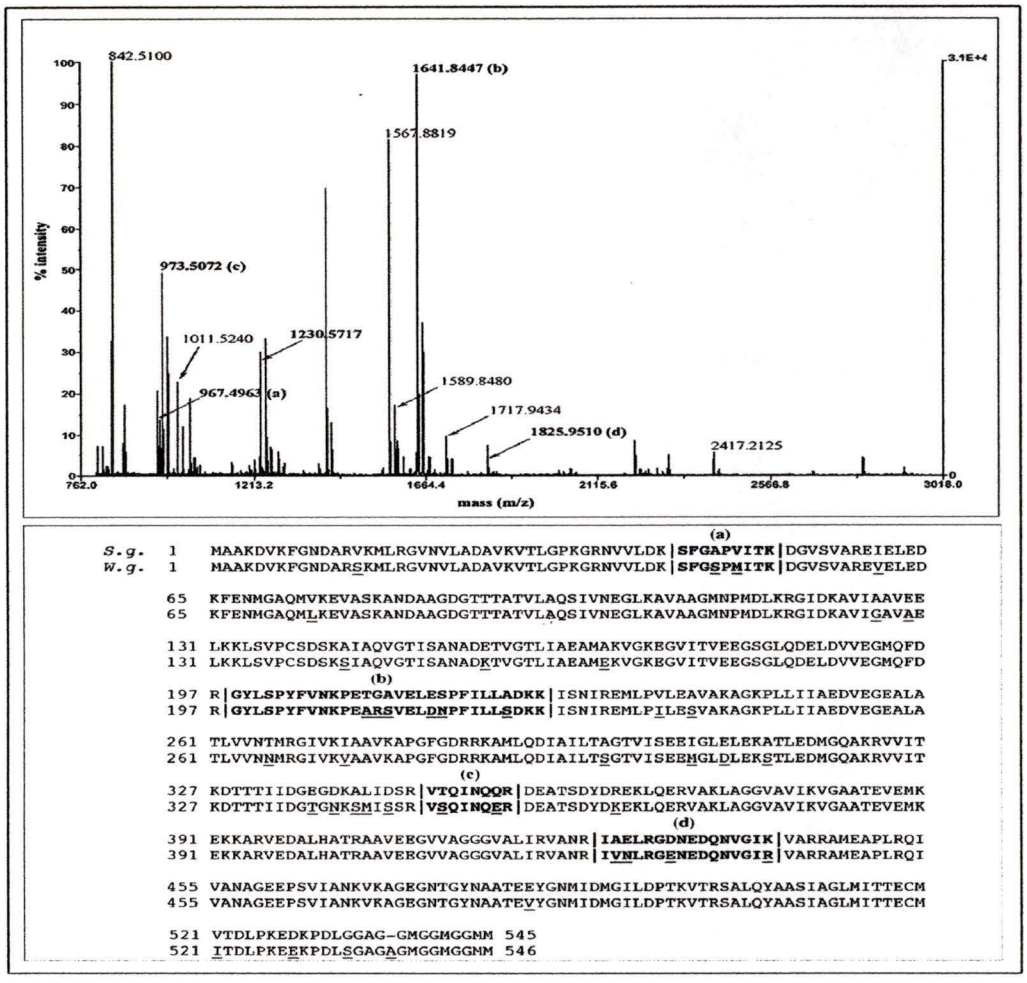


Figure 2.5. ***Peptide mass spectrum of the digested 60 kDa protein from teneral <u>G. m. morsitans</u> midgut and correlation of masses to the amino acid sequences of GroEL for <u>S. glossinidius</u> (S.g.) and <u>W. glossinidia</u> (W.g.).***

The major 60 kDa protein spots (arrowhead in Figure 2.1, Panel B) were used for tryptic digestion and MALDI-TOF mass spectrometry to obtain the peptide masses. Upper Panel: Peptide mass profile showing the major peptide peaks. For simplicity, only the masses of the major peptide peaks that matched predicted masses of GroEL peptides are shown. Although not all matching peptides are represented in the figure, 30% coverage of the total protein sequence was achieved using the Matrix Science (Mascot) algorithm and 36% coverage using the Protein Prospector (MS-Fit) algorithm. Peptides labelled a, b, c and d, are unique to <u>W. glossinidia</u>. Lower Panel: Comparison of amino acid sequences of GroEL from <u>S. glossinidius</u> and <u>W. glossinidia</u>. Amino acids that differ between the two sequences are underlined. The unique <u>W. glossinidia</u> peptides identified by mass spectroscopy are in bold and designated a, b, c and d, corresponding to the peaks designated in the spectrum in the upper panel.

Twenty-three of the 73 peptide masses used to search the NCBI nr.7.1.2001 database identified bacterial GroEL. Depending on the search algorithm used I obtained either 30% coverage (Mascot; Matrix Science, London, UK) or 36% coverage (MS-Fit; Protein Prospector, San Francisco, CA) of the protein. Four of the peptide masses were unique to the *W. glossinidia* peptide fingerprint (designated a, b, c, and d in Figure 2.5), unequivocally determining that the GroEL-like molecule was from this endosymbiont.

3.5. Comparison of bacteriome and bacteriome-severed midgut by 2-dimensional gel electrophoresis

Although both *S. glossinidius* and *W. glossinidia* are inhabitants of tsetse midgut, *W. glossinidia is* restricted to the bacteriome. I therefore compared the proteins of the bacteriome with those of the remaining midgut tissue by 2-D gel electrophoresis (Figure 2.6). It was striking that the major 60 kDa protein was restricted to the bacteriome (ellipse, Panel B).

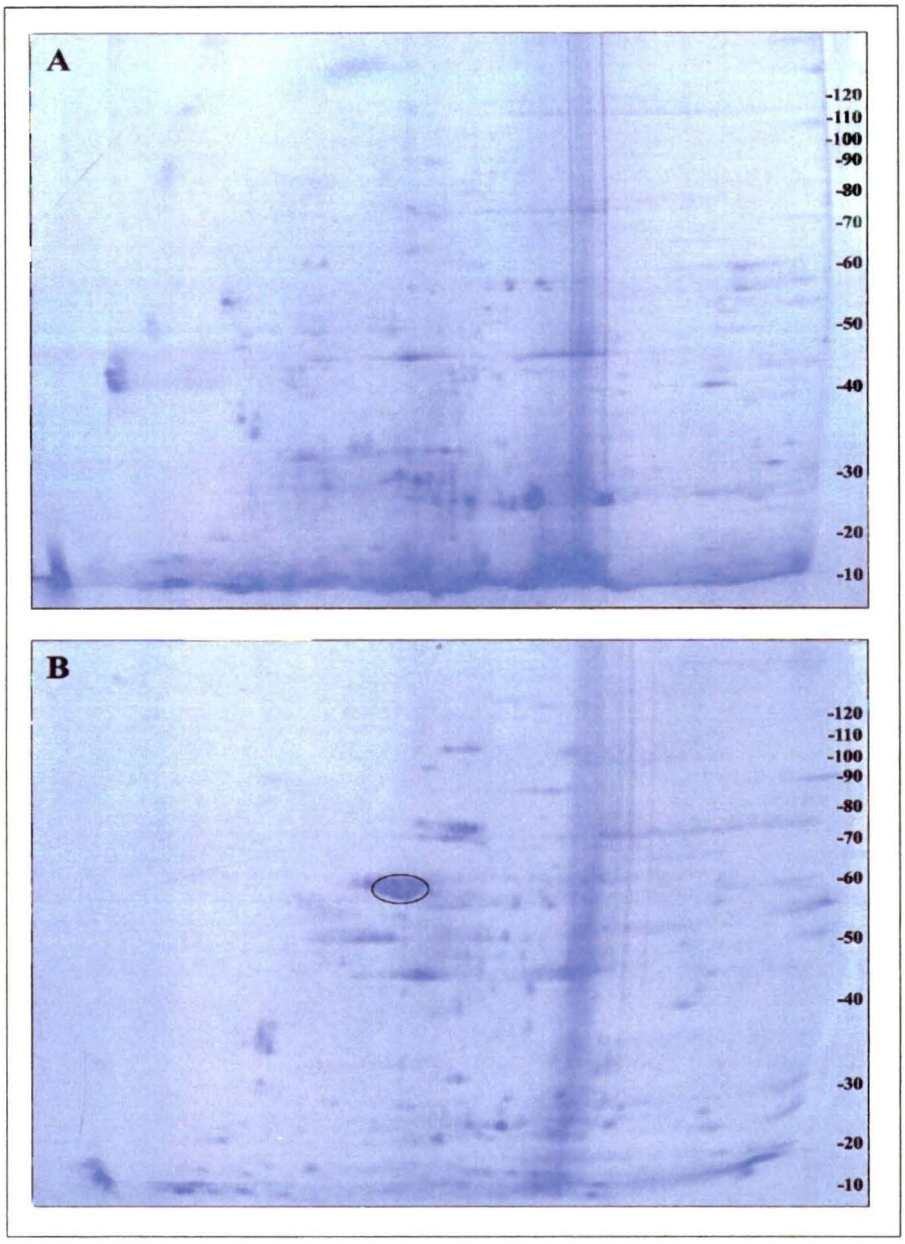


Figure 2.6. ***Two-dimensional polyacrylamide gel analysis of proteins in bacteriome and midgut lacking bacteriome from teneral <u>G. m. morsitans</u>.***

Proteins were stained using colloidal Coomassie Brilliant Blue G-250. Panel A: Midguts with bacteriomes removed. Panel B: Bacteriomes. The <u>W. glossinidia</u> GroEL position is within the ellipse. Its identity was confirmed by MALDI-TOF mass spectrometry. Both gels were run simultaneously using the ISO-DALT system and are shown with the acid end to the left.

3.6. Immunodetection of the 60 kDa chaperonin in bacteriome and midgut lacking bacteriome

To detect GroEL–like proteins in the two midgut fractions, immunoblotting was performed using an anti-GroEL mAb of known specificity as a probe. The anti-Hsp60 mAb that I used recognises an epitope that is conserved in both eukaryotes and prokaryotes. The results are shown in Figure 2.7.

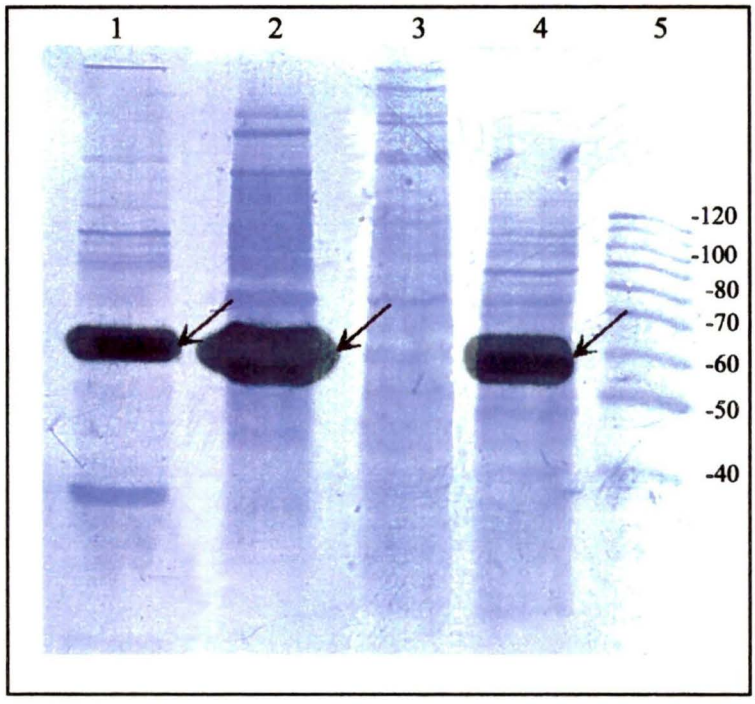


Figure 2.7. ***Immunoblot analysis of midgut proteins from teneral G. m. morsitans.*** *Proteins were separated using a 10% gel, blotted onto PVDF membrane and probed with anti-GroEL mAb SPA-807 (StressGen). The autoluminogram pattern (dark bands indicated by arrows) is shown superimposed on the same PVDF membrane which was subsequently stained with GelCode® Blue to reveal the exact location of the immunoreactive protein bands. Lane 1, S. glossinidius (positive control); Lane 2, bacteriome; Lane 3, midgut minus bacteriome; Lane 4, whole midgut; Lane 5, molecular mass standards (10 kDa ladder).*

The autoluminogram, which is superimposed over the same blot after subsequent counterstaining with GelCode® Blue, showed single, predominant, strong bands at 60 kDa in the bacteriome fraction (arrow, lane 2) and in whole midgut (arrow, lane 4). The midgut without bacteriome was negative (lane 3). As a positive control, *S. glossinidius* was loaded into Lane 1, and as expected, a single 60 kDa band (arrow), previously identified as a GroEL-like protein by N-terminal sequencing (Table 2.1) was detected.

4. Discussion

Tsetse flies, like most hematophagous insects, can simultaneously harbour multiple organisms such as bacteria, nematodes, protozoa, and viruses thus potentially complicating tsetse genome analysis (Maudlin and Welburn, 2001). Similarly, the presence of multiple organisms within certain tsetse tissues, such as the midgut, makes tsetse proteome analysis more complex. The midgut of fed flies also contains proteins from host blood and tsetse proteins induced by feeding. For these reasons, in the work reported here, I decided to focus on the midgut molecules of unfed, newly emerged (teneral) tsetse.

I set out to identify the most abundant proteins in the insect midgut and made an effort not to remove soluble molecules contained in the lumen. It was my reasoning that abundant molecules may be important for health and survival of the flies and that some of them may potentially interact with trypanosomes ingested during the first infected bloodmeal. It is an interesting phenomenon that tsetse are most susceptible to trypanosome infection as teneral flies that feed on infected blood immediately following eclosion. Otieno *et al.*, (1983) reported that the highest susceptibility to *T. b. brucei*

infection was observed when membrane-fed *G. m. morsitans* were given an infective meal 1-8 hours after emergence. This increased susceptibility suggests that the midgut either contains molecules that augment the establishment of a mature infection (perhaps the *Wigglesworthia* GroEL?) or that there is an absence of trypanolytic factors within the midgut upon eclosion. It is of course possible that an incompletely formed peritrophic membrane in teneral flies (Lehane and Msangi, 1991) may also play a central role in this increased susceptibility.

Both total and PBS-soluble proteins in teneral tsetse midgut were separated by 1-D and 2-D gel electrophoresis. Although both procedures revealed proteins that ranged from less than 20 kDa to greater than 200 kDa, it was clear that 2-D separation gave a better indication of the amount of individual constituents of the midgut proteome. This was due to the separation, by isoelectric point differences, of proteins of the same apparent molecular mass that appeared as single bands on the 1-D gel profiles and to the enhanced resolution achieved by gradient gels in the second dimension. Thus the 2-D profiles clearly revealed the 60 kDa protein as the most abundant protein in the midgut. I therefore selected this predominant protein for identification by Q-TOF mass spectrometric analysis. Database searching using peptide mass maps obtained after in-gel tryptic digestion showed that several peptides matched those of a GroEL-like protein isolated from a weevil endosymbiont, *Sitophilus oryzae* principal endocytobiote (SOPE) as well as SymL from an aphid symbiont, *Buchnera aphidicola*. Biosynthetic protein labelling experiments demonstrated that greater than 40% of the *ex vivo* protein synthesis was directed to the GroEL-like protein produced by the endosymbiont (Charles *et al.*, 1997) further supporting my observations of GroEL over-expression.

That the major tsetse soluble midgut protein was most closely related to GroEL-like chaperones of bacteria was surprising. However, two Gram-negative symbionts are known to reside within the tsetse midgut: the primary endosymbiont *Wigglesworthia glossinidia* and the secondary symbiont *Sodalis glossinidius*. Interestingly, immunological characterization of these symbionts (isolated from non-teneral flies) suggested that abundant proteins synthesized by both were chaperonins (Aksoy, 1995a). Since my preliminary mass spectrometry results indicated that the 60 kDa protein was of bacterial origin, I decided to ask the question, is this midgut protein indeed a microbial chaperonin or is it produced by the tsetse? To date, no tsetse heat shock protein or chaperonin has been identified or biochemically characterized.

All *Glossina* sp. harbour *Wigglesworthia*, an obligate endosymbiont that resides within a specialized midgut organelle, known as the bacteriome (Aksoy, 1995b; Aksoy *et al.*, 1995). These bacteria are implicated in the metabolism of B-complex vitamins essential for tsetse survival (Nogge, 1981). In contrast, the symbiont *Sodalis* is found in a range of host tissues, including extracellularly in the midgut lumen and intracellularly in the midgut epithelium (Aksoy *et al.*, 1997; Dale and Maudlin, 1999). Unlike *Wigglesworthia*, a species-specific and age-specific population density has been observed with *Sodalis* (Cheng and Aksoy, 1999; Maudlin, 1990). However, Welburn and Gibson (1989) have previously reported stable *Sodalis* populations in teneral *G. m. morsitans*, the species I have chosen to investigate. Based on a recent analysis of the *Sodalis* genome (Akman *et al.*, 2001), a role for *Sodalis* in the biosynthesis of cofactors and vitamin metabolites has also been proposed, although these bacteria, unlike *Wigglesworthia*, may not be necessary for tsetse survival. Of great interest, however, is that there are several

reports of increased trypanosome infections in flies with high *Sodalis* (previously referred to as rickettsia-like organism) loads (Maudlin and Ellis, 1985).

Because it could be cultured *in vitro* in the absence of host cells or feeder cells, I first focused on *Sodalis glossinidius* in my pursuit of endosymbiont 60 kDa chaperonin. My objective was to determine the sequence of the *Sodalis* GroEL gene product in order to predict the tryptic peptide masses for mass spectrometry analysis. To do this I isolated the *Sodalis* GroEL gene, determined the DNA sequence and translated it into an amino acid sequence. Multiple alignments of GroEL protein sequences of enteric bacteria show high sequence identity in the N-terminal apical domains and in the C-terminal equatorial domains, permitting the design of both forward and reverse primers for use in gene isolation by polymerase chain reaction. In addition, the codon usage in the sequences from the multiple alignment was appropriately biased for *Sodalis*, as judged by considering the few previously submitted GenBank *Sodalis* sequences. Using these primers, a GroEL-like gene from *Sodalis* was successfully cloned and sequenced. This sequence was aligned with the GroEL sequences of *Wigglesworthia* (from *G. brevipalpis*) and *Sodalis* (*G. m. morsitans* isolate) that had been independently sequenced by one of our collaborating laboratories (Yale). These were subsequently submitted to the NCBI GenBank (accession numbers AF321516 and AF321517). Hypothetical tryptic peptides were generated from the translated DNA sequences and compared to the tryptic peptide masses generated by MALDI mass spectrometric analysis of the in-gel digested proteins. The GroEL from *Wigglesworthia glossinidia* was unambiguously identified, by the presence of four unique peptides, as the predominant molecule within the midgut of *G. m. morsitans*. This result was subsequently confirmed by comparing the 2-D gel patterns of

midguts (which may contain *Sodalis*, but not *Wigglesworthia*) to dissected bacteriomes (containing only *Wigglesworthia*). Only the bacteriomes contained the 60 kDa protein. Finally, immunoblotting with an anti-GroEL monoclonal antibody confirmed that the 60 kDa GroEL-like protein was present only in the bacteriome and not in the rest of the midgut, again indicating that it originated from *Wigglesworthia.* It is surprising that *S. glossinidius* GroEL was not detected in the midgut lacking the bacteriome by either 2-D gel electrophoresis or by the more sensitive immunoblotting technique. Perhaps the population of this microbe in this strain of teneral *G. m. morsitans* is too low to allow the detection of GroEL. It will be interesting to examine the midguts of fed flies to see if the numbers of *S. glossinidius* change after feeding, particularly since it has been reported that there is a correlation between the number of these symbionts and susceptibility to trypanosome infection (Welburn and Maudlin, 1991).

GroEL (also known as Hsp60) belongs to a well-characterized sequence-related group of molecular chaperones that are constitutively expressed but are also stress inducible. Members of this group of chaperones are involved in assembly of oligomeric protein complexes, folding and refolding of nascent polypeptides, and transmembrane transport of newly synthesized proteins (Filichkin *et al.*, 1997). For a comprehensive review of chaperonin structure and function consult Ranford *et al.* (2000). GroEL is essential for cell survival and growth, as *GroE* operon deletions are lethal in *E. coli* (Fayet *et al.*, 1989). This is likely to be the case for *Wigglesworthia* as well. The question remains, what is the role of this highly expressed protein in the midgut of *G. m. morsitans*? There are several plausible reasons why symbiont chaperonins are overexpressed in insect-symbiont systems. First, they may be required for stress-induced

chaperonin activity necessary for survival in a hostile environment such as that of the tsetse midgut with its array of proteases, changes in protein compositions and temperature shifts. Second, in other obligate, mutualistic and pathogenic systems involving endosymbionts, non-chaperonin activities have been observed. These include: prevention of aggregation of and /or disassembly of invading microbes (Filichkin *et al.*, 1997), protection from proteolytic degradation (Evans *et al.*, 1992), mucus-mediated interaction with host cells (Ensgraber and Loos, 1992) and adherence to (Frisk *et al.*, 1998) or invasion of host cells (Garduno *et al.*, 1998). In this regard, it may be of benefit to *Wigglesworthia,* which has a significantly reduced genome and thus is absolutely dependent on the host (Akman and Aksoy, 2001), to produce large amounts of the chaperone in order to transport host proteins across its membrane. In addition, the endosymbiont may secrete the chaperone. This could bind any anti-bacterial peptides or proteins produced by the tsetse thus protecting the microbes from the insect immune system. Indeed, GroEL has previously been observed to bind anti-bacterial peptides secreted by insects (Otvos *et al.*, 2000). Third, and perhaps most interesting, is the recent discovery that *Enterobacter aerogenes,* normally an enteric pathogen, exists as an endosymbiont within the larval salivary glands of the ant-lion *(Myrmeleon bore).* Here, *E. aerogenes* produces a GroEL that functions as an insect toxin to help paralyse the ant-lion's insect prey (Yoshida *et al.*, 2001). This GroEL contains four key residues that are crucial for toxicity but do not interfere with protein-folding capacity (Figure 2.8). Interestingly, the *Wigglesworthia* GroEL sequence includes 3 of the 4 crucial residues and the *Sodalis* GroEL contains all four.

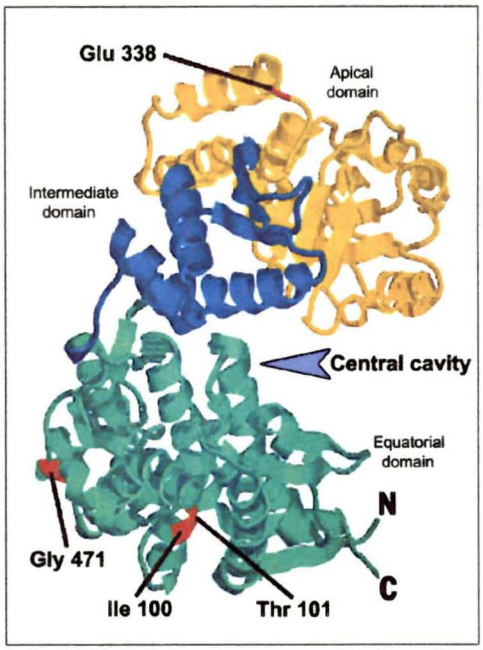


Figure 2.8. ***Three-dimensional structure of a GroEL subunit.***
Yellow: apical domain. Blue: intermediate domain. Green: equatorial domain. The four amino acids crucial for conferring toxicity are shown in red (modified from Yoshida et al., 2001).

This observation suggests the intriguing possibility that the GroEL-like proteins in *Wigglesworthia glossinidia* and in *Sodalis glossinidius* also may function as toxins. Why these organisms would require such an activity is unknown. Perhaps the amino acid differences are important for selection of trypanosomes able to establish a midgut infection.

Chapter 3. Surface expression and secretion of a chaperonin from the tsetse symbiont, *Sodalis glossinidius*

1. Introduction

Three species of symbiotic bacteria have been identified in tsetse flies (*Glossina spp.*). One of these, *Wigglesworthia glossinidia,* is the primary endosymbiont found within the bacteriome and is required for tsetse viability. Another, *Wolbachia*, is a secondary symbiont common to most arthropods. It is localized to the reproductive tissues and is known to influence reproductive abnormalities such as cytoplasmic incompatibility (CI), feminization and parthenogenesis (Hurst and Randerson, 2002). The third bacterium, *Sodalis glossinidius*, a secondary symbiont, was recently officially classified as a new bacterial taxon and species and has not been well defined (Dale and Maudlin, 1999). This Gram negative, non-motile microbe inhabits the midgut of tsetse, is maternally inherited and is a member of the family Enterobacteriaceae. *Sodalis* lives in the gut lumen, existing intracellularly in epithelial cells and as free-living forms. Reinhardt *et al.* (1972) first described *Sodalis* as small rickettsia-like organisms (RLO) that were separated from the cytoplasm of midgut cells by a clear lytic zone. Pinnock and Hess (1974) confirmed these observations and reported the presence of this pleomorphic microbe in other tissues such as the fat body and ovaries. The development of a *Sodalis*-specific PCR-assay further revealed that this symbiont resided in the midgut, hemolymph and milk gland of teneral (unfed, newly emerged) flies (Cheng and Aksoy, 1999).

Despite its detection almost 30 years ago, little is known about the biochemistry of *S. glossinidius*. *Sodalis* is microaerophilic and utilises enzymatically-digested proteins as a nitrogen source. Its primary carbon sources are both N-acetyl-D-glucosamine and raffinose. The organism does not produce catalase (characteristic of several microaerophiles) which, when combined with the results of additional phenotypic tests, indicates that *Sodalis* has a reduced biochemical profile when compared to other members of the family Enterobacteriaceae (Dale and Maudlin, 1999). *Sodalis* is relatively difficult to isolate and to culture and thus, as might be expected, only a few *Sodalis* proteins have been identified. These include an exochitinase (NCBI accession number CAA72201) and an endochitinase (Welburn *et al.*, 1993), both of which are thought to influence tsetse susceptibility to midgut infection by trypanosomes. A GroEL-like protein from *Sodalis* has been immunologically detected using a monospecific antiserum raised against the chaperonin from an aphid endosymbiont, *Buchnera aphidicola* but was not biochemically characterized (Aksoy, 1995a). Recently, 85% of the *Sodalis* genome was characterized by hybridisation to *E. coli* gene macroarrays (Akman *et al.*, 2001) and many potential gene products were identified. The use of *E. coli* arrays of course did not allow the identification of molecules unique to *Sodalis*, as exemplified by the failure to detect genes of the type III secretion system (Dale *et al.*, 2001) and genes encoding the two chitinases.

The relationship between *Sodalis* and tsetse flies is unknown although a mutualistic relationship is possible, perhaps involving the provision of metabolites. This idea is supported by evidence that genes encoding enzymes required for vitamin synthesis are found in *Sodalis* (Akman *et al.*, 2001) as in *W. glossinidia*, the tsetse primary

endosymbiont (Akman and Aksoy, 2001). Mutualism is also supported by the observations that in some tsetse species, longevity is reduced in flies lacking *Sodalis* (Dale and Welburn, 2001). In addition, puparia that are heavily infected with *Sodalis* show increased survival in adverse conditions (Baker *et al.*, 1990). Of perhaps greater interest is the hypothesis that *Sodalis* may modulate transmission of trypanosomes. Increased trypanosome infections in the tsetse midgut were reported in both lab-reared and wild tsetse possessing high numbers of *Sodalis* (Maudlin and Ellis, 1985). These investigators compared susceptible and refractory lines of *Glossina morsitans morsitans* and found that the midguts of the susceptible line were prone to the heaviest *Sodalis* infections. Supporting these laboratory observations, examination of midguts from wild flies also revealed a correlation between an increase in trypanosome infections and high numbers of *Sodalis*. Indeed, it was stated that a wild fly carrying *Sodalis* was six times more likely to be infected with trypanosomes than a *Sodalis*-free fly (Maudlin *et al.*, 1990). To confuse the issue, others did not observe a direct relationship between *Sodalis* population densities and tsetse refractoriness or susceptibility (Moloo and Shaw, 1989). Later studies on puparial development at lower temperatures revealed a significant reduction of *Sodalis* numbers in the teneral flies. The decreased symbiont population was correlated to an increased tsetse refractoriness to trypanosome infections (Welburn and Maudlin, 1991).

The suggestion that a bacterial symbiont may influence the vectorial capacity of an insect is not unprecidented. The whitefly, *Bemisia tabaci*, is a vector of the tomato yellow leaf curl (TYLC) virus and its primary endosymbiont produces large quantities of a molecular chaperone. Morin *et al.* (1999) proposed that this endosymbiotic chaperonin

bound to and protected viral particles from proteolytic degradation during passage through the whitefly midgut. When whiteflies were fed an anti-endosymbiont chaperonin antiserum prior to viral infection, the viral transmission to control plants was reduced by greater than 80%. The same group later demonstrated a direct interaction between the coat protein of the virus and the chaperonin, although the micro faunal contribution to transmission competency was not resolved (Morin *et al.*, 2000). It is therefore plausible that the African trypanosome could have evolved an elaborate transmission mechanism, similar to the viral exploitation of symbiotic chaperonin molecules, to increase survival during passage through the tsetse. Because of the potential importance in disease transmission, it is necessary to examine the complex relationships between the symbionts, trypanosomes and tsetse. In particular, since this bacterium and the trypanosomes co-inhabit the tsetse midgut, *Sodalis* may influence parasite establishment in the teneral fly, and ultimately, transmission. As a first step in understanding the *Sodalis*-tsetse-trypanosome relationship, I set out to identify surface molecules of *Sodalis*.

2. Experimental procedures

2.1. Bacteria

Escherichia coli 0157:H7, *Escherichia coli* Y1090, *Salmonella typhimurium* S736, *Shigella sonnei*, *Citrobacter freundii* 8090 and *Enterobacter aerogenes* were obtained from the culture collection of the Department of Biochemistry and Microbiology at the University of Victoria. The origin of these strains has been described previously (Feutrier *et al.*, 1986; Doran *et al.*, 1994). Stock cultures were maintained and stored on

Luria Bertani agar plates at 4°C. Working cultures were grown in 5 ml Luria broth with shaking (200 rpm) for 16 h at 37°C. *Sodalis glossinidius* was isolated, cloned and cultured as described in detail below.

2.2. Axenic culture and cloning of Sodalis glossinidius

S. glossinidius were isolated at the University of Alberta according to a protocol modified from that of (Welburn and Maudlin, 1987). In brief, six teneral *G. m. morsitans* Westwood, line 231, were surface sterilised using a series of washes with bleach, ethanol and sterile water. The head (posterior to the brain) was punctured using a sterile stainless steel pin and after removing 2-4 μl of hemolymph from each fly, the pooled hemolymph was added to a single 5 ml culture tube previously seeded with *Aedes albopictus* cells (cell line C6/36). After centrifuging the pooled hemolymph and cells for 3 min at 1800 *x* g, the supernatant was discarded, thus removing inhibitory melanization factors in the hemolymph. Pelleted cells were resuspended in 0.5 ml serum-free Mitsuhashi and Maramorosch Insect (MMI) medium (Sigma-Aldrich Canada, Oakville, ON) and transferred to a non-vented T25 flask (Corning, Cambridge, MA) seeded with a layer of *A. albopictus* cells and incubated in the dark at 27°C for 20 days. An axenic culture was established by adding 20% of the volume of the original culture to serum-free MMI medium. Serum-free axenic cultures, established since 02-2000, were maintained by passaging every 6-8 days into fresh MMI medium. Cultures were monitored by microscopic examination (oil immersion 1000X). The identity of the cultures was confirmed by 1-D gel analysis (Fig. 2A) since *Sodalis* yields a characteristic protein banding pattern after staining with colloidal Coomassie Brilliant Blue. For cloning, one

ml of a 5-day old *Sodalis* culture (0.18 OD_{600}) was transferred aseptically onto MMI plates (1% w/v agar; Difco, Detroit, MI), spread across the surface and allowed to absorb into the agar for 15 min at room temperature. The plates were placed in an unvented BBL GasPak™ anaerobic jar containing a BBL CampyPak *Plus*™ system (Becton Dickinson, Sparks, MD) to generate microaerophilic conditions and incubated at 27°C for 23-25 days or until white papillated colonies were evident. Large, single colonies (>5 mm diameter) were assessed for purity by light microscopy and confirmed as *S. glossinidius* by colony polymerase chain reaction using the *Sodalis* plasmid-specific primer set GP01 F/R designed by O'Neill *et al.* (1993). To establish pure cultures, the colonies were cored from the agar plate, added to 0.5 ml of fresh MMI medium in a sterile 15 ml screw cap tube (Sarstedt, Inc., Newton, SC) vortexed for 1 min, and the suspension incubated at 27°C for 5 days. The cultures were maintained by doubling the volume with fresh MMI medium as required.

2.3. Tsetse and isolation of bacteriomes

The *G. m. morsitans* Westwood, from which midgut bacteriomes were obtained, were from several sub colonies maintained at the University of Alberta. All of these colonies descended from material originating near Kariba, Zimbabwe. For a brief history of the colonies, see Gooding and Jordan, 1986. The colonies were maintained at 24.5°C and approximately 60% relative humidity. Tsetse were fed on rabbits every other day using a protocol that conformed to guidelines of the Canadian Council on Animal Care. Intact bacteriomes (formerly called mycetomes) were isolated from the midguts of both female and male *G. m. morsitans*. Teneral (i.e. newly eclosed, unfed) adults aged 24 to 48 hours

were placed in a glass shell vial and immobilised by cooling on ice for 30 minutes. After removing the wings and legs, the flies were submersed in sterile saline (0.85% NaCl) and positioned ventral side up under a dissecting microscope. Incisions were made along the sides of the abdomen, using fine tipped surgical scissors, to expose the digestive tract. To isolate intact midguts, the entire organ was gently teased free of adhering fat body and severed immediately posterior to the crop and anterior to the proctodaeum. Following this, the bacteriome was severed from the midgut, snap frozen in sterile Eppendorf microcentrifuge tubes (pre-cooled on dry ice) and shipped on dry ice to the University of Victoria for storage at -80°C until use. With all preparations, excised tissues were removed from the dissecting tray and excess saline blotted away to minimize protein dilution.

2.4. Electron microscopy

Sodalis glossinidius were grown in MMI medium at 27°C for 7 days. Cells were gently washed and resuspended in PBS. A Formvar (0.5%) coated copper grid was placed on a drop of cell suspension for one minute to adsorb bacteria to the surface. The grid was subsequently stained with 0.5% phosphotungstic acid (pH 7.0) for 35 seconds. A Hitachi 7000 transmission electron microscope was used at 75kV to view the sample. Photomicrographs were scanned as reverse TIFF files to enhance contrast and to decrease background intensity.

2.5. Derivation of monoclonal antibodies

Two, 6-8 week old, female BALB/c mice were immunized with *S. glossinidius*. One mouse received 0.1 ml of live organisms adjusted to an OD_{600} of 1.0 in PBS and the other, the same amount of heat-killed (100°C, 5 min) organisms. The mice were injected subcutaneously with 50 µl of the designated suspension in each of the hind legs. The first injection was administered in incomplete Freund's adjuvant and, over a one-month period, four subsequent injections were given without adjuvant. Three days after the last injection, the mice were sacrificed by cervical dislocation, the heart severed and the blood collected using a sterile Pasteur pipette. Popliteal lymph nodes were removed, single cell suspensions made and the pooled cells from both mice were fused with X63-Ag8.6.5.3 parental myeloma cells. Single step selection and cloning of hybridomas was performed using the ClonaCell-HY™ system (StemCell Technologies Inc., Vancouver, B.C.) following the instructions of the manufacturer. Essentially the cell fusion mixture was diluted and plated in a semi-solid methylcellulose containing HAT selective medium and B-cell growth factors, allowing a single-step selection and cloning of the hybridomas. After isolation and growth of individual clones, hybridoma tissue culture supernatants were first screened by indirect enzyme-linked immunosorbent assay (ELISA) on a variety of solid-phase adsorbed antigens.

2.6. Enzyme-linked immunosorbent assays

Hybridoma supernatants were screened in indirect ELISA (Tolson *et al.*, 1989) using three different antigen preparations to detect mAbs that bound to either *S. glossinidius*

surface or internal antigens or to antigens secreted or shed into the culture supernatant. Separate sets of ELISA plates (Falcon 3915 PRO-BIND™ Assay Plates, Becton-Dickinson, Oxnard, CA) were coated by drying 0.1 ml per well of *Sodalis* (1:100 dilution in water of 1.0 OD_{600}), the same concentration of *Sodalis* after heating at 100°C for 5 min, or a 1/10 dilution of *Sodalis* culture supernatant. Screening of hybridoma supernatants was also performed on *E. coli* and on human transferrin to detect cross-reacting antibodies and non-specific "sticky" antibodies respectively. Once hybridoma clones were selected, isotyping of secreted mAbs was performed using tissue culture supernatants and Instant Chek™ monoclonal antibody isotyping kit (EY Laboratories, Inc., San Mateo, CA).

2.7. Flow cytometry and immunofluorescence microscopy

Sodalis were harvested from 6 day old cultures by centrifugation, washed twice in PBS and adjusted with PBS to give an OD_{600} of 1.0. Other species of bacteria were grown as overnight cultures in Luria-Bertani broth and adjusted to the same density with PBS. For each immunoreaction, 100 µl of bacterial suspension were transferred to fresh 1.5 ml Eppendorf microcentrifuge tubes and centrifuged at 2900 *x g* to pellet the bacteria. Primary antibodies (undiluted hybridoma tissue culture supernatants) were added in 50 µl volumes to resuspend the bacteria and the tubes incubated for 30 min at room temperature (RT). After washing the cells twice by microcentrifugation with 1.0 ml PBS, the pellets were resuspended in 50 µl of a 1:50 dilution of goat-anti-mouse IgG/IgM (H + L) FITC conjugate (Caltag Laboratories, Burlingame, CA). The mixtures were incubated for 30 min at RT, washed twice with 1.0 ml sterile PBS, resuspended in 500 µl PBS and

flow cytometry was performed using a FACSCalibur™ flow cytometer (Becton-Dickinson, San Jose, CA). For every sample 1 x 10^5 events were analysed. The same suspensions, 10 µl volumes, were placed on a glass slide and viewed using fluorescence microscopy. For photography, the second antibody used was labelled with a 1:500 dilution of Alexa Fluor 488 (Goat anti-mouse IgG (H+L); Molecular Probes, Eugene, OR). The cells were attached to poly-L-lysine (Sigma Chemical Co. St. Louis, MO) coated slides and were examined and photographed using a Zeiss UV microscope fitted with a digital camera and a 100 X oil immersion objective.

2.8. One-dimensional gel electrophoresis

Sodalis cultures aged 5-9 days and other Enterobacteriaceae (overnight cultures) were washed twice in PBS. One ml of each suspension was adjusted to an OD_{600} of 0.2 with PBS. Culture supernatants were filter sterilized (0.2 µm Acrodisc® syringe filter, Pall Gelman Laboratories, Ann Arbor, MI) prior to centrifugation for 1 hour at 100,000 *x g* at 4°C. All samples were resuspended in equal volumes of 2X Laemmli sample buffer (Laemmli, 1970), heat denatured for 5 min, and briefly centrifuged except for the bacteriome. A single tsetse bacteriome was ground in 40µl 1X Laemmli sample buffer using sterile silica sand and a micropestle. The suspension was boiled for 10 minutes at 100°C and microcentrifuged for 1 min to pellet the sand and insoluble material. Each sample (20 µl) was loaded onto 10% acrylamide, 1-dimensional SDS-PAGE 0.75 mm minigels and run on a Mini-Protean II apparatus (Bio-Rad Laboratories, Hercules, CA). Proteins were electrophoresed at 50V for 30 min and then at 90V for an additional 90 min. A 10 kDa molecular mass ladder (#10064-012; Gibco BRL, Burlington, ON) was

run on each gel. Following electrophoresis, gels were stained with colloidal Coomassie Brilliant Blue (CBB) G-250 (see later) or transferred onto membranes for subsequent immunoblotting.

2.9. Immunoblotting

Immunoblotting using BioTrace™ polyvinylidene (PVDF) membrane (Pall Corporation, Ann Arbor, MI) was performed as previously described (Beecroft *et al.*, 1993) except that SuperSignal Dura chemiluminescence substrate (Pierce Chemical Company, Rockford, IL) was used. As primary antibodies, hybridoma tissue culture supernatants (anti-*Sodalis* mAbs) were used either undiluted or at a 1:2 dilution. A 1:50,000 dilution of secondary antibody (goat anti-mouse IgG/IgM-horseradish peroxidase conjugate; Caltag Laboratories, South San Francisco, CA) was used throughout. Kodak Biomax MR film (Eastman Kodak Company, Rochester, NY) was used to detect chemiluminescence. After development of the autoluminograms, the proteins were stained on the PVDF membrane with GelCode® Blue and the two patterns (film and PVDF) overlaid to reveal the precise location of the immunoreactive protein bands in relationship to the entire protein profile.

2.10. Immunoaffinity purification of Sodalis antigens

Antigens recognized by the anti-*Sodalis* mAbs were isolated using micro-immunoadsorbent columns by a procedure modified from Pearson and Anderson (1980). Protein A-agarose beads (100 µl; EY Labs, San Mateo, CA) were mixed with 2 ml of selected mAbs supernatants and incubated for one hour at 25°C with end-over end

mixing using a Labquake Shaker (Labindustries, Berkeley, CA). The beads were washed two times with PBS and pelleted by centrifugation. Either solubilized *Sodalis* lysate or *Sodalis* culture supernatant was added to the beads for overnight mixing at 4°C. The lysates were prepared as follows: 2.0 ml of a well established culture of *Sodalis* (0.28 OD_{600}) were transferred to a 5 ml conical centrifuge tube and centrifuged at 1250 *x g* for five minutes at 25°C. Cells were washed twice with 5 ml PBS followed by resuspension of the pellet in 2 ml of solubilization buffer (0.1 M glycine, 0.1% NP-40 in PBS, pH 7.4) and 10 µl/ml of 100X protease inhibitor cocktail (500 µM iodoacetamide, 100 µM pepstatin A, 100 µmM leupeptin, 50 µM EDTA, 100 µM AEBSF, 100 µg/ml aprotinin in PBS). The mixture was vigorously vortexed, centrifuged for 5 min at 15, 800 *x g* and the soluble fraction added to the mAb-Protein A-beads. Alternatively, *Sodalis* culture supernatant was centrifuged for 10 min at 15, 800 *x g* to remove cellular debris, syringe filtered (0.2 µm) and 2 ml was transferred to the mAb-Protein A-bead mixture. As negative controls, the same incubation mixtures lacking *Sodalis* lysate or supernatant were prepared. After a 1 h incubation at 25°C, the samples were centrifuged at 1250 *x g* for 5 min and the immunoadsorbents were transferred to Nanosep™ MF 0.45 µm microporous centrifugal filters (Pall Filtron Corp, Northborough, MA). The adsorbent beads were washed 5 times with 500 µl PBS/0.1% Tween 20 at 11,000 *x g* for 3 min. Finally, the washed beads were heated at 100°C for 10 min in 30 µl of Laemmli buffer (Laemmli, 1970), the solubilised proteins were eluted by a 2 min centrifugation at 5,000 *x g* and analysed by 1-D gel electrophoresis.

2.11. Two-dimensional gel electrophoresis

High resolution two-dimensional sodium dodecyl sulphate polyacrylamide gel electrophoresis was performed using the ISO-DALT multiple 2-D system (Anderson and Anderson, 1978a; Anderson and Anderson, 1978b) as previously described (Anderson *et al.*, 1985). One ml of a 5 day old *Sodalis* culture (0.20 OD_{600}) was twice washed in PBS and then resuspended in 20 µl of SDS mix (2% (w/v) SDS, 1% DTT (w/v), 10% (v/v) glycerol, 50 mM CHES, pH 9.5) and mechanically disrupted using a 1 cc syringe equipped with a 21G needle. Samples were denatured at 100°C for 5 min, centrifuged for 30 sec at 10,000 *x g* to remove insoluble material and subsequently loaded onto pre-focused tube gels containing pH range 3-10 ampholines (Pharmalyte 3-10, Amersham Pharmacia, Upsala, Sweden). First dimension isoelectric focusing was conducted at 800 V for 18 h (14,400 Vh). Following electrophoresis, the tube gels were equilibrated for 15 min at room temperature in equilibration buffer and were immediately mounted onto 10-16.5% gradient SDS-PAGE slab gels with the acidic end positioned to the left. Electrophoresis was performed at 4°C and run at 1 Amp until the dye front was about 1 cm from the bottom of the gel (about 5 hours). After electrophoresis, the gels were fixed and stained with colloidal Coomassie Brilliant Blue G-250.

2.12. Protein staining with colloidal Coomassie Brilliant Blue G-250

Both 1-D and 2-D gels were agitated gently in fixative (50% (v/v) ethanol, 3% (v/v) ortho phosphoric acid) for 1-4 days at room temperature, washed three x 30 min in distilled water and allowed to equilibrate in Neuhoff's solution (16% (w/v) ammonium sulphate, 25% (v/v) methanol, 5% (v/v) ortho phosphoric acid; (Neuhoff *et al.*, 1988) for one hour

with gentle agitation. One gram of Coomassie Brilliant Blue (CBB) G-250 (SERVA Blue G, research grade; SERVA Electrophoresis, Heidelberg, DRG) was sprinkled into the Neuhoff's solution and staining continued for 2-3 days. Once dark protein spots were visible, gels were scanned and the protein spots cored or the intact gels transferred into a 20% (w/v) ammonium sulphate solution for storage at 4°C (Neuhoff *et al.*, 1988). Digital images of both 1-D and 2-D CBB-stained gels were captured by scanning at 300 dpi using a colour scanner (UMAX Astra 3400, Fremont, CA). The images were stored and manipulated in TIFF format using Photoshop™ 5.5 graphic software (Adobe Systems Inc., San Jose, CA).

2.13. Reduction, alkylation and tryptic digestion of proteins

Protein bands or spots of interest were cored from gels using 4 mm plastic straws and were either transferred to 1.5 ml Eppendorf microcentrifuge tubes (rinsed with 70% (v/v) ethanol to remove any contaminants) for tryptic digestion or to 96 well sterile tissue culture plates (one spot per well in 10 µL of 20% w/v ammonium sulphate) for storage at -20°C. For analysis by mass spectrometry, 2-D protein spots were de-stained (50% (v/v) methanol/ 5% (v/v) acetic acid) overnight, reduced with 50 mM DTT at 56°C for 30 min and alkylated at 45°C for 30 min in the dark with 100 mM iodoacetamide (Hale *et al.*, 2000; Kinter, 2000). Following reduction and alkylation, protein spots were digested overnight at 37°C with 20 ng/µl modified porcine sequence grade trypsin (Promega, Madison, WI) according to the manufacturer's directions. Peptides were extracted from the gel pieces using one wash with 30 µl of 100 mM sodium carbonate. The eluates were

reduced to a final volume of 20 µl in a vacuum centrifuge (Speed Vac Concentrator, Savant, Hicksville, NY) prior to analysis by mass spectrometry.

2.14. Nanospray MS/MS

Tryptic peptides were desalted using glass capillary needles (Protana Inc., Staermosegaardsvej, Denmark) packed with C18 resin (Perceptive POROS R2, 50 µm bead) and were extracted into sample needles using 1.0 µl 50% (v/v) methanol 1% (v/v) formic acid. Nanospray electrospray ionisation (ESI) was used to introduce ions into a PE-SCIEX Q-STR*i* quadrupole time-of-flight mass spectrometer (Applied Biosystems, Foster City, CA). Data were managed with Bioanalyst Software (PE-SCIEX, Boston, MA). Peptide fragmentation data searching was performed using the Mascot MS/MS Ions Search algorithm (Matrix Science; London, UK: http://www.matrixscience.com/).

2.15. MALDI-TOF mass spectrometry

Peptides from each trypsin-digested sample were desalted using a ZipTip (C18 resin; P10, Millipore Corporation, Bedford, MA). For each sample, 1.0 µl of the desalted peptide mixture was mixed (1:1) with the matrix, alpha-cyano-4-hydroxycinnamic acid spiked with 1 fm each of internal standards (Sigma Chemical Co. St. Louis, MO), bradykinin, fragment 2-9 (FW 904.4681) and adrenocorticotropic hormone, fragment 18-39 (FW 2465.1989) and spotted onto a Voyager, 100 position, stainless steel MALDI plate (Applied Biosystems, Foster City, CA). An Applied Biosystems Voyager DE-STR mass spectrometer (Applied Biosystems, Foster City, CA) running in delayed extraction, reflectron mode was used to acquire MALDI-TOF data. Selected peptide masses were

submitted to MS-Fit (Protein Prospector software package; San Francisco, CA: http://prospector.ucsf.edu/) and Mascot (Matrix Science, London, UK: http://www.matrixscience.com/) for database searching and determination of peptide mass maps.

3. Results

3.1. Growth and cloning of Sodalis glossinidius

Sodalis grown axenically in serum-free medium displayed features of density dependent cultures. In liquid medium the organisms were slow growing and reached a maximum density (0.30 OD_{600}) after approximately 10 days. Microscopic examination of five day old cultures revealed small, Gram negative rods that aligned end to end to form long chains. When the bacteria in these cultures were negatively stained and examined by transmission electron microscopy, the rods were estimated to be approximately 0.75 μm by 2.75 μm in size and lacked both cilia and flagella (Figure 3.1). *Sodalis* from older liquid cultures (>5 days) gradually lost their filamentous, chain-like morphology and favoured an arrangement into pairs and singles. With increasing culture age, the cells also became highly pleomorphic with some of the individual rods developing what appeared to be centrally located spherical bodies, perhaps storage vesicles. Spherical cells were also observed in the extracellular material, possibly resulting from the disintegration of the cell wall or the formation of transport vesicles (Li *et al.*, 1998). The cultures grew surprisingly well in long-term serum free conditions, in contrast to *Sodalis* cultures (previously described by Welburn and Dale, 1997) that survived only two or three passages in serum-free liquid medium. Interestingly, light microscopy of culture

supernatants from old cultures revealed an accumulation of fragile, spindle shaped clusters of crystals that were encircled by aggregations of live bacteria.

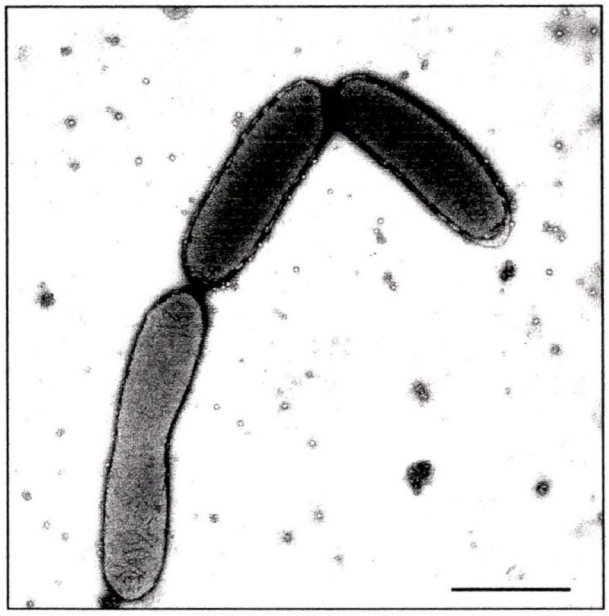


Figure 3.1. ***Transmission electron micrograph of <u>Sodalis glossinidius</u> isolated from a five day old liquid culture.***
Note pleomorphism and characteristic linear alignment of the cells. The scale bar represents 1.5 micron. Magnification is 8000 X.

I successfully isolated several clones of *Sodalis* from a serum-free liquid culture after inoculation onto Mitsuhashi-Maramorosch Insect (MMI) agar (1% w/v) plates and incubation under microaerobic conditions. After 14 days, hundreds of small (<1 mm) colonies could be seen by the naked eye. As observed with other Gram negative bacteria, the colonies were translucent and white. The primary form grew as flat colonies with irregular edges and rough papillated surfaces (Figure 3.2). These rough colonies were designated phase I. After 3 weeks, pinhead-sized, opaque, white colonies with smooth, shiny, convex surfaces (designated phase II) appeared. Bacteria in these phase II

colonies were plated on fresh MMI agar plates and growth continued as smooth, phase II colonies. Reversion to the original rough phenotype was not observed. Surprisingly, when plates containing both phases were incubated aerobically in atmospheric conditions, only the phase II forms maintained their morphology and continued to grow, albeit slower than previously observed. In contrast, large rough phase I colonies slowly transformed into a smooth phase II phenotype possessing a mucoid-like surface. Growth of the smaller colonies was arrested under aerobic incubation.

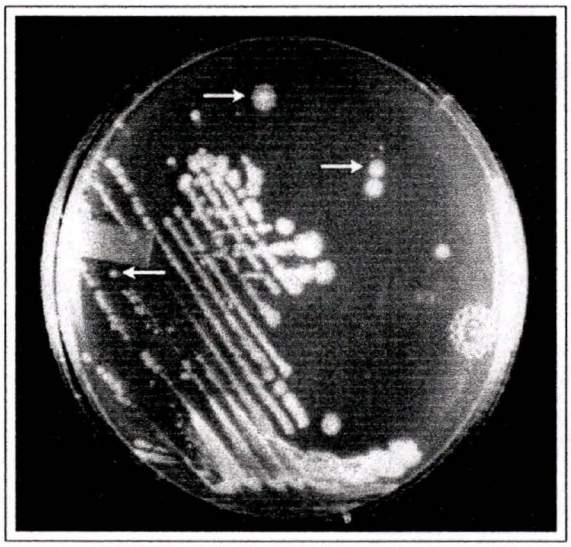


Figure 3.2. ***Cloned S. glossinidius growing on semi-solid (1% agar) Mitsuhashi and Maramorosch Insect medium plates.***
The arrows indicate colonies of different sizes that were assessed for purity by light microscopy and PCR.

3.2. Analysis of the S. glossinidius proteome

Cloned *Sodalis* were grown to log-phase (6 day serum-free, liquid culture) and their protein profiles determined by 1-D and 2-D gel electrophoresis (Figure 3.3). The 1-D gel banding pattern showed predominant proteins at approximately 60 and 35 kDa

(Figure 3.3, Panel A, lane 1). This precise pattern was observed repeatedly with bacteria from both log-phase and older, stationary-phase cultures. The characteristic gel profile was subsequently used as a diagnostic indicator of *Sodalis*, since when compared to the 1-D gel patterns of other Gram negative bacteria (for example *E. coli*, Figure 3.3, Panel A, lane 2), it was unique. Both rough (phase I) and smooth (phase II) colonies were identified as *Sodalis* using both gel banding patterns and colony PCR employing species-specific primers. The 2-D gel profile of log-phase *Sodalis* revealed a large number of

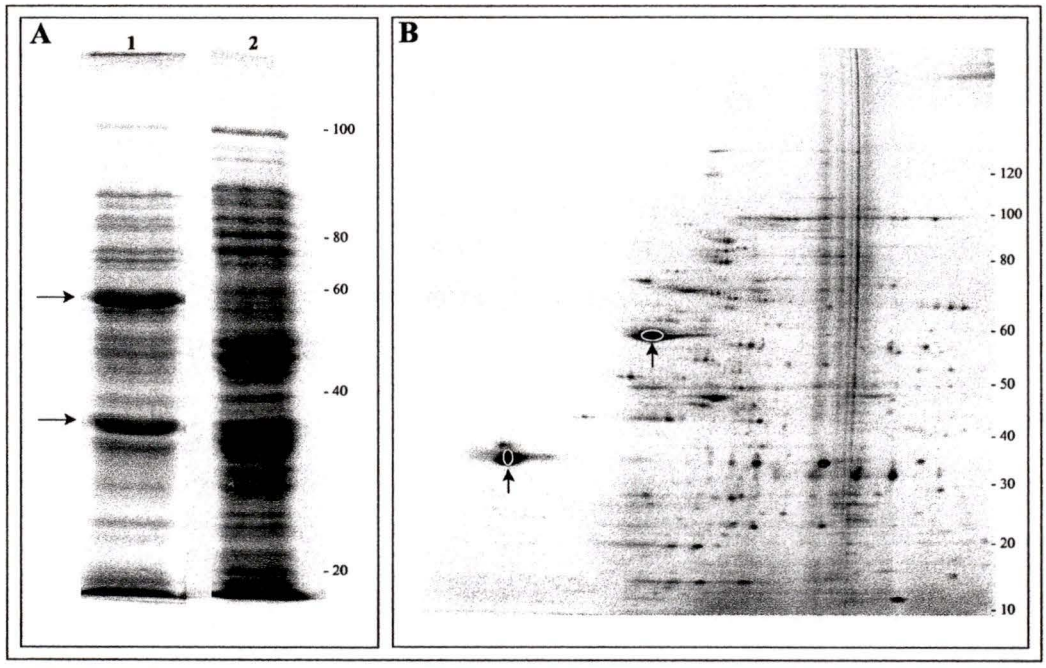


Figure 3.3. ***Protein profiles of broth-reared Enterobacteriaceae.***
Proteins were stained using colloidal Coomassie Brilliant Blue G-250. Panel A: One-dimensional SDS-PAGE. Lane 1, S. glossinidius lysate (cells from 1.0 ml of culture at an OD_{600} of 0.2). The arrows indicate the characteristic predominant 60 kDa and 35 kDa proteins. Lane 2, E. coli lysate from overnight culture adjusted to OD_{600} of 0.2. Panel B. Two-dimensional polyacrylamide gel analysis of Sodalis lysate. The ellipses outline the major 60 kDa and 35 kDa protein spots used for mass spectroscopy. Wide range ampholytes, pH 3-10, were used in the first dimension and a 10-16.5% gradient gel in the second dimension. The gel is oriented with the acid end to the left. Molecular mass standards are shown on the right of each gel.

protein spots after staining with colloidal CBB (Figure 3.3, Panel B). Although several major spots and spot constellations were observed, one at 60 kDa (upper arrow) was the most abundant in the proteome determined with the wide-range (pH 3-10) ampholines. There were several major spots at approximately 35-38 kDa (the major one indicated by the lower arrow) that would have contributed to the intense band at this apparent mass in the 1-D gel profile.

3.3. Characterization of monoclonal antibodies and their specific antigens

After 10-14 days of growth in the semi-solid Clonacel-HY^{TM} medium, 252 hybridoma clones were selected and sub cultured. Supernatants from these hybridomas were tested in ELISA using *Sodalis* lysates or *Sodalis* culture supernatants as solid-phase antigen. After selection of different hybridomas based on their mAb binding specificity, 52 stable clones were chosen for further analysis. Surface binding mAbs were detected by flow cytometry and by immunofluorescence microscopy and to my surprise, 29 (56%) showed positive surface binding. Of the 52 mAbs, 24 were positive in immunoblots on *Sodalis* lysates and 9 on filter-sterilised *Sodalis* culture supernatants. Based on binding patterns in ELISA, surface immunoreactivity and reactivity in immunoblots, 5 of the mAbs were selected for detailed study. The summarised results are shown in Table 3.1. All five mAbs showed strong surface binding to living *Sodalis* as shown by flow cytometry (Figure 3.4). Indeed with one mAb, 1H1, the fluorescence intensity was 1000-fold higher than with the negative mAb control. A bimodal distribution of fluorescence was observed with this mAb whereas the others exhibited a single peak distribution.

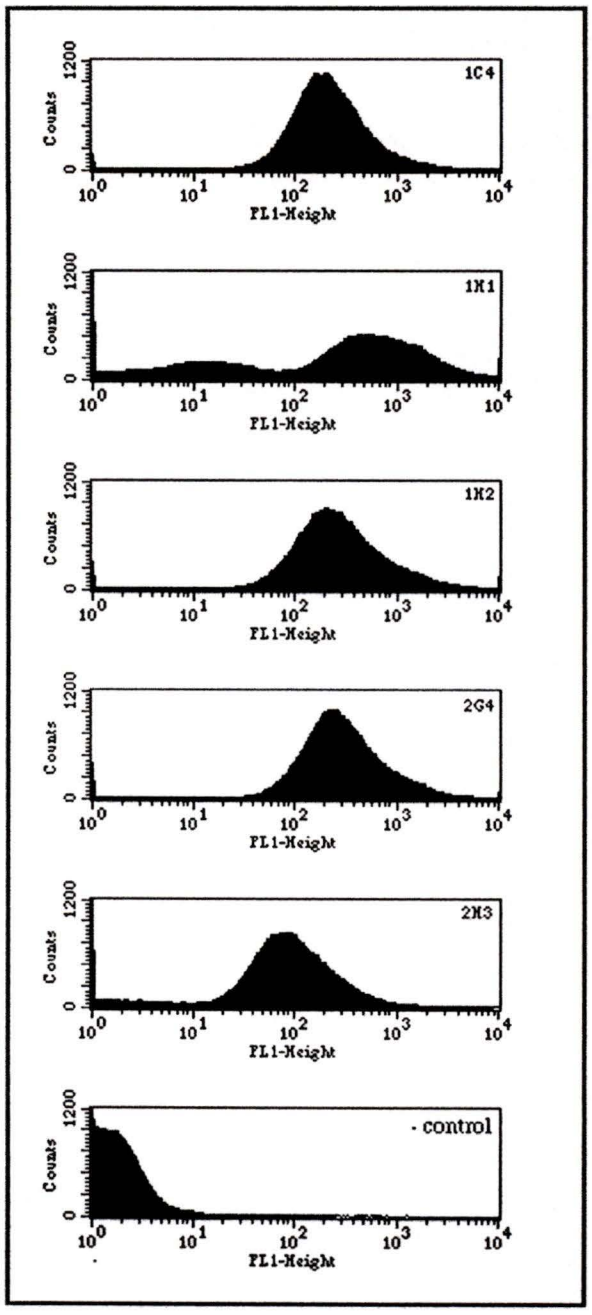


Figure 3.4. ***Flow cytometry analysis of anti-<u>Sodalis</u> monoclonal antibodies.***
The mAbs 1C4, 1H1, 1H2, 2G4, 2H3 and anti-human transferrin (negative control) were tested for binding to the surface of live <u>S. glossinidius</u>. Undiluted hybridoma tissue culture supernatants were used with a 1:50 dilution of fluorescein-conjugated goat anti-mouse IgG/IgM secondary antibody. For every sample, 1 x 10^5 cells/particles were analysed.

When examined by epifluorescence microscopy, all of the mAbs showed what appeared to be surface fluorescence on *Sodalis*. With 4 of the mAbs, very bright, uniform fluorescence was observed around the entire periphery of the cells as shown in the example in Figure 3.5 below.

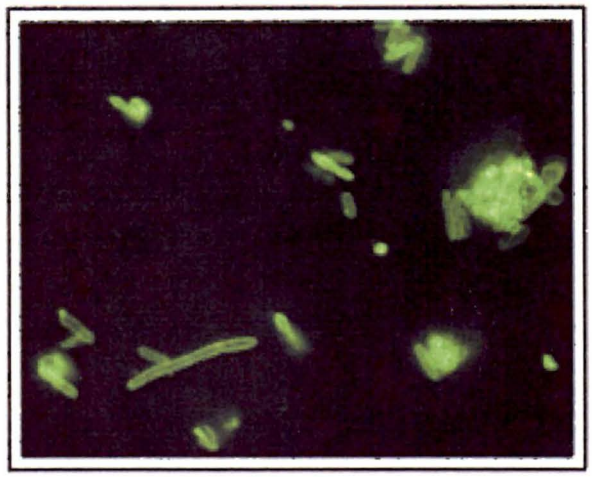


Figure 3.5. ***Photomicrograph of surface membrane immunofluorescence on S.glossinidius.***
Immunoreactions were performed on live broth-reared Sodalis cultures. Surface-specific mAb 2G4 (neat) and a 1:500 dilution of Alexa Fluor 488 goat anti-mouse IgG secondary antibody were used followed by fixation of labelled cells to poly-L-lysine coated glass slides.

However, with mAb 1H1 some of the younger cells in chains were only weakly or non-fluorescent whereas single cells, doubles and older, disintegrating cells showed moderate cell surface labelling. A surprising observation, and one that may explain the bimodal fluorescence profile seen with this mAb, is that bright extracellular fluorescence was seen as star-like clusters (not shown) in the medium. At first I thought this was cell debris but upon closer examination most of the bacteria were still intact and the extracellular material was of crystalline composition. The immunofluorescence profiles suggested that

in older cultures the antigen detected by mAb 1H1 was abundant and existed both on the cell surface and as a particulate form in the culture medium.

When the mAbs were tested in immunoblots on *Sodalis* lysates, proteins of 15, 20, 35 and 60 kDa were detected (summarized in Table 3.1). It is interesting that strong

Table 3.1. ***Summarised data: characterization of selected anti-Sodalis monoclonal antibodies.***

Clone[a]	Isotype[b]	ELISA screening[c]	Western blot (kDa)[d]	Flow cytometry[e]	IFA[f]
1C4	IgG2b	sur, int	20	+4	4+ surface
1H1	IgG2a	sur, int, sol	60	+2	3+ crystals
1H2	IgG2b	sur, int	15	+4	4+ surface
2G4	IgG2b	sur, int	35, 15	+4	4+ surface
2H3	IgG2b	sur, int	35, 15	+4	3+ surface

[a] *Sodalis* specific mAbs are underlined.
[b] Determined by antigen-capture ELISA.
[c] Antigens were intact *Sodalis*, lysed *Sodalis* or conditioned medium (harvested from a week old culture). Positive mAb binding to these antigens was labelled as surface (sur), internal (int) or soluble (sol) respectively. Absorbance values at least ten-fold greater than background were considered positive.
[d] Performed on *Sodalis* lysates.
[e] Determined by flow cytometry on live *Sodalis*. Fluorescence intensity was graded as very strong (+4) to weak (+1) or negative (-).
[f] Epifluorescence microscopy was performed on *Sodalis* adsorbed to poly-L-lysine coated slides. Crystals refer to the extracellular spindle-like crystal structures. The mAb 1H1 was weak on bacterial surfaces in young cultures.

reactions were seen with bands at 35 kDa and 60 kDa, since proteins of these masses were predominant in the 1-D and 2-D gel profiles. MAb 1H1 bound to a 60 kDa protein and since, in immunofluorescence experiments this mAb also detected the crystal-like extracellular antigen described above, immunoblotting experiments were performed on *Sodalis* culture supernatants to characterize the shed or released material. The results are

shown in Figure 3.6. An intense 60 kDa band was detected with mAb 1H1 in supernatant from a 10 day old culture both before (lane 1) and after (lane 2) high speed ultracentrifugation (125,000 *x g* for 1 hour), indicating that the protein was soluble.

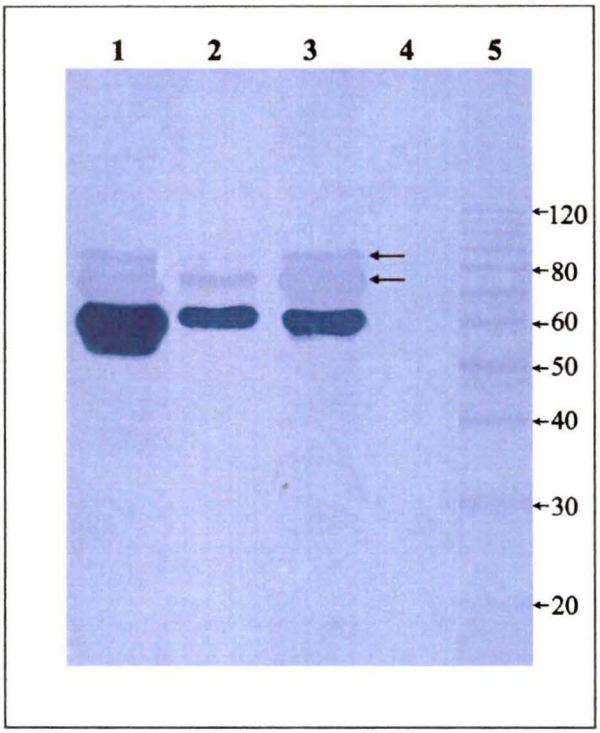


Figure 3.6. ***Immunoblot analysis of antigens in Sodalis culture supernatants.*** *Proteins were separated using a 10% gel, blotted onto PVDF membrane and probed with anti-Sodalis mAb 1H1 (previously determined by microaffinity purification and mass spectrometry to recognize a 60 kDa GroEL homologue). The autoluminogram pattern is shown superimposed on the same PVDF membrane which was subsequently stained with GelCode® Blue to reveal the exact location of the immunoreactive protein bands. Lane 1, non-centrifuged 10d supernatant; Lane 2, ultracentrifuged 10d supernatant; Lane 3, ultracentrifuged 21d supernatant; Lane 4, naïve culture medium; Lane 5, molecular mass standards (10 kDa ladder). Accumulation of the 60 kDa immunoreactive product and two other proteins at 80 kDa and 90 kDa is evident by comparing the older, 21 day culture, Lane 3 (arrow) with the younger culture (Lane 2).*

The amount of this soluble protein was increased in the clarified supernatant from a 21-day old culture (lane 3). No reactivity was seen in normal (naïve) culture medium (lane 4). Counterstaining of the immunoblot membrane with GelCode® Blue showed that in

addition to the 60 kDa immunoreactive band, soluble proteins of approximately 70 and 80 kDa were also present in the culture medium (arrows, lane 3) and that the amounts increased with the age of the cultures (compare lanes 2 and 3).

The specificity of the 5 mAbs was tested on 5 different species of live Enterobacteriaceae by flow cytometry and immunofluorescence microscopy. Three of the mAbs bound only to the surface of *Sodalis* (not shown). Immunoblotting on lysates of the 5 species showed that mAb 2G4 (Figure 3.7, Panel A) and mAb 2H3 (data not shown) recognised a 15 kDa protein in both *Sodalis* and *E. aerogenes.* MAb 1H1 bound to a 60 kDa protein only in *Sodalis* (Figure 3.7, Panel B). The immunoblot membranes were counterstained with GelCode® Blue so that the protein profiles of the different bacterial species could be seen. The unique profile of *Sodalis* (lane 8 in both A and B) is evident. The mAbs were also tested in immunoblots on solubilised tsetse midgut bacteriomes known to harbour the primary tsetse endosymbiont, *W. glossinidia.* Only mAb 1H1 reacted with a protein in the bacteriome (Figure 3.7, Panel C), suggesting that it cross-reacts with the 60 kDa protein of *W. glossinidia* and thus is not *Sodalis*-specific. Although mAb 2G4 bound molecules in *E. aerogenes*, surprisingly, it did not cross-react with proteins in the bacteriome. None of the remaining 3 mAbs (1C4, 1H2 and 2H3) showed reactivity with the bacteriome proteins thus confirming that they were specific for *Sodalis*.

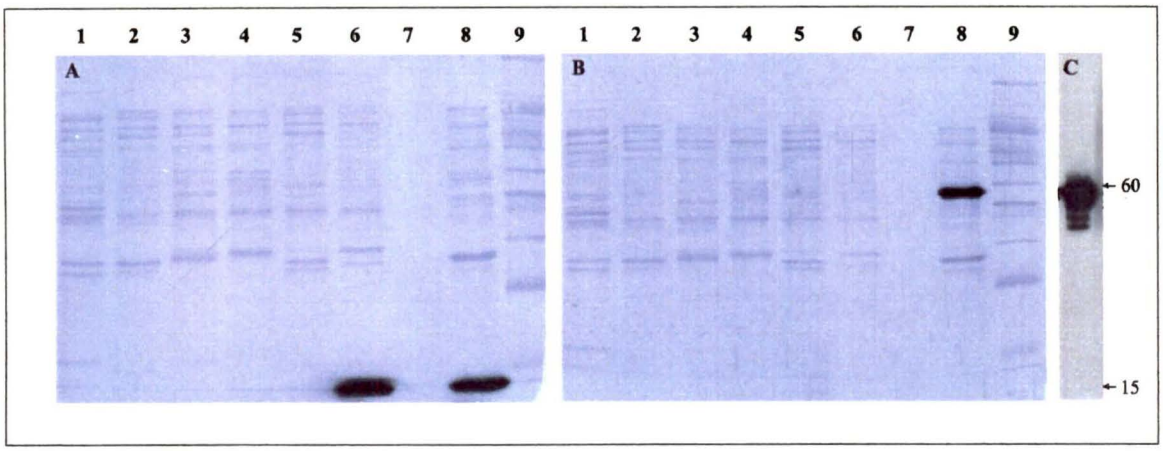


Figure 3.7. ***Immunoblot analysis of anti-Sodalis mAbs against a panel of Enterobacteriaceae.***

Lane 1, E. coli 0157:H7*; Lane 2, E. coli* Y1090*; Lane 3, Citrobacter freundii; Lane 4, Salmonella typhimurium; Lane 5, Shigella sonnei; Lane 6, Enterobacter aerogenes; Lane 7, 1X Laemmli buffer; Lane 8, Sodalis glossinidius positive control; Lane 9, 10 kDa ladder. Panel A: mAb 2G4; Panel B: mAb 1H1; Panel C: mAb 1H1 on solubilised tsetse bacteriome.*

To identify the molecules recognised by selected mAbs, immunoaffinity purification of the bound ligands was performed prior to mass spectrometry. Micro-immunoadsorbents containing protein-A-agarose were used since the mAbs were shown in ELISA to bind protein A. Eluted molecules were separated by 1-D gel electrophoresis and visualised by sensitive colloidal CBB staining. The results are shown in Figure 3.8. Presumptive immunoglobulin heavy chains and light chains from the mAbs were clearly seen as bands at approximately 52-55 kDa and 25-27 kDa respectively (antigen controls, lane 2 in each of panels A, B and C). MAb 1H1 bound to an abundant 60 kDa molecule from the *Sodalis* lysate (asterisk, lane 1, panel A) and from culture supernatant (arrow 3, lane 1, panel B). The presumptive heavy and light chain bands (arrows 1 and 2 respectively, lane 2, panel A) and the 60 kDa band (arrow 3, lane 1, panel B) selected for in-gel tryptic digestion and identification by mass spectrometry (see below). Results from the surface-binding mAb 2G4 are shown in panel C. A prominent band was seen at

36 kDa (arrow 4, lane 1, panel C). This band (arrow 4) was also chosen for in-gel tryptic digestion and identification by mass spectroscopy (see below). Interestingly, a band at the same position (although of lower intensity) was also seen in the eluate from the mAb 1H1 immunoadsorbent (lane 1 of panel A) but not the culture supernatant, implying that this was a non-specifically bound sticky protein in the cell lysate.

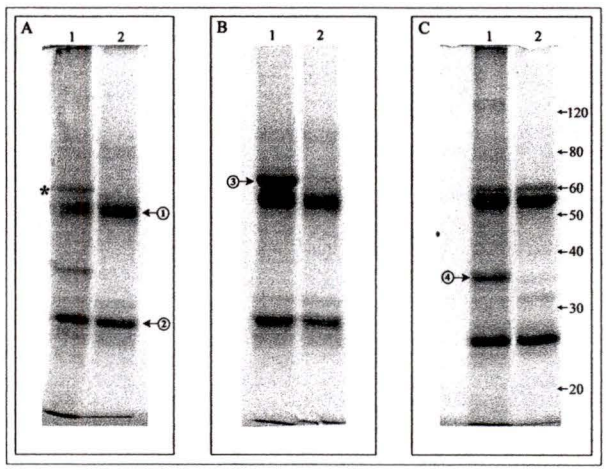


Figure 3.8. ***SDS-PAGE analysis of S. glossinidius molecules purified using Protein-A agarose microimmunoaffinity columns.***

Samples that were cored and processed for mass spectrometry are denoted with a numbered arrow. In each of the panels (A, B and C), lane 1 shows the affinity-purified molecules and lane 2 represents the antigen control. Panel A: Eluate from mAb 1H1 column-purified Sodalis lysate. The asterisk indicates an isolated 60 kDa antigen. Arrows 1 and 2 represent the presumptive IgG heavy and light chain respectively. Panel B: Eluate from mAb 1H1 eluted antigen from Sodalis culture supernatant. Arrow 3 identifies a 60 kDa band comparable to the asterisk band in panel A. Panel C represents eluate from mAb 2G4 purified Sodalis lysate. An isolated 36 kDa band, denoted by arrow 4, may be a non-secreted sticky protein as it also weakly present in lane 1, panel A.

3.4. *Mass spectroscopic identification of Sodalis surface molecules*

In-gel reduction, alkylation and tryptic digestion were performed on each of the four gel bands from the micro-immunoadsorbent columns (indicated by the arrows in

Figure 3.8). Both MALDI-TOF and Q-TOF mass spectrometry were used for protein identification by peptide mass mapping and peptide sequencing, respectively. The results are shown in Table 3.2. Band 1 was identified as immunoglobulin heavy chain IgG2a. This confirms the isotyping results for this mAb. Band 2 was identified as Ig light chain. Band 3, specifically adsorbed by mAb 1H1, was unequivocally GroEL from *Sodalis glossinidius*. The results confirm that a soluble form of GroEL is shed or secreted, since this protein was isolated from *Sodalis* culture supernatant. The mass spectrometry results obtained with band 4 were unexpected and interesting. The protein was identified as a 36 kDa porin. Since this band was also evident in eluates from several of the mAb immunoadsorbents, it either non-specifically binds to the adsorbents or the protein interacts with the mAbs or the Protein A on the adsorbent. Nevertheless, several of the mAbs also bind to this molecule in immunoblots, indicating that they may be specific for porin epitopes.

Table 3.2. Mass spectrometric identification of proteins eluted from micro-immunoadsorbent columns made using anti-Sodalis monoclonal antibodies.

Protein[a]	Signature masses[b]	Protein coverage[c] (%)	Protein identity	Mass[e] (kDa)
1	ASGYTFTDYYINWVKQKPGQGLK TTAPSVYPLAPVCGDTTGSSVTLGCLVK APQVYVLPPPEEEMTK	23	IgG2a heavy chain[f]	57
2	ASNLEGVTPAR DSTYSMSSTLTLTK DIVLTQSPASLAVSLGQR	21	Ig kappa light chain	24
3	ANDAAGDGTTTATVLAQSIVNEGLK AIAQVGTISANADETVGTLIAEAMAK AGKPLLIIAEDVEGEALATLVVNTMR QIVANAGEEPSVIANK	27	*S. glossinidius* GroEL[d]	58
4	DGNKLDLYGK LAFAGLKFGDAGSFDYGR YVDLGATYYFNK TTGVATYR	13	*E. aerogenes* Omp36 porin	35

[a] Proteins from the gel bands indicated by arrows in Figure 3.8.

[b] In-gel tryptic digestion was performed and extracted peptides were analysed by MALDI-TOF and Q-TOF mass spectrometry.

[c] The percentage of the identified protein sequence covered by peptide masses is shown. The mass map search algorithms used were MS-FIT and Mascot.

[d] Unequivocal protein identification based on an exact match of peptides to NCBI accession number AF321517. Other protein identities based on homology to sequences in the NCBInr database.

[e] Masses are given as an approximation calculated from protein hits in the NCBInr database and spot position on the 2-D gel.

[f] Protein identity based solely on MALDI-TOF mass mapping analysis.

4. Discussion

Although slow-growing even in the presence of feeder cells, *Sodalis* could be cultured in serum-free medium and cloned by streaking on MMI agar. In microaerophilic conditions, the bacteria grew as rough, phase 1 colonies. After long incubation periods or after stressing the bacteria by growth in aerobic conditions, smooth, phase II colonies grew. The phase II forms did not revert to phase I colony types upon subculture, suggesting that irreversible phase transition occurred. These observations contradict those of Dale and Maudlin (1999) who reported that morphological changes in colonies were population dependent and not the result of mutation or phase transition. The phase variation phenomenon that I observed with older *Sodalis* colonies has also been reported in other species of bacteria where prolonged culture *in vitro* and diminished nutrient availability triggered phase variation. *Lactococcus lactis,* when cultured in spent medium, demonstrated higher resistance to environmental stress. When the morphotypes were examined, the original parental plasmids had been rearranged into a new plasmid (Kim *et al.*, 2001). As *Sodalis* harbours plasmids, it may be that this symbiont has developed a similar plasmid-mediated adaptation for survival, one that is phenotypically characterized by phase transition. Phase transition can ensure that organisms are well able to adapt in order to survive changing environmental conditions. For example, Deziel *et al.*, (2001) recently described a mutant phenotype of *Pseudomonas aeruginosa* with increased aerotolerance capabilities. Different morphotypes, as observed with the entomopathogenic nematode symbiont, *Photorhabdus luminescens*, usually have diverse biological and biochemical properties indicative of the roles that each mutant plays

(Bondi *et al.*, 1999). Phenotypic diversity has also been implicated in increased pathogenicity and cell-surface associated virulence (Weiser, 1993). Unfortunately, the significance or function of the phase variation observed in *Sodalis* is only speculative. However, the distinct colony morphologies are likely adaptive by nature and phenotypically represent mechanisms of survival, immune evasion or pathogenicity (Henderson *et al.*, 1999).

Is it possible that *Sodalis*, which displays characteristics of both pathogenic and symbiotic organisms, is a pathogen that is evolving towards a symbiotic relationship with tsetse? Pinnock and Hess (1974), originally thought *Sodalis* was a pathogen, based on electron microscopic examination of tsetse tissues. These authors suggested that lytic zones, surrounding *Sodalis* outer membranes, might be indicative of an adverse reaction between the host tissue and the bacteria. In addition, the highly infected fat body in *Glossina pallipides* showed signs of cellular disruption and degeneration that could not be explained as procedural artefacts. This cellular disruption was also observed when *Sodalis,* isolated from the hemolymph of tsetse, were cultured on a layer of *Aedes albopictus* (mosquito) feeder cells (Welburn *et al.,* 1987). Since *Sodalis* exists both extracellularly in the hemolymph and intracellularly within tsetse midgut epithelial cells, the invasive form may represent a pathogenic remnant.

With most Gram-negative pathogenic bacteria, adhesion to target cells and subsequent colonization are important steps in the process of host cell infection. It is interesting that GroEL (Hsp60), the major *Sodalis* surface molecule identified in the current study, is used by several bacteria to adhere to and invade host cells. Surface disposed GroEL has been described as either an invasin or as an adhesin in, *Helicobacter*

pylori (Yamaguchi *et al.*, 1996; Vanet and Labigne, 1998), *Legionella pneumophila* (Garduno *et al.*, 1998), *Salmonella typhimurium* (Ensgraber and Loos, 1992) and *Haemophilus ducreyi* (Frisk *et al.*, 1998). In addition, it has been reported that exposure to a stressful environment may enhance cell attachment and invasion by microbes utilizing the adherence properties of Hsp60 (Hennequin *et al.*, 2001). These authors suggested that GroEL could be translocated to the bacterial surface despite the fact that it lacks a classical signal peptide sequence. Their results with *Clostridium difficile* indicated that an extracellular release of GroEL was followed by surface adsorption. They also suggested a type III secretion mechanism for Hsp60, which also would result in transport of the chaperone to the bacterial outer membrane. It is of interest that a *S. glossinidius* mutant, having a defect in the chromosomal *invC* gene (a component of the type III secretion system of many pathogens) was incapable of cellular invasion (Dale *et al.*, 2001). My results showing the presence of *Sodalis* Hsp60 on the cell surface and in culture supernatant are compatible with an active secretion mechanism. Phylogenetic analysis of other pathogenic enteric secretion systems also suggests that prior to adaptation to a symbiotic lifestyle within the tsetse, *Sodalis* may have been an insect pathogen. The size of its genome, as determined by Akman *et al.*, (2001), is 2 Mb, more than twice the size of the genome of the primary symbiont, *Wigglesworthia glossinidia.* The larger genome of *Sodalis* and its ability to adapt to culture *in vitro* suggest that it has not become fully symbiotic and that, perhaps, some pathogenic capabilities remain.

Other than that of an adhesin or an invasin, are there other roles that GroEL plays? One hypothesis is that Hsp60 acts as a defense against anti-microbial peptides. Insects favour an innate immune response over an acquired response, as it is 100 times

more energy efficient. This translates into a faster reaction time to invasive organisms (Barra *et al.*, 1998). Indeed, only 6 hours after bacterial challenge, immune molecules were detected in the haemolymph of *Drosophila* (Uttenweiler-Joseph *et al.*, 1998). The predominant innate insect defense mechanism is characterized by the production of antimicrobial peptides. On average, an insect produces between 10-15 antimicrobial peptides, each exhibiting a unique activity spectrum and tissue specificity in response to injury or infection. Antimicrobial peptides are usually synthesized in the fat body, although anterior midgut tissue is implicated in the haematophagous insect, *Stomoxys calcitrans* (Lehane *et al.*, 1997). In the tsetse, antimicrobial peptide genes were recently identified in the fat body after immune stimulation with both *E. coli* and *T. b. rhodesiense* (Hao *et al.*, 2001). This concentrated zone of antibacterial peptide production may explain the destruction of the fat body (containing *Sodalis*) reported by Pinnock and Hess (1974). In addition, the production of peptides in the midgut and subsequent secretion in the hemolymph of tsetse may also affect *Sodalis* residing in both of these tissues. It is thought that the peptides destroy bacteria by interacting with the outer membrane lipids, consequently resulting in either pore-induced lysis or structural collapse. As an example of this, a lepidopteran cecropin was shown to initiate pore formation in *E. coli* cell membranes (Lockey and Ourth, 1996).

Otvos *et al.*, (2000) identified GroEL (Hsp60) as one of several molecules on the bacterial membrane that bound peptides but did not trigger pore-assisted lysis. If indeed GroEL is a target protein for antimicrobial peptides, how does *Sodalis* survive as an over-expresser of GroEL? As I have shown, *Sodalis* produces a membrane bound form as well as a soluble form of GroEL. It is possible that the soluble form acts as an antibacterial

peptide sponge thus protecting the cell surface from the action of the peptides. In addition, the membrane-bound form may alter the cell surface enough to inhibit antimicrobial activities. Supporting this hypothesis is a study on the mechanisms of three distinct antimicrobial peptides by Liang and Kim (1999). One of the selected strains of *E. coli* was a GroE+/DnaK+ over-expressing mutant. To their surprise, this strain resisted the actions of all three anti-microbial peptides and led to the conclusion that the mechanisms of antimicrobial immunity are determined not only by the peptide composition, but also by features of the bacterial membrane. *Sodalis* may have adapted to life within the tsetse by cloaking itself with molecules of GroEL in order to increase resistance to an antimicrobial assault.

Although GroEL appears to be involved with immune evasion and bacterial survival, it may also contribute to symbiosis between the host and the microbe. Recent findings by Yoshida *et al.*, (2001) describe a chaperonin with dual functionality. In the salivary glands of larvae of the antlion, *Myrmeleon bore*, the Gram negative endosymbiont, *Enterobacter aerogenes,* produces a GroEL homologue that functions as both a molecular chaperone and an insect paralysing toxin. The antlion typically captures its insect prey and paralyses it by injecting salivary venom. As this toxin was found in the saliva, it was thought to be secreted and thus soluble. Exploiting this attribute, the investigators purified the protein from *E. aerogenes* culture medium, performed amino-acid sequencing to determine protein identity and consequently sequenced and cloned the *groE* gene. A comparison of this sequence to the non-toxic *E. coli groE* sequence identified 11 differing amino acid residues. Further mutation analysis revealed that four residues were critical for toxicity (Table 3.3). Importantly, toxicity was conferred to

E. coli when toxic residues replaced the four benign amino acids. They concluded that the symbiont GroEL homologue might induce paralysis in the prey by interfering with eukaryotic receptors.

Table 3.3. Comparison of GroEL sequences from several species of symbiotic and pathogenic Gram negative bacteria identifying differences between toxin-associated amino acid residues.

Species	Key residue identity and position in the GroEL protein sequence				NCBI Accession Number
	V100[a]	N101[a]	D338[a]	A471[a]	
E. aerogenes (*M. bore* isolate)[b]	***V***	***N***	***D***	***A***	AAL09389
S. oryzae principal endosymbiont	***V***	***N***	***D***	***A***	AAB97670
S. glossinidius (*G. m. morsitans* isolate)[c]	***V***	***N***	***D***	***A***	AAK92204
S. glossinidius (*Gb* isolate)[d]	***V***	***N***	***D***	***A***	AAG49528
W. glossinidia (*Gb* isolate)[d]	***V***	***N***	N	***A***	AAK07427
E. aerogenes[f]	***V***	***N***	E	***A***	BAA25215
S. typhimurium[e]	I	T	E	G	BAA94286
E. coli 0157:H7	I	T	E	G	AAG59342

Species-specific GroEL sequences were obtained from the PubMed database developed by the National Center for Biotechnology Information (NCBI http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=PubMed). Protein sequences were aligned to the endosymbiotic *E. aerogenes* (AAL09389) using the EMBL multiple alignment program Clustal W (http://www.ebi.ac.uk/clustalw/) to determine matching residue positions.

[a] Amino acid residues crucial for GroEL toxicity were determined by residue substitutions in *E. coli* mutants (Yoshida *et al.*, 2001). Those sequences above the centerline contain all four toxic residues.
[b] Endosymbiont inhabiting *Myrmeleon bore* larval salivary glands. The GroEL homologue produced by these bacteria acts as a paralytic insect toxin.
[c] Endosymbiont isolated from *Glossina morsitans morsitans* hemolymph
[d] Endosymbiont isolated from *Glossina brevipalpis* hemolymph
[e] *S. typhimurium* strain X3306
[f] *E. aerogenes* strain JM1235

It is of great interest that the GroEL homologue released by *Sodalis* possesses all four toxin-specific residues whereas the *Wigglesworthia* GroEL sequence and a non-symbiont strain of *E. aerogenes* possess only three. The *Sitophilus oryzae* (weevil)

principal endosymbiont (SOPE) also produces a GroEL containing the four crucial residues (Table 3.3). As neither the tsetse nor the cereal weevil feed on other insects, it is a mystery as to why their symbionts would secrete a GroEL that may demonstrate toxicity towards insects. Since the major protein expressed by *Sodalis* is a GroEL homologue, both soluble and membrane-bound, it is probable that a non-chaperonin function exists.

Chapter 4. Identification of a midgut-associated EP repeat protein from *Glossina palpalis palpalis*

1. Introduction

Each day, more than 100 people and 10,000 cattle in Africa die, due to the direct effects of African trypanosomiasis (Hursey, 2001). However this number is grossly underestimated as, at present, there is a resurgence of human sleeping sickness in sub-Saharan Africa, affecting greater than 500,000 Africans (Smith *et al.*, 1998; Barrett, 1999). Currently the most effective methods for disease management are tsetse population control techniques, such as aerial insecticide spraying, bait trapping and sterile male release. However, there are problems associated with implementing such practices including financial limitations, sustainability, and environmental impact (Allsopp, 2001). Thus, there is thus a need for new vector and parasite control methods. Efforts to develop a vaccine targeting the bloodstream forms of the parasite have been unsuccessful. The ability of the parasite to undergo antigenic variation has frustrated the research efforts of many scientists and in some cases has raised doubt regarding the feasibility of developing a vaccine. Some of those persisting with African trypanosomiasis vaccine development have turned their attention to studying the interactions between the procyclic trypanosome and its vector, the tsetse. Research, involving entomology, biochemistry, immunology, protein chemistry, molecular biology and bioinformatics is necessary to fully investigate and exploit the molecules within the tsetse midgut and those unique to the forms of the parasite that transit through the tsetse vector.

When bloodstream form trypanosomes enter the tsetse midgut environment after the uptake of an infective bloodmeal, they immediately begin to differentiate into insect forms (see the trypanosome life cycle, Figure 1.5, page 12). This transformation is marked by the disappearance of the variant surface glycoprotein (VSG) coat encasing the bloodstream forms and the synthesis of a new polyanionic coat comprised of stage-specific procyclic surface molecules collectively known as the procyclins. A schematic view of the array of procyclins on the surface of *T. brucei* spp. is shown in Figure 4.1.

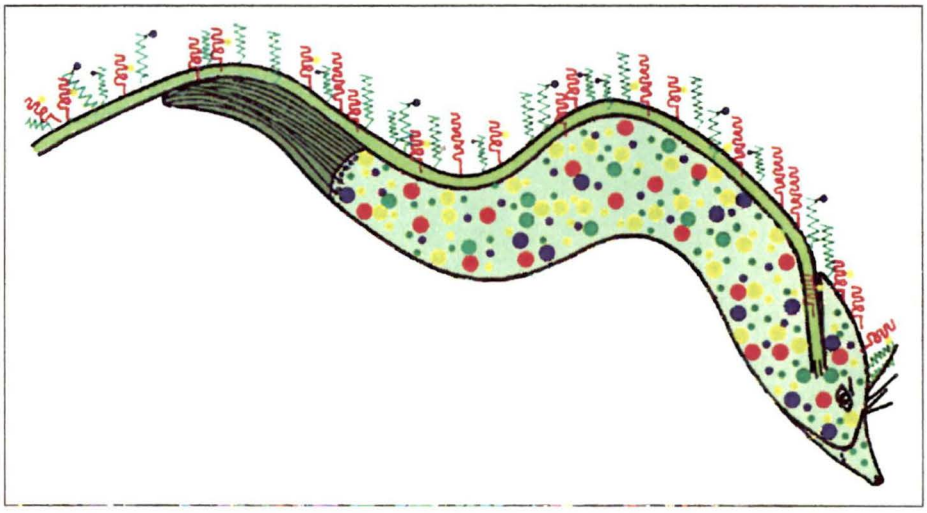


Figure 4.1. ***The "Spiny Norman" model of procyclin expression: a schematic representation of the procyclin surface coat of <u>Trypanosoma brucei spp.</u> procyclic or epimastigote forms.***

The "coat of many colours" shows the different forms of procyclins expressed during parasite transit through the tsetse vector. Red surface structures are the EP procyclins showing different EP lengths with yellow N-linked carbohydrates attached. Green surface structures are the GPEET procyclins with blue phosphates attached to threonine residues. The surface molecules cover the entire trypanosome, although for clarity, only part of the coat is shown here. The large carbohydrate structures on the GPI anchor that form the glycocalyx are not shown (from Pearson, 2001).

There are four genes that encode the two major classes of procyclins, the GPEET and EP forms. The highly phosphorylated GPEET procyclins have Gly-Pro-Glu-Glu-Thr

pentapeptide repeats and exist in two isoforms, GPEET and GPEET2. The EP procyclins have seven different isoforms: EP1-1, EP1-2, EP1-3, EP2-1, EP2-2, EP2-3, and EP3 that vary in the length of the internal EP dipeptide repeat region (18-30) and in glycosylation patterns (Acosta-Serrano *et al.*, 2000).

Recently, mass spectrometry experiments have shown that within 24 hours post-ingestion by tsetse, the trypanosomes co-expressed both glycosylated EP procyclins and phosphorylated GPEET procyclins. As development proceeded within the midgut (and *in vitro*) GPEET expression was down regulated and the glycosylated EP forms completely replaced the GPEET form on the cell surface (Vassella *et al.*, 2001). These dominant EP procyclins persisted until the trypanosomes reached the salivary glands and epimastigote differentiation commenced. Unfortunately, too few epimastigote forms were isolated to identify which procyclins were expressed in response to maturation signals within the salivary gland (Acosta-Serrano *et al.*, 2001).

One function of the procyclins may be to provide protection against the protease-rich environment of the midgut. The EP repeat sequences in the EP procyclins are resistant to proteolytic degradation by most known proteases. When bloodstream forms lacking the procyclic armor are exposed to proteases (*in vivo* and *in vitro)*, they are rapidly degraded (Sbicego *et al.*, 1999). However, evidence in contradiction to this hypothesis was obtained from procyclin knockouts. Removal of the EP procyclin coat in deletion mutants not only had insignificant effects on morphology and growth, but their sensitivity to proteases also remained unaffected. It is possible however, that the other cell-surface procyclins (GPEET forms) still protected the membranes of procyclic mutants as it was impossible to generate a null GPEET knockout (Ruepp *et al.*, 1997).

In response to unknown environmental cues, differential procyclin expression occurs (Vassella *et al.*, 2001). This may signify a role in stage-specific signaling. As a trypanosome replicates within the tsetse midgut and becomes established, the level of expression of the procyclins fluctuates in what appears to a programmed response to extracellular signaling. Sbicego *et al.*, (1999) also proposed that the procyclins might be involved in a type of developmental regulation. They found that EP procyclin null mutants were completely incapable of establishing heavy midgut infections. The authors suggested that the EP mutants were impaired in their ability to mature *in vivo* due to distorted parasite-tsetse interactions (Ruepp *et al.*, 1997). This implies that differentiation and maturation may rely on interactions between the EP procyclins and molecules within the insect vector. The roles of the different procyclins and their isoforms are still unclear. How do the procyclins contribute to trypanosome differentiation, establishment and maturation in the tsetse? How does the tsetse contribute to the survival or destruction of the parasite and does the interaction involve the procyclins at all?

It is this latter question that initiated the analysis of a procyclin-like molecule expressed in the midgut of the tsetse. Here I report on the cloning and sequencing of genes encoding an EP repeat protein from both *Glossina morsitans morsitans* and *Glossina palpalis palpalis.* Although there is a high sequence similarity between the two tsetse EP proteins, *G. p. palpalis* expresses a form that contains 27 additional EP repeat units. Computer analysis of the translated protein sequences using multiple alignment tools and protein prediction algorithms have revealed that these proteins are unique to the tsetse. Does the presence of this midgut molecule provide a clue to trypanosomal

immune evasion mechanisms? Is it possible that this tsetse protein signals trypanosome differentiation events or is the EP portion of the molecule only a physical structure used to prevent clumping of parasites by imparting charge repulsion and ensuring parasite dispersal in the midgut?

2. Experimental procedures

2.1. Tsetse

The *G. m. morsitans* Westwood and the *Glossina palpalis palpalis,* from which midguts were obtained, were from several subcultures maintained at the University of Alberta. All of these colonies descended from material originating near Kariba, Zimbabwe. For a brief history of the colonies, see Gooding and Jordan (1986). Throughout my experiments, the colonies were maintained at 24.5°C and approximately 60% relative humidity. Tsetse flies were fed on rabbits every other day using a protocol that conformed to guidelines of the Canadian Council on Animal Care.

Intact midguts were prepared from mixtures of female and male of both species. Teneral (i.e. newly eclosed, unfed) adults aged 24 to 48 hours were placed in a glass shell vial and immobilised by cooling on ice for 30 minutes. After removing the wings and legs, the flies were submersed in sterile saline (0.85% NaCl) and positioned ventral side up under a dissecting microscope. Incisions were made along the sides of the abdomen, using fine tipped surgical scissors, to expose the digestive tract. To isolate intact midguts, the entire organ was gently teased free of adhering fat body and severed immediately posterior to the crop and anterior to the proctodaeum. With all preparations, excised tissues were removed from the dissecting tray and excess saline blotted away, taking care to minimise loss of midgut contents. After transferring the midgut tissue to sterile Eppendorf microcentrifuge tubes (pre-cooled on dry ice) and freezing, the tubes were stored at -80°C, shipped on dry ice to the University of Victoria and then stored at -80°C until used for protein fractionation.

2.2. *One-dimensional gel electrophoresis*

Midguts (15 µl) were resuspended in 2X Laemmli sample buffer (Laemmli, 1970) and 25 µl of each were loaded onto 10% acrylamide, 0.75 mm 1-dimensional sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) minigels and run on a Mini-Protean II apparatus (Bio-Rad Laboratories, Hercules, CA). Proteins were electrophoresed at 50V for 30 min and then at 90V for an additional 90 min. A 10 kDa molecular weight ladder (#10064-012; Gibco BRL, Burlington, ON) was run on each gel. Following electrophoresis, gels were either stained with GelCode® Blue (Pierce Chemical, Rockford, IL), silver stained (Merril *et al.*, 1981) or transferred onto membranes for subsequent immunoblotting.

2.3. *Immunoblotting*

Immunoblotting using BioTrace ™ polyvinylidene (PVDF) membrane (Pall Corporation, Ann Arbor, MI) was performed as previously described (Beecroft *et al.*, 1993) with the following exceptions: SuperSignal Dura chemiluminescence substrate (Pierce Chemical Company, Rockford, IL), undiluted primary antibody (mAb 247; mouse anti-EP repeat procyclin, (Richardson *et al.*, 1986; Richardson *et al.*, 1988) and a 1:50,000 dilution of secondary antibody (goat anti-mouse IgG/IgM-horseradish peroxidase conjugate; Caltag Laboratories, South San Francisco, CA). Kodak Biomax MR film (Eastman Kodak Company, Rochester, NY) was used to detect chemiluminescence.

2.4. Screening of cDNA libraries

A *Glossina morsitans morsitans* whole fed fly cDNA expression library was kindly donated by Dr. Isabel Roditi (University of Bern, Bern, Switzerland). The library was constructed using the λZipLox™ System (Gibco BRL, Paisley, Scotland) according to the manufacturer's instructions. Briefly, poly (A)+ mRNA from whole *G. m. morsitans* was isolated, reverse transcribed to yield second strand cDNA, and double stranded cDNA was directionally cloned into the λgt11 cloning vector. The library host strain, *E. coli* Y1090(ZL) (Gibco BRL, Burlington, ON, Canada), was propagated in Luria-Bertani (LB) medium (1 % tryptone, 0.5 % yeast extract, 1% NaCl (w/v) pH 7.0) supplemented with 0.2% (w/v) maltose and 10 mM $MgSO_4$. A *Glossina palpalis palpalis* cDNA library was constructed using RNA isolated from whole unfed (teneral) flies according to the protocol in the Zap Express™ cDNA Synthesis Kit Instruction Manual (Stratagene, La Jolla, CA, USA). Six plates of each of the cDNA libraries were prepared by adding the λgt11 library or the Lambda ZAP Express Library (10^5 pfu/plate) to 100 μl of an overnight *E. coli* culture adjusted to an OD_{600} of 0.5 in 10 mM $MgSO_4$. *E. coli* Y1090(ZL) or *E. coli* XL1-Blue MRF' were used as hosts, respectively. The mixtures were incubated for 15 min at 37°C, added to 9 ml molten (51°C -55°C) NZY top agarose (0.5% NaCl , 0.2% $MgSO_4$, 0.5% yeast extract, 1% casein hydrolysate, 0.75% agarose (w/v) pH 7.5) and poured onto pre-warmed NZY agar (1.5% w/v) in 150 mm plates. The solidified plates were incubated at 37°C for 5 hours, overlaid with 137 mm Duralon UV™ noncharged nylon membranes (Stratagene) and bacteria/phage transferred onto the membranes for 2 min at 37°C. Membranes were oriented with needle marks and processed according to the manufacturers' protocol. To prevent membranes from

sticking to the top agar upon removal, inverted plates were chilled at 4°C for 2h. Duplicate lifts were prepared by removing the master membrane and overlaying the plates with a second membrane, followed by a 10 min incubation at 37°C. Template plates were stored at 4°C. Membrane lifts were denatured for 2 min at room temperature in denaturing solution (1.5 M NaCl, 0.5 M NaOH) followed by neutralization for 5 minutes in a solution of 1.5 M NaCl, 0.5 M Tris, pH 7.5. The membranes were then briefly rinsed in 2X SSC buffer (3 M NaCl, 0.3 M sodium citrate, pH 7.0) at room temperature and blotted using filter paper. Phage DNA was crosslinked to the membranes using the "autocrosslink" setting on a Stratagene UV Stratalinker™ 2400 (120,000 μJ of UV). Once crosslinked, filters were air dried overnight, sealed in Saran wrap and stored at 4°C until further use.

2.5. Screening of cDNA libraries with ^{32}P labeled oligonucleotides

The dried, crosslinked plaque lifts were prepared by first washing for 1 h at 25°C in 3X SSC, 0.1% (w/v) SDS to remove any probe-interfering agents such as residual agar. An additional 30 min wash was conducted at 55°C in the same solution. The membranes were then transferred to the pre-hybridization solution (6X SSC, 2X Denhardt's solution, 0.25% SDS) for 3 h at 55°C with shaking while the oligonucleotide probe was being labeled. The oligonucleotide probe (used to screen both cDNA libraries) was a 58-mer construct, since shorter probes used previously gave negative hybridization results. Approximately 100 nmol of the oligonucleotide (5'-[GAGCCA]$_6$GACCCAGAGCCAGAGCCTGACC-3'; made by Gibco BRL), corresponding to an EP dipeptide coding region, were labeled with [γ-^{32}P] adenosine 5'

triphosphate (Amersham Pharmacia Biotech, Inc., Baie d'Urfé, PQ) using T4 polynucleotide kinase (New England Biolabs). Three µl of the probe were labeled at the 5' end in a 50 µl reaction volume using 50 µCi of [γ^{32}-P] dATP (specific activity 6000 Ci/mM) and 3 units of T4 PNK in1X PNK buffer (70 mM Tris-HCl, 10 mM $MgCl_2$, 5 mM DTT (pH 7.6)). The reaction was incubated at 37°C for 1 h followed by the removal of unincorporated nucleotides on a Sephadex® G-25F gel permeation column (Amersham Pharmacia Biotech, Inc., Baie d'Urfé, PQ). The probe was then added to 50 ml of hyrbridization solution (6X SSC, 2X Denhardt's solution, 0.25% SDS) and added to the membranes. The hybridization was conducted in large glass culture dishes for 24 hours in a shaking (75 rpm) dry incubator set at 55°C. Post-hybridization, the probe was removed and stored at -20°C. To remove background radiation, the membranes were washed three times (45 min, 40 min, 30 min) with mild agitation in wash solution (1X SSC, 0.1% SDS) at 55 °C. The wet membranes were then sealed in heavy-duty plastic bags (taking care to remove all bubbles) and placed inside erased Phosphor Screen cassettes (Molecular Dynamics, Sunnyvale, CA) for 24 hours. Autoradiography was performed using a STORM 820 Imager and images were analyzed with the ImageQuant analysis software (Molecular Dynamics).

2.6. Isolation of RNA from tsetse midguts

Glossina palpalis palpalis poly A^+ midgut mRNA was either obtained from Dr. Serap Aksoy (Yale University, NY, USA) or was isolated using a MicroPoly(A)Pure™ mRNA Isolation Kit (Ambion Inc, Austin, TX). An Eppendorf microcentrifuge tube containing five teneral midguts from *G. p. palpalis* was thawed on ice and briefly centrifuged to

pellet the tsetse tissues. Excess PBS (acquired during midgut dissection) was removed and 50 µl of ice cold lysis buffer (proprietary solution containing guanidinium isothiocyanate; (Ambion Inc.) were added to the midguts. The resultant slurry was transferred to an RNAse free microtube containing an additional 50 µl of lysis buffer and 50 mg autoclaved silica sand. The tissues were ground in the microtube for two minutes using a disposable Pellet Pestle™ (KONTES Glass Company, Vineland, NJ). The lysate was briefly agitated using a vortex mixer and diluted with 220 µl of dilution buffer (10 mM Tris pH 7.5, 1 mM EDTA). To pellet the sand and particulate debris, the microtube and contents were centrifuged at 12,000 *x g* for 15 min at 4°C. The lysate was transferred to a clean microtube containing 20 mg of oligo (dT) cellulose resin (Ambion Inc) and incubated at room temperature with gentle agitation for 1 h to allow hybridization of the poly(A) tails to the resin. The mixture was then centrifuged at 3,000 *x g* for 3 min at room temp, the supernatant was removed and the pellet was washed sequentially with high salt (Binding Buffer) and low salt (Wash Buffer) solutions. The resin (resuspended in 400 µl wash buffer) was loaded into a spin column and washed 4 times with wash buffer. The elution buffer (10 mM Tris pH 7.5, 1 mM EDTA) was pre-warmed to 65°C and 100 µl were added to the column. The column was centrifuged at 5,000 *x g* for 30 sec at room temperature and the eluted poly(A)$^+$ RNA was collected into a clean microtube. A second elution with 100 µl of elution buffer was performed, yielding a final volume of 200 µl of poly (A)$^+$ RNA. To concentrate the mRNA, 20 µl of 5 M ammonium acetate were added to the eluate followed by 1 µl of glycogen (5 mg/ml) and 550 µl of ice cold 100% EtOH. The solution was gently agitated and incubated at

-20°C overnight to precipitate the mRNA. The following day, the solution was centrifuged at 14,000 *x g* for 20 min at 4°C to pellet the mRNA. The EtOH was removed and the pellets were air dried for 10 minutes prior to resuspension in 20 µl DEPC-treated water/0.1 mM EDTA. Finally, the concentration of mRNA was determined by spectrophotometric analysis at OD_{260} with the conversion: 1 OD_{260} = 40 ng/µl. The RNA was stored in 3 µl aliquots (310 ng/µl) at -70°C prior to further use.

2.7. First strand 5'-RACE amplification of cDNA

Poly(A)$^+$ RNA from midguts of teneral *G. p. palpalis* was used as the template for cDNA synthesis. The SMART™ RACE cDNA Amplification Kit (CLONTECH Laboratories, Inc., Palo Alto, CA, USA) was selected for the isolation of the complete 5' sequence of the EP repeat gene from *G. p. palpalis*. First-strand cDNA was synthesized according to the manufacturer's protocol. In brief, 0.93 µg (3 µl) of the purified poly(A)$^+$ RNA were incubated for 2 min at 70°C in a 5 µl reaction mixture containing 1 µl (10 µM) of a modified oligo (dT) primer (5'-$(T)_{25}N_{-1}N$-3') and 1 µl (10 µM) of the SMART II oligonucleotide (5'-AAGCAGTGGTAACAACGCAGAGTACGCGGG-3"). The reaction tubes were first cooled on ice. Two µl of 5X first-strand buffer (250 mM Tris-Cl, pH 8.3, 375 mM KCl, 15 mM $MgCl_2$), 1 µl of 20 mM dithiothreitol (DTT, Gibco BRL), 1 µl of dNTP mix containing 10 mM of each nucleotide at neutral pH (Amersham Pharmacia Biotech, Inc., Baie d'Urfé, Quebec) and 200 units of SuperScript™ RNase H$^-$ reverse transcriptase (Canadian Life Technologies, Burlington, ON) were cooled and then added to the PCR reaction tubes. The resulting mixtures were incubated at 42°C for 90

min. First-strand reaction mixtures were diluted to 200 µl final volume with Tricine-EDTA buffer (10 mM Tricine-KOH , pH 8.5, 1 mM EDTA) and heated for 7 min at 72°C. Prior to second-strand cDNA synthesis, the cDNA:RNA hybrids were incubated for 5 min with 4 units of RNase H (United States Biological Corp, Cleveland, OH, USA) at 37°C. This last step removes RNA that could interfere with subsequent PCR amplification reactions. The samples were analysed on 1% agarose gels next to a lane of human placental total RNA (the kit control). Positive cDNA synthesis reaction mixtures were stored at -70°C for no longer than 2 months.

2.8. PCR amplification of the EP repeat 5' sequence from G. p. palpalis cDNA

For each PCR amplification, the reaction mixtures (25 µl total volume) consisted of 25 pmol of gene-specific forward and reverse primers, 2.5 µl of 10X *Taq* DNA polymerase buffer (200 mM Tris-HCl pH 8.4, 0.5 M KCl) (Canadian Life Technologies, Burlington, ON), 2 µl of the cDNA template, 1 µl of dNTP stock containing 10 mM of all four of the dNTPs, each at a final concentration of 10 mM pH (Amersham Pharmacia Biotech, Inc., Baie d'Urfé, Quebec), 0.5 µl – 6.0 µl of 10 mM $MgCl_2$ and sterile PCR grade water to final volume of 24.5 µl. When reaction tubes reached the denaturation temperature of 72°C, 0.5 µl of recombinant *Taq* polymerase (supplied by Dr. Chris Upton, Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada) was added to the mixture. For all 5'RACE PCR reactions, 2 µl of the forward universal primer mix (UPM) recognizing the 5' SMART oligo sequence were substituted for the gene specific forward primer. A primer set designed to amplify the alpha tubulin gene from *Glossina* (gift from Dr. Serap Aksoy) was used as a cDNA positive control. The

human placenta cDNA and complementary forward and reverse transferrin primers supplied by CLONTECH Laboratories, Inc. (Palo Alto, CA, USA) served as the positive amplification control. The sequences of the primers used are shown in Table 4.1. All gene specific primers were designed to be less than 30 nucleotides in length, to have a GC content between 50-70% and to have melting temperatures greater than 65°C as determined by the nearest neighbour method. The software program, Primer Premier (PREMIER Biosoft International, Palo Alto, CA) aided in the design of these primers while eliminating those constructs predisposed for false priming sites, dimers and hairpin formations. The oligonucleotides were constructed using the preferred codon usage of tsetse collected from previously published *G. m. morsitans* DNA sequences (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=Glossina+morsitans+morsitans+%5Bgbinv%5D&aa=1&style=GCG). The nested primer, Gp5.3 unique (encoding the amino acid sequence CLIEGLPEPE) used the sequence information obtained from the cDNA library clone and includes the last 18 nt unique region and the first two dipeptide repeats. The tsetse alpha tubulin primers (from Dr. Serap Aksoy) amplified a PCR product of approximately 360 bp that encodes a 121 amino acid sequence located centrally in the *G. m. morsitans* protein (NCBI accession number AAG49533).

The polymerase chain reactions (PCR) were performed in a thermal cycler (GeneAmp PCR System 2400; Perkin-Elmer, Norwalk, CN) for 35 cycles at 94°C for 30 sec (denaturation), at 55°C for 45 sec (annealing), and 65°C for 2 min (extension), followed by a 10 min final extension at 70°C. The upstream noncoding and downstream unique

coding regions were amplified using the 5' forward universal primer mix and the 3' nested Gp 5.3 reverse primer.

2.9. Agarose gel electrophoresis

Unless otherwise indicated, 1% (w/v) agarose gels were used. Molecular biology grade agarose LE (Promega, Madison, WI, USA) was dissolved in 0.5X Tris-acetate/EDTA (TAE) electrophoresis buffer (20 mM Tris acetate, 0.5 mM EDTA pH 8.5) to prepare the gel. Prior to gel casting, 5 µl of ethidium bromide solution (stock 10 mg/ml, Gibco BRL, Burlington, ON, Canada) were added to the molten agarose once it had cooled to 55°C. *Hind* III digested lambda DNA standards (New England Biolabs (NEB), Beverly, MA, USA), a 100 bp DNA ladder (NEB) or a 1:1 mixture of the two products served as molecular size standards for each gel. The ethidium-bromide labeled DNA was visualized and recorded using a ChemiImager™ 4000 equipped with AlphaEase™ image processing and analysis software (Alpha Innotech Corporation, San Leandro, CA, USA). The images obtained were saved digitally in TIFF format.

2.10. Cloning and transformation

Amplified PCR products were gel purified using a QIAEX II Gel Extraction Kit (Qiagen Inc., Mississauga, ON) and cloned into *E. coli* TOP 10 cells (TOPO cloning Kit; Invitrogen, Carlsbad, CA) for blue/white screening without IPTG induction. Ten transformants were grown overnight in Luria-Bertani broth supplemented with ampicillin (100 µg/ml) and plasmid purification was performed using a Qiaprep Spin MiniPrep Kit (Qiagen Inc.) according to the manufacturers' instructions. Of the ten clones, four were

submitted to the Centre for Environmental Health, University of Victoria, for automated dideoxynucleotide sequencing.

2.11. Radiolabelling of the EP specific probe, Gp5.3, for northern blot analysis

To determine if the transcript encoding the EP repeat in *G. p. palpalis* was expressed in midgut tissue, a radiolabelled probe encoding the EP rich sequence was used. The probe was excised from the phagemid vector pBK-CMV previously confirmed to carry the DNA insert coding the 3' end of the *G. p. palpalis* EP repeat sequence. *E. coli* XLOLR (clone Gp5.3) was grown in Luria-Bertani broth supplemented with Kanamycin A (50 mg/μl, Sigma Chemical Co., St. Louis, MO, USA) overnight at 37°C. Phagemids were isolated from the bacteria using a Qiaprep Spin MiniPrep Kit (Qiagen Inc., Mississauga, ON, Canada) according to the instructions of the manufacturer. To purify the insertion sequence, a double restriction digest was performed, using 5 units *Xho* I (Gibco) and 10 units of *Eco* RI (New England Biolabs) with 1 μg of phagemid DNA as target. After an incubation at 37°C for 2 h, the digestion mixture was electrophoresed on a 1% (w/v) OmniPur, low melt agarose gel (EM Science, Gibbstown, NJ, USA) made with 1X TAE buffer and 50 μg ethidium bromide. The successful separation of the insert from the phagemid vector was verified by briefly visualizing the DNA in the gel with UV light (FotoDyne). The band corresponding to insert DNA (740 bp) was excised and the DNA was extracted using a QIAEX II Gel Extraction Kit (Qiagen Inc., Mississauga, ON). The DNA was eluted with 10 mM Tris-HCl (pH 8.5) and isolation was confirmed again by gel electrophoresis. The final concentration of the Gp5.3 insert used for further radiolabelling was 32.5 ng/μl (determined with the conversion: 1 OD_{260} = 50 ng/μl). Two

µl of the Gp5.3 DNA were denatured by boiling for 2 min. The probe was labeled at the 5' end in a 50 µl reaction volume using 50 µCi of [γ^{32}-P] adenosine 5' triphosphate (specific activity 6000 Ci/mM; Amersham Pharmacia Biotech, Inc., Baie d'Urfé, PQ) and 3 units of T4 polynucleotide kinase (New England Biolabs) in the appropriate PNK buffer dilution (supplied by NEB with enzyme). The reaction mixture was incubated at 37°C for 90 min followed by the removal of unincorporated nucleotides on a Sephadex® G-25F gel permeation column (Amersham Pharmacia Biotech, Inc., Baie d'Urfé, PQ). The amount of radionucleotide incorporation was evaluated by liquid scintillation counting as 26,260 cpm/µl of probe. The Gp5.3 radiolabelled probe was stored at -20°C until required.

2.12. Northern blot analysis

To determine whether the *G. p. palpalis* EP repeat mRNA transcripts were found in tsetse midgut tissue, northern blot analysis was performed. A 1.4% (w/v) agarose gel in 10 mM Na_2HPO_4 buffer, pH 6.8 containing 1 µl of 10 mg/ml ethidium bromide (Gibco BRL, Burlington, ON, Canada) per 100 ml of buffer was poured. *G. p. palpalis* poly $(A)^+$ RNA (0.1 µg) was diluted in sterile distilled water to a volume of 10 µl and mixed with 2 µl of sterile 6X loading buffer (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 30% (v/v) glycerol, 1.2% (w/v) SDS, 60 mM sodium phosphate, pH 6.8). The mixture was incubated at 75°C for 5 min to rapidly denature the RNA instead of using traditional glyoxal or formaldehyde denaturation (Pelle and Murphy, 1993). Samples were immediately loaded onto the gel and electrophoresed for 3 hours at 5 V/cm. Constant recirculation of the sodium phosphate buffer was maintained by stirring with small

magnetic stirring bars in each end of the gel apparatus. Following electrophoresis, the RNA was vacuum transferred to Hybond N^+ nylon membrane (Amersham Pharmacia) pre-wetted in transfer buffer (20X SSC, pH 7.0) using a VacuGene™ XL Vacuum blotting System (Pharmacia Biotech AB, Uppsala, Sweden). The vacuum was stabilized between 50 and 55 mbar and the RNA was transferred for 30 min. Following transfer, the membrane was briefly washed in 2X SSC buffer. A UV-transilluminator (FotoDyne) was used to ascertain the efficiency of RNA transfer to the membrane following blotting. The transferred RNA was then fixed to the nylon membrane using a Stratagene UV Stratalinker™ 2400 (120,000 μJ of UV). Once crosslinked, the membranes were air dried and stored between layers of filter paper at 4°C until use. The prehybridization, hybridization and washing steps were performed as previously described for cDNA library screening. The blot was exposed for 72 h. Autoradiography was performed using a STORM 820 storage phosphor imaging system (Molecular Dynamics, Sunnyvale, CA, USA).

2.13. DNA sequencing

Pure double-stranded plasmid (pZLI Lambda ZipLox™ vector and pCR II TOPO™ cloning vector) or phagemid pBK-CMV ZAP Express™ vector DNA containing an EP repeat gene fragment were submitted to the Centre for Environmental Health, University of Victoria, Victoria, BC and automated dideoxynucleotide sequencing was performed using a NEN Global IR2 DNA Sequencer (LI-COR, Inc., Lincoln, NB, USA). The traces for each sample (generated by e-Seq vs.2.0 (LI-COR, Inc.) were visually examined for errors and ambiguous regions. The sequence data from four positive clones was selected

for assembly and alignment using the SeqmanII vs. 5.03 sequence alignment software suite from DNASTAR, Inc. (Madison, WI, USA) for derivation of the final consensus sequence.

2.14. Database searches, sequence alignments and computer characterization

Database searches were performed using the National Centre for Biotechnology Information (NCBI) Basic Local Alignment Search Tool algorithm (BLASTP® 2.2.1, updated Apr-13-2001) (Altschul *et al.*, 1997) on all non-redundant GenBank CDS (coding derived subsequence) translations, Brookhaven Protein Data Bank (PDB), SwissProt (Swiss Institute of Bioinformatics, Geneva, Switzerland), PIR (Protein Information Resource Washington, DC) and PRF (Protein Research Foundation Osaka, Japan). The nucleotide sequences were conceptually translated into protein sequences using the software program, Primer Premier (PREMIER Biosoft International, Palo Alto, CA). Multiple sequence alignments of the polypeptide sequences for the EP repeat proteins and the trypanosomal EP procyclins were performed using Clustal W™ (http://www.ebi.ac.uk/clustalw/) multiple alignment program (Thompson *et al.*, 1994). Antigenic analysis of protein sequences was accomplished by using the DNASTAR protein analysis software program Protean™ (Lasergene, Madison, WI). TMHMM vs. 2.0 (transmembrane detection using the hidden Markov model approach) was used to predict transmembrane helices in proteins (Centre for Biological Sequence Analysis, http://www.cbs.dtu.dk/services/TMHMM) (Moller *et al.*, 2001). SignalP vs 2.0 (http://www.cbs.dtu.dk/services/SignalP-2.0/#submission) at the Center for Biological Sequence Analysis, Technical University of Denmark, predicted the probability and

location of signal sequences (Nielsen *et al.*, 1999). The analysis of possible protein motifs and post-translational modifications (phosphorylation and glycosylation sites) was performed by: ScanProsite (http://ca.expasy.org/tools/scnpsite.html), a tool which explores the PROSITE database for biologically significant sites, patterns and profiles (Hofmann *et al.*, 1999) (maintained by the ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics), ppsearch, (Prosite pattern search, http://www2.ebi.ac.uk/ppsearch/) at the European Bioinformatics Institute (EBI) (Hofmann et al., 1999) and ProfileScan (http://hits.isb-sib.ch/cgi-bin/PFSCAN), the protein server created by the Swiss Institute for Experimental Cancer Research (ISREC). The experimentally verified O-glycosylation database, O-GLYCBASE vs 5.00 (Gupta *et al.*, 1999), provided by the Center for Biological Sequence Analysis (Technical University of Denmark), was searched using the NetOGlyc 2.0 algorithm (http://www.cbs.dtu.dk/services/NetOGlyc/) for potential O-glycosylation sites (Hansen *et al.*, 1998). The annotated DNA and protein sequence for the *Glossina palpalis palpalis* EP repeat protein (Gp5.3) was submitted to the National Centre for Biotechnology Information database (NCBI, Bethesda, MD) using the GenBank sequence submission tool (http://www.ncbi.nlm.nih.gov/Genbank/index.html). The GenBank accession numbers for the 1125 bp open reading frame were assigned AY077716 and AAL82540 for the 374 aa protein product.

3. Results

3.1. SDS-PAGE and immunoblot analysis of tsetse midguts

Immunoelectron microscopy on tsetse tissues had previously revealed strong labeling in midguts with a mAb that recognizes EP repeat epitopes (Tetley *et al.*, 1987). To determine if immunoreactivity was present in teneral midgut tissue, immunoblot analysis was performed on both *G. m. morsitans* and *G. p. palpalis*. The EP-specific monoclonal antibody used, TBRP1/247, was originally prepared by immunizing mice with procyclic forms African trypanosomes (Richardson *et al.*, 1988). The immunoblot results are shown in Figure 4.2, Panel A. The autoluminogram reveals distinct immunoblot patterns when the two tsetse species are compared. One major, broad band of 47-52 kDa was detected in the midgut of *G. p. palpalis* (Lane 1) whereas a band of similar intensity and shape was detected in *G. m. morsitans* (Lane 2). However, the apparent mass was much lower at 41-46 kDa. The broad band may be due to glycosylation or non-typical protein migration in SDS-PAGE as has been observed with the EP procyclins of *T. brucei* spp (Richardson *et al.*, 1988). Analysis of the protein composition by SDS-PAGE of midguts from the two tsetse species (Figure 4.2, Panel B) revealed unique protein profiles both *G. m. morsitans* and *G. p. palpalis*. Silver staining of the proteins showed that the *G. p. palpalis* midgut (Lane 1) contained a large number of proteins with masses greater than 100 kDa. In the *G. m. morsitans* midgut sample (Lane 2) the majority of proteins were less than 100 kDa. In both profiles, bands corresponding to those detected by immunoblotting were not obviously predominant.

To identify and characterize the tsetse antigens recognized by the anti-EP repeat mAb, genes encoding the immunoreactive molecules were identified by cDNA library screening.

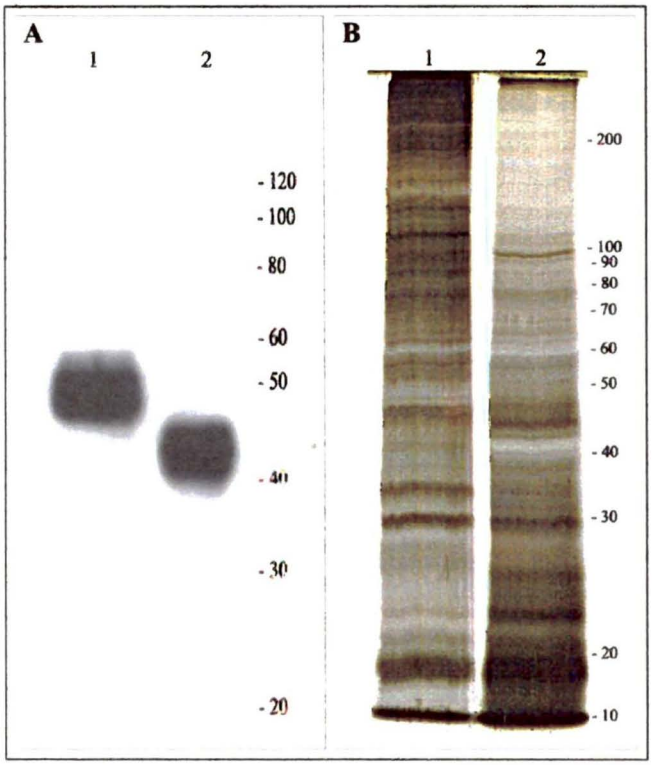


Figure 4.2. ***SDS-PAGE and immunoblot analysis of solubilized midgut from teneral <u>Glossina palpalis palpalis</u> and <u>Glossina morsitans morsitans</u>.***
A single midgut was solubilised in 40 µl of Laemmli buffer and 20 µl of this mixture were separated in a 10% gel. Panel A: Immunoblot analysis using anti EP-repeat mAb 247. Panel B: Silver stained gel profiles. Lane 1. <u>G. p. palpalis</u>. Lane 2. <u>G. m. morsitans.</u> Positions of molecular mass markers (in kilo Daltons) are indicated on the right of each panel.

3.2. Bacteriophage cDNA library screening and northern blot analysis

To obtain a DNA sequence for the *G. p. palpalis* EP repeat gene, bacteriophage expression libraries prepared from adult teneral (unfed) tsetse were screened with a radiolabelled oligonucleotide probe encoding 19 tandem EP repeats. An example of a

plaque lift showing positive hybridization is shown in Figure 4.3. The screening experiment identified multiple plaques from each library that exhibited reactivity with the probe. Five *G. m. morsitans* control plaques (Panel A) were selected for further analysis. Secondary amplifications of the positive clones from one of the plaques shown in Panel A were performed to verify positivity (Panel B). Ten plaques from *G. p. palpalis* were also selected for further analysis (not shown). One positive phage per amplified plate was selected for cloning into the appropriate *E. coli* strain.

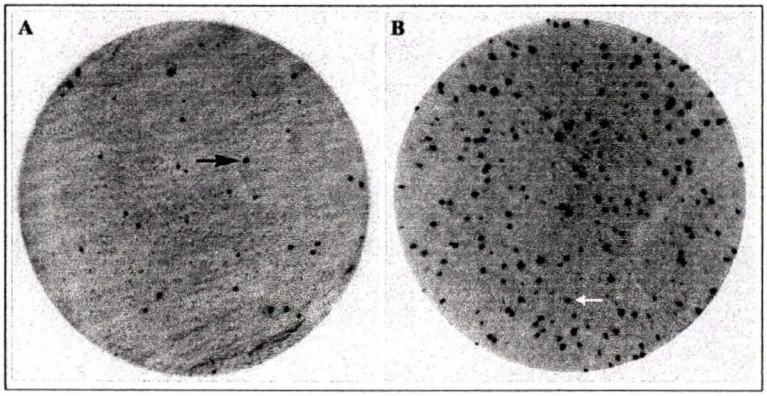


Figure 4.3. ***Representative plaque lifts from a <u>G. m. morsitans</u> cDNA library screen using a ^{32}P-labelled oligonucleotide probe encoding 19 tandem EP dipeptide repeats.***

Specific binding of the probe to membrane cross-linked plaque DNA was visualized with a STORM PhophorImager. Panel A: Primary screen. The black arrow indicates a plaque selected for amplification. Panel B: Secondary screen. The white arrow designates the plaque used for further sequence analysis.

After growth in *E. coli*, the pBK-CMV phagemids containing *G. p. palpalis* DNA inserts and pZL1 plasmids with *G. m. morsitans* DNA insertion sequences were purified and digested with *Eco* RI|*Xho* I and *Hind* III|*Kpn* I, respectively, and analyzed by agarose gels to estimate the size of the inserted sequences. All insert sizes ranged from 550 to 750 bp (data not shown). A total of four *G. m. morsitans* and six *G. p. palpalis* clones were

submitted for DNA sequencing. One *G. p. palpalis* clone (Gp5.3), containing a 739 bp sequence encoding the C-terminus and EP rich region of the protein, was selected for use as a probe for northern blotting on tsetse midgut RNA. Poly $(A)^+$ RNA was isolated from the midguts of teneral *G. p. palpalis*, thus confirming eukaryotic derivation (and eliminating RNA from bacterial origin) and tissue specificity. The northern blot showed that *G. p. palpalis* EP repeat mRNA transcripts exist within the midgut of an unfed fly (Figure 4.4). A more detailed analysis of the mRNA (transcript size, stage-specific expression etc...) was not performed as it was deemed more important to focus on obtaining the complete sequences of the EP proteins from midguts of both *G. m. morsitans* and *G. p. palpalis*.

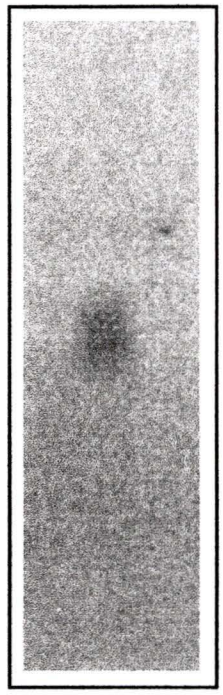


Figure 4.4. ***Northern blot analysis of teneral <u>G. p. palpalis</u> midgut poly $(A)^+$ mRNA for transcripts encoding an EP repeat sequence.***
The blot was probed with a ^{32}P-labelled cloned pBK-CMV phagemid insert containing the 3' DNA sequence (739 bp) of the EP repeat gene from <u>G. p. palpalis</u>.

3.3. Nucleotide sequence of DNA encoding the G. p. palpalis EP repeat protein

The entire nucleotide sequence encoding the *G. p. palpalis* EP repeat protein, including upstream and downstream flanking sequences, was obtained through a series of PCR amplifications of mRNA. Poly $(A)^+$ RNA was isolated from teneral *G. p. palpalis* midguts for 5' RACE cDNA synthesis. A universal forward primer mix and a gene specific reverse primer, Gp5.3 unique (designed from a combination of the *G. m. morsitans* EP repeat gene (previously sequenced in the laboratory of Drs. Isabel Roditi, Bern, and David Barry, Glasgow) and the 5' end of the *G. p. palpalis* cDNA library derived sequence), were used to identify the target sequence (Table 4.1). Reverse primer Gp5.3 unique encodes the peptide sequence CLIEGLPEP that is located in the center of the EP protein and serves as the overlap point between the two cloned fragments.

Table 4.1. Polymerase chain reaction primers used to amplify the G. p. palpalis EP repeat DNA sequence or to amplify control sequences.

Primer Name	Sequence (5'-3')	Specificity	T_m (°C)
Universal Primer Mix (F)	long: CTAATACGACTCACTATAGGGCA AGCAGTGGTAACAACGCAGAGT short: CTAATACGACTCACTATAGGGC	5' SMART oligo	>65
Gp5.3 unique (R)	CTCTGGTTCTGGCAAACCCTCAATTAAGCA	nested primer upstream of EP repeat sequence	78
Gsp2 (F)	ACGTATTCATTTCCCTTTGG	tsetse alpha tubulin	61
Gsp2 (R)	AATGGCTGTGGTGTTGGACAA	tsetse alpha tubulin	67
TFR Primer (F)	CLONTECH	human transferrin receptor	>65
TFR Primer (R)	CLONTECH	human transferrin receptor	>65

The initial round of PCR amplifications yielded a 702 bp fragment (Figure 4.5) that corresponds to the predicted product size (based on the *G. m. morsitans* sequence

information). To generate sufficient quantities of the PCR product for cloning and further sequencing, the band from lane 6 was gel eluted and re-amplified using the same primer set and an optimized salt concentration (0.8 mM). The amplified PCR fragments were cloned into the pCR II TOPO™ Cloning vector and 12 white colonies (transformants) were picked for plasmid isolation. Insert analysis by *Eco* RI restriction digest of the purified plasmids confirmed the successful cloning of the target sequence (Figure 4.6, Panel A).

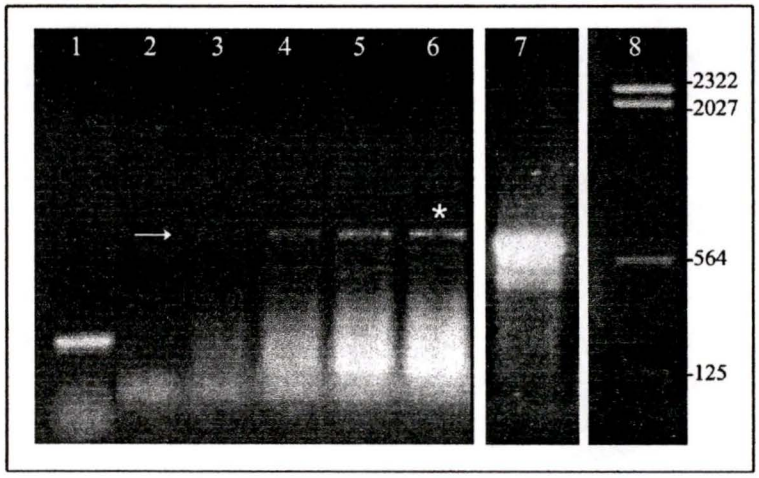


Figure 4.5. ***Agarose gel analysis of PCR amplified DNA fragments encoding the 5' region of the <u>G. p. palpalis</u> EP repeat gene.***

Lane 1: Positive control: 360 bp fragment amplified using nested primers specific for the tsetse alpha-tubulin gene. Lane 2: Negative control (lacking forward universal primer). Lanes 3, 4, 5, 6: Amplification of target sequence (arrow) with 5' RACE universal primer mix, 3' nested gene-specific reverse primer and varying final $MgCl_2$ concentrations: 0.2 mM, 0.4 mM, 0.6 mM and 0.8 mM respectively. Lane 7: Re-amplification of gel-eluted PCR product from Lane 6 (asterisk). Lane 8: <u>Hind</u> III digest of lambda DNA (molecular size standards). The sizes (in base pairs) are indicated on the right of the figure.

Of the 12 clones isolated, 9 contained insertions of similar sizes (lanes 2-7, 9-11). Interestingly, the insertion sequences have an internal *Eco* RI recognition site that

generates two cleavage products (arrow and *) which, when combined, equaled the predicted fragment size. The remaining three clones were rejected since gel electrophoresis of the uncut plasmids revealed size discrepancies indicative of empty plasmid cassettes (Panel B, 12a, 13a).

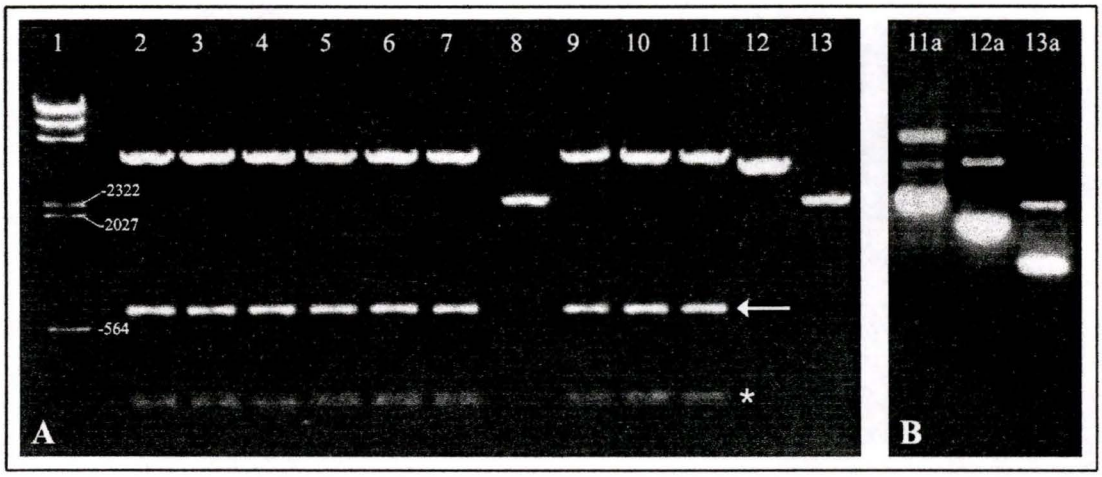


Figure 4.6. ***Agarose gel analysis of TOPO cloning vectors containing G. p. palpalis DNA inserts.***

Panel A; Lane 1: Molecular size standards (lambda phage Hind III digest). Lanes 2-13: Eco RI restriction digest of the cloning vector, pCRII, isolated from 12 colonies transformed with the Gp5.3 insert (encoding an EP repeat sequence). Panel B; Lanes 11a – 13a, whole plasmid preps (undigested), correspond to lanes 11 –13 in Panel A. The plasmid in Lane 11a contains the complete Gp5.3 insert. Lanes 12a and 13a signify either empty vector cassettes or inserts that contain mutations at the endonuclease cleavage site. Note: The 700 bp Gp5.3 5' insertion sequence contains an internal Eco RI site as shown by the creation of two digestion fragments at 600 bp (arrow) and 100 bp (*).

The 5' RACE 702 bp PCR fragment was sequenced and aligned with the 739 bp fragment encoding the 3' *G. p. palpalis* EP repeat region, the stop codon and an additional downstream segment (preceding the poly A tail). An 18-mer DNA segment overlap joined the two sequences together to form the entire cDNA transcript (Figure 4.7). The open reading frame (blue highlight) encodes a 374 residue anionic protein with

a unique N-terminal region and a C-terminal EP-rich region comprised of 85 dipeptide repeats. The translated product has a calculated molecular mass of 41,877 Da and the translated cDNA sequence, in comparison, has a molecular mass of 52,726 Da, as computed by the ProtParam prediction tool (see Materials and Methods).

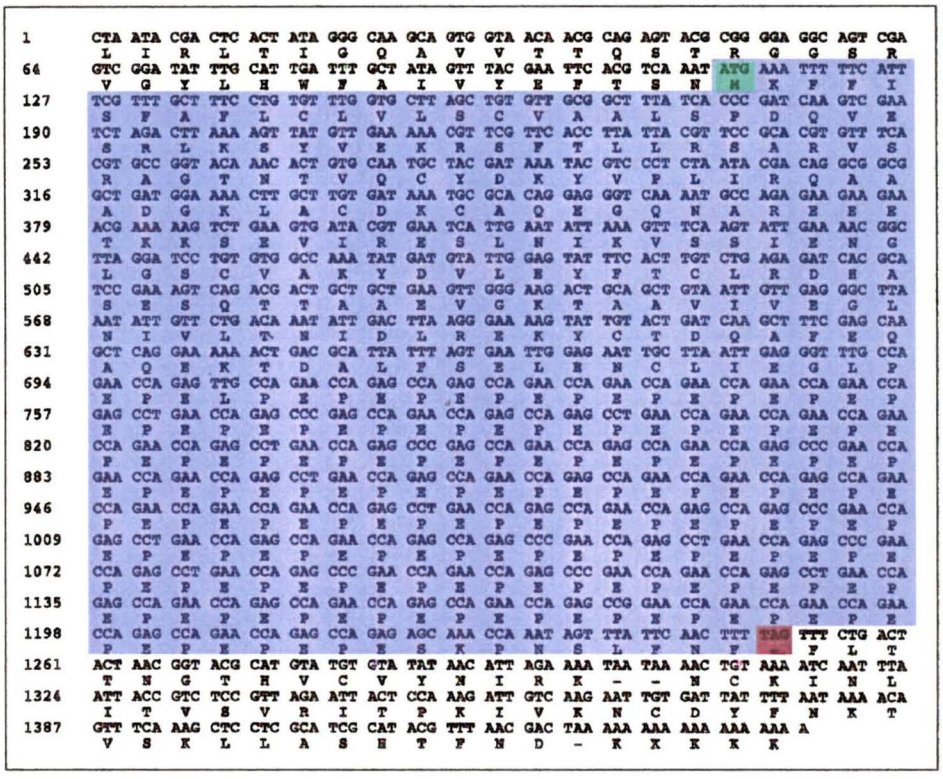


Figure 4.7. ***The nucleotide sequence and translated amino acid sequence of <u>G. p. palpalis</u> cDNA and protein: the glutamic acid-proline (EP) repeat molecule.***

Blue = coding region. Green = initiation codon and N-terminal methionine. Red = stop codon.

3.4. Sequence alignment analysis

Searching the non-redundant protein and DNA databases with the complete *G. p. palpalis* EP repeat sequence, using the BLAST algorithm (Altschul *et al.*, 1997), revealed the only significant hit to be the predicted tsetse (*G. m. morsitans*) EP protein (NCBI accession number CAC86027). Other lower scoring hits identified bacterial

proteins containing glutamic acid:proline-rich domains. The two outstanding examples are a *Cyanobacterium synechocystis* hypothetical protein slr1753 (PIR reference S74322) containing 104 EP dipeptide repeats and a hypothetical *E. coli* 0157:H7 (NCBI accession NP_311534) protein with 25 EP repeats.

Sequence alignment (Figure 4.8) of the tsetse EP protein (*G. m. morsitans*) and the *G. p. palpalis* EP repeat protein was performed using the Clustalw alignment tool (Thompson *et al.*, 1994). Despite the fact that the *G. p. palpalis* protein contains 27 more EP repeats, the overall alignment of the open reading frames from both tsetse species shows a 83 % sequence similarity, indicating that the protein is conserved between the savannah and riverine tsetse subgenera.

```
Gmm   -----------------------------------MKFFISFAFLCLVLSCVAALSPD 23
Gpp   LIRLTIGQAVVTTQSTRGGSRVGYLH-FAIVYEFTSNMKFFISFAFLCLVLSCVAALSPD 59
                                          ************************

Gmm   QVESRLKSYVENRSFALLRSGRVARAGTNTVQCYDKYVPLIRQAAAEGKFASDKCAQEGQ 83
Gpp   QVESRLKSYVEKRSFTLLRSARVSRAGTNTVQCYDKYVPLIRQAAADGKLACDKCAQEGQ 119
      ***********:***:****.**:************************:**:*.********

Gmm   NAREEETKKSEVIRESLNVKVASIEKGLGACVAKYDVLEYFTCLRDHASESQTAAGEVGK 143
Gpp   NAREEETKKSEVIRESLNIKVSSIENGLGSCVAKYDVLEYFTCLRDHASESQTTAAEVGK 179
      ******************:**:***:***:************************:*.****

Gmm   TSAVIVEGLNIVLTNIELREKYCIDQAFEQAQEKSDILFTELENCLIEGLPEPEPEPEPE 203
Gpp   TAAVIVEGLNIVLTNIDLREKYCTDQAFEQAQEKTDALFSELENCLIEGLPEPEPEPEPE 239
      *:***************:******* ********** :* **:*********************

Gmm   PEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE 263
Gpp   PEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE 299
      ************************************************************

Gmm   PEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPQPEPE------------ 311
Gpp   PEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE 359
      ********************************************:****

Gmm   ----------------------------------------SKPNSLFNF--------- 320
Gpp   PEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPESKPNSLFNF-FLTTNGTH 418
                                               *********

Gmm   ------------------------------------------------------
Gpp   VCVYNIRK--NCKINLITVSVRITPKIVKNCDYFNKTVSKLLASHTFND-KKKKK 470
```

Figure 4.8. ***Comparison of the translated amino acid sequences of the midgut EP repeat proteins of G. m. morsitans (Gmm) and G. p. palpalis (Gpp).***
Clustalw 1.01 multiple sequence alignment program was used to produce the comparative alignment. (*) indicates an identical amino acid. (:) indicates a highly conserved residue change. (.) indicates a moderately conserved amino acid substitution.

The *G. p. palpalis* EP repeat protein sequence was also aligned with five trypanosome EP procyclin sequences (EP1-2, EP2-1, EP3-2, EP3-3, EP3-4). Surprisingly, the alignment showed sequence homology at five N-terminal sites in addition to the EP-rich region. The alignment with one of the procyclins, EP2-1 is shown to illustrate this (Figure 4.9).

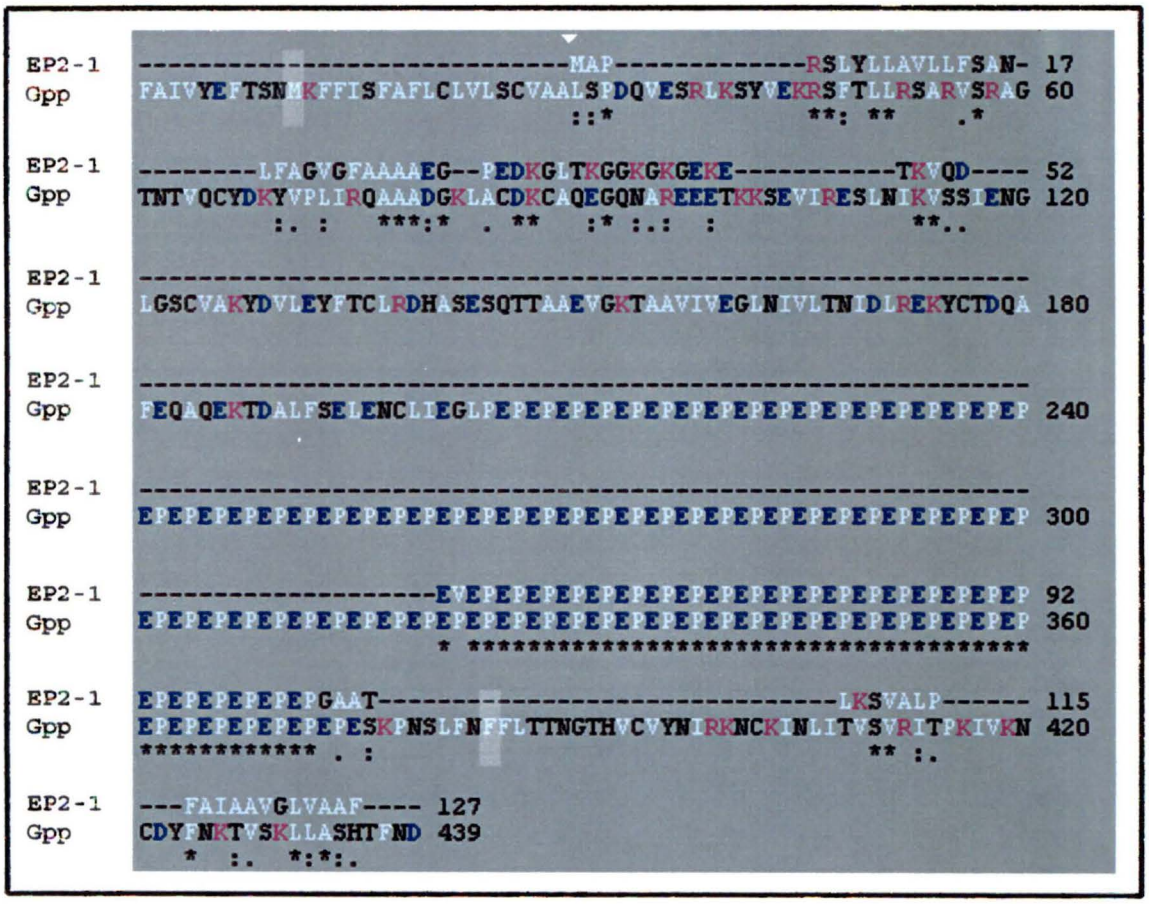


Figure 4.9. ***An alignment of the trypanosomal procyclin EP2-1 and the G. p. palpalis EP repeat protein sequences.***
The deduced amino acid sequence of G. p. palpalis EP repeat protein (Gpp) was aligned to the surface protein EP2-1 procyclin precursor sequence from T. b. brucei AnTat 1.1 (NCBI accession number AAK62890) using Clustal W. (*) indicates an identical amino acid. (:) indicates a highly conserved residue change. (.) indicates a moderately conserved amino acid substitution. The initiation methionine and the last residue (F) of the open reading frame are indicated by highlighting. The (▾) indicates the predicted cleavage site (VAA-LS) of the tsetse signal peptide. The colour of each residue designates the nature of the amino acid: acidic (blue), basic (pink), small and hydrophobic (white) and hydrophilic (black).

3.5. Computer assisted protein prediction analysis

The deduced amino acid sequences of the *Glossina* EP repeat proteins were submitted to several online structure and function prediction programs. Some of the significant physical properties are tabulated below (Table 4.2). For all of the sequences, a 6-7 kDa discrepancy (approximately) exists between the predicted mass (based on amino acid composition) and the observed mass determined by immunoblotting (Figure 4.2, Panel A). The migration of these highly anionic proteins in SDS-PAGE may be aberrant, as the negatively charged SDS is likely to be repelled by the EP repeat region, subsequently altering the charge to mass ratio. This has been observed with the EP procyclins of African trypanosomes (Richardson *et al.*, 1988). Nevertheless, the calculated M_r and the observed M_r independently confirmed a 6 kDa difference between the ORF's encoding the EP proteins of *G. m. morsitans* and *G. p. palpalis*. The 27 additional EP repeats in the *G. p. palpalis* protein totals 6202 Da. The theoretical isoelectric point is the pH value where the net charge of the protein is neutral. Although this is valuable information for comparative analysis, the probability of this uncharged state existing within the acidic environment of the midgut is low. Hydropathy indices were used to predict transmembrane domains and surface-exposed regions on these proteins. Of the three sequences submitted to ProtParam for this analysis, all demonstrated negative values indicative of high hydrophilicity and potential B-cell antigenicity, at least for immunoglobulin binding (Table 4.2, right hand column). The *G. p. palpalis* EP repeat had a hydrophilicity value of -1.253, which is much higher than that of the *G. m. morsitans* protein with a value of –1.006. This is due to the additional 54 polar residues in the EP-rich region of the *G. p. palpalis* protein.

Table 4.2. ***Summary of the physical parameters predicted for the G. m. morsitans and G. p. palpalis EP repeat protein sequences***

Sequence	Total residue number	calculated M_r (kDa)[a]	predicted M_r (kDa)[b]	theoretical isoelectric point (pI)[a]	hydropathy index [a,c]
G. m. morsitans ORF	320	35.7	43	4.03	-1.006
G. p. palpalis ORF	374	41.9	49	3.87	-1.253
G. p. palpalis full length transcript	470	52.7	-	4.19	-1.020

[a] computed using the ExPASy tool Protein Parameter sequence prediction tool, ProtParam.

[b] estimated from the central position of the immunoreactive bands compared to molecular mass standards (Figure 4.2).

[c] the mean value of the hydrophobicity at each amino acid position. It is based on the Kyte and Doolittle (1982) hydrophobicity scale where computed values range from 4.5 (hydrophobic) to -4.5 (hydrophilic).

Following the computer analysis of the EP proteins described above, their predicted chemical properties were examined using another series of online tools. Figure 4.10 shows a composite diagrammatic representation of the structure and the potential signature sequences predicted for the *G. p. palpalis* EP repeat protein ORF. The TMHMM program (Sonnhammer *et al.*, 1998; Krogh *et al.*, 2001) was used to locate transmembrane helices in the protein. Only one predicted transmembrane region, spanning 18 residues, was found in the N-terminal region and its location and size is indicative of a signal peptide. To further substantiate this possibility, the SignalP tool (Nielsen *et al.*, 1997; Nielsen *et al.*, 1999) was used to measure the likelihood and location of signal peptide cleavage sites within the EP protein sequence. The *G. p. palpalis* EP ORF was predicted to encode an N-terminal signal peptide (Figure 4.10, red box) with a cleavage site (▾) at VAA-LS, 19 residues downstream from the methionine

residue. This suggests that the tsetse EP protein is likely to be extracellular (green line, Figure 4.10) and is probably secreted into the hemolymph.

ScanProsite, which searches the annotated ProSite database (Hofmann *et al.*, 1999), was used to identify other potential protein motifs. No N-linked glycosylation sites (*) were predicted for the EP repeat protein. Other modifications, such as 11 phosphorylation and 4 myristoylation sites, were predicted but were not included in the figure in order to maintain clarity. The potential O-linked glycosylation sites were predicted by the NetOGlyc algorithm (Hansen *et al.*, 1998), based on the O-GLYCBASE vs. 5.0 database containing experimentally confirmed O-glycosylated proteins (Gupta *et al.*, 1999). O-glycosylation involves the addition of a GalNAc-residue to the hydroxyl group of a serine or threonine. As these glycans are commonly found clustered in Ser/Thr-rich domains and may be associated with protease resistance (Van den Steen *et al.*, 1998), the EP repeat proteins were examined for potential O-glycosylation sites. Despite the presence of 11 threonine and 18 serine residues in the open reading frame of the *G. p. palpalis* EP repeat protein (7.8 % of total protein composition), the sequence scored below the calculated threshold levels (score potential < threshold indicates low probability of O-glycosylation). However, as this database is established on the analysis of mammalian proteins, arthropod proteins may give less definitive scores. In light of this, the residue Thr137, which scored just under the threshold, requires further analysis to support O-glycosylation at this site.

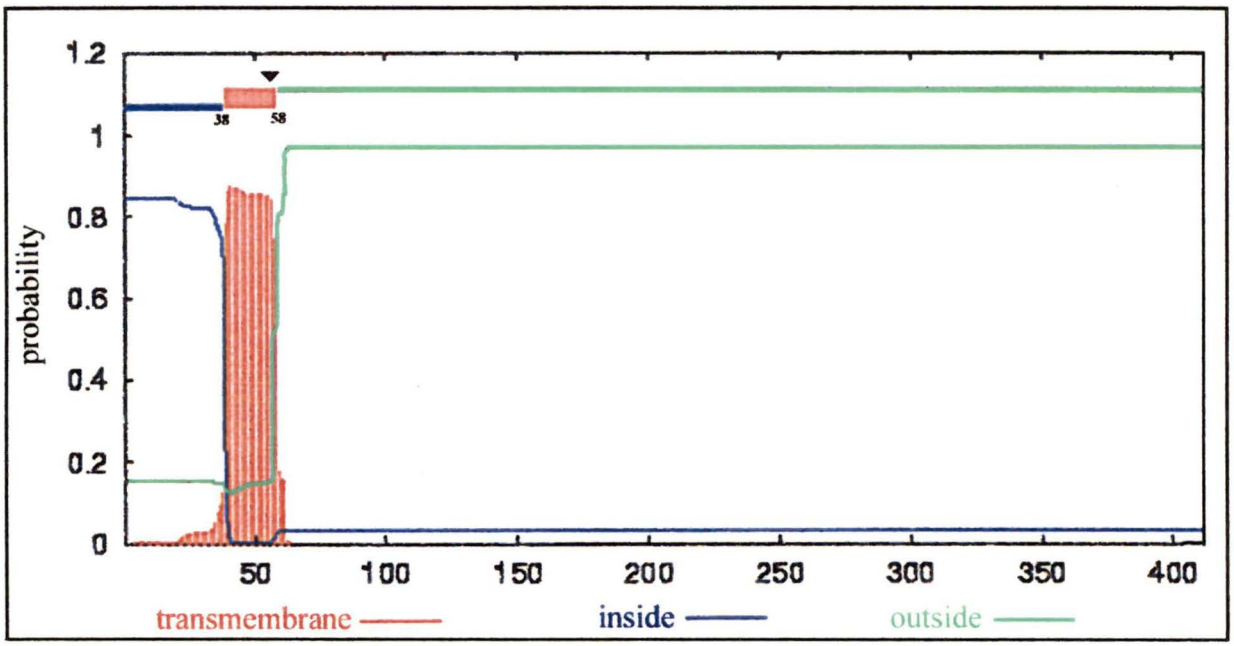


Figure 4.10. Analysis of the G. p. palpalis EP repeat protein sequence using multiple algorithms for prediction of protein motifs.

The probability of transmembrane helices (y-axis) is plotted against the protein sequence (amino acid residue number, x-axis). The prediction program, TMHMM, calculated a single transmembrane region spanning residues 39 to 57. Using the SignalP algorithm, a cleavage site (▾) was predicted at residue 55, suggesting that the hydrophobic peptide is a leader sequence.

The *G. p. palpalis* EP protein sequence was further analysed for its antigenic potential using the Protean™ protein analysis software component of DNASTAR Inc., (Madison, WI, USA). A combination of four algorithms was used to predict potential B-cell and T-cell epitopes. Potential B-cell epitopes were examined first. The Hopp-Woods algorithm (Hopp and Woods, 1983) predicts regional hydrophilicity by assuming that antigenic sites contain regions of charged side chains that interact with an aqueous environment. However, the Jameson-Wolf algorithm is the most informative as it draws a composite plot using information from five algorithms that assess regional hydrophilicity (Hopp and Woods, 1983), surface probability (Emini *et al.*, 1985), side-chain rigidity (Karplus and Schulz, 1987), and secondary protein structure (Chou and

Fasman, 1978; Garnier *et al.*, 1996). The results of B-cell epitope analysis are shown in the upper two plots in Figure 4.11. The highlighted region (blue) represents the hydrophilic, surface-disposed EP repeat region that is probably antigenic and could contain multiple B-cell epitopes. The Jameson-Wolf plot also shows several other potential B-cell epitopes dispersed along the entire protein sequence.

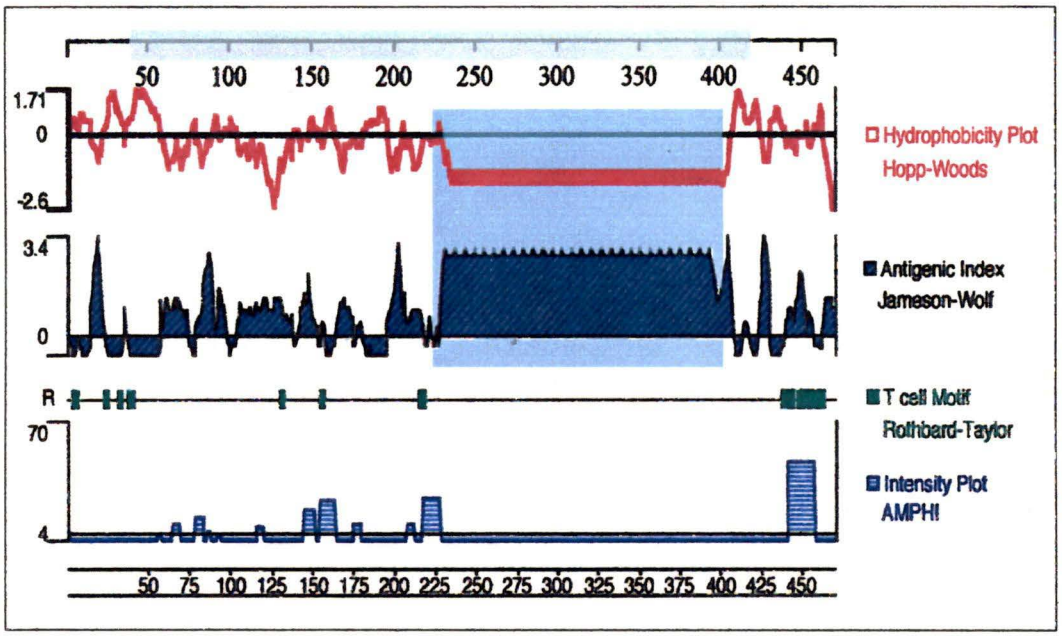


Figure 4.11. Antigenic analysis of the <u>G. p. palpalis</u> EP repeat protein sequence. *The Hopp-Woods plot (red) shows hydrophobic regions. Surface-exposed antigenic regions are predicted by the Jameson-Wolf (black) program. The Rothbard-Taylor (green) and the AMPHI (blue) algorithms predict T-cell epitopes. The open reading frame region is indicated in gray and the light blue box highlights the hydrophilic EP repeat region that is a predicted B-cell epitope. Top and bottom scales represent the residue position.*

The Rothbard-Taylor algorithm detects T-cell epitopes by searching a sequence for residue periodicity common to small T-cell motifs (Rothbard and Taylor, 1988). The AMPHI algorithm preferentially scans the sequence for amphipathic helices, a structural pattern of MHC binding motifs commonly observed at helper T-cell antigenic sites. The

C-terminus of the *G. p. palpalis* EP repeat protein showed a potential T-cell motif. This was however, not in the open reading frame of the encoded protein.

4. Discussion

As previously stated in chapter 2, few molecules within the tsetse midgut have been biochemically characterized (p.16). Since an increased incidence of trypanosome infection is related to the age of the fly (the teneral susceptibility phenomenon), examination of molecules unique to teneral flies was the initial focus of my research. The use of teneral flies also eliminates potential bloodmeal contamination, thus simplifying analysis. I first screened midgut lysates from the riverine species, *G. p. palpalis* and the savannah species *G. m. morsitans* by immunoblotting with monoclonal antibody 247. This was done because of a serendipitous finding in an immunogold electron microscopic study (Tetley, L., personal communication) that an anti-EP repeat mAb (247) bound to structures in teneral flies. MAb 247 is a species-specific and stage-specific antibody that recognizes EP repeat structures of procyclin molecules. In my experiments, the mAb recognized a broad band on immunoblots of midgut lysates and there was size discrepancy of approximately 6 kDa between the immunoreactive bands of *G. p. palpalis* and *G. m. morsitans*. It is unlikely that these EP repeat-containing proteins are from the peritrophic matrix since the mild solubilization procedures utilized to prepare the midgut lysates were not harsh enough to extract matrix proteins (East *et al.*, 1993). The apparent difference in molecular masses could have been caused by either polypeptide variations or differences in post-translational modifications. Therefore, to determine the primary structure of the EP-containing proteins, I cloned and sequenced the encoding genes from both *G. p. palpalis* and *G. m. morsitans*. The identification of the relevant gene fragments from cDNA libraries made from fed and teneral flies confirmed that the expression of this protein was independent

of the ingestion of a bloodmeal. This is unique as a majority of the midgut molecules previously characterized are expressed only in response to tsetse feeding (Cappello *et al.*, 1998; Yan *et al.*, 2001a). By performing additional PCR amplifications, the entire open reading frame of the EP genes and both upstream and downstream noncoding regions were identified and sequenced. The longer *G. p. palpalis* EP repeat protein was analysed using computer prediction algorithms and characterization programs to investigate hypothetical structures and functions. First, a putative hydrophobic, N-terminal signal sequence was predicted. Second, the C-terminal domain was predicted to be highly hydrophilic, and thus almost certainly confers solubility to the protein since it is rich in polar residues. It is interesting that all 8 of the cysteine residues (comprising 2.1% of the total amino acid composition) are in the N-terminal region of the open reading frame, suggesting that this domain may be highly folded.

Recently, analysis of midgut microvillar molecules of the mosquito revealed 28 glycoproteins, eight of which expressed O-linked oligosaccharides. While the role of these glycoproteins has not been determined, oligosaccharide side chains have been proposed to be parasite receptors (Wilkins and Billingsley, 2001). In the tsetse EP gene sequences reported in my work, the only two N-terminal glycosylation consensus sequences were outside of the open reading frame, thus the mature protein is not likely N-glycosylated. However, post-translational addition of O-linked glycan chains to the hydroxyl groups of specific amino acids can produce multivalent, typically immunogenic, glycoproteins (Hounsell *et al.*, 1996). Such glycans are commonly found in clusters and their presence physically confers stability, by adding increased protease resistance and rigidity to the protein (Van den Steen *et al.*, 1998). Considering that O-linked

oligosaccharides have previously been characterized in the tsetse midgut (Lehane *et al.*, 1996) and the *G. p. palpalis* EP protein is Ser/Thr-rich, inflexible and protease resistant, the probability of O-glycosylation appears to be high. Despite the large number of Ser/Thr residues, the NetOglyc prediction program did not identify any potential O-glycosylation residues with high scores. However one residue, Thr137, showed a weak potential which was below the threshold value determined by the algorithm. However, it is possible that relying on mammalian proteins for the design of prediction algorithms may be misleading.

The use of computer-driven algorithms to search for protein sequences likely to induce both cellular and humoral immune responses significantly aids the selection of molecules, and fragments of those molecules, that may be suitable as vaccine targets. T-cell antigenic peptides contain amino acid sequences that allow the peptide to bind in the MHC class I and II grooves, with a hydrophilic side interacting with the T cell receptor and a hydrophobic side facing the MHC molecule. Using various prediction algorithms, such as AMPHI and Rothbard-Taylor algorithms, a preliminary evaluation of the T-cell antigenicity of an amino acid sequence can be conducted. No prominent T-cell motifs were predicted for the EP repeat protein. On the other hand, the Jameson-Wolf antigenic index which is predictive for B-cell epitopes, identified the EP region to be highly antigenic.

It is known that both the species of tsetse and the trypanosome strains have an influence on the efficiency of parasite transmission (Moloo *et al.*, 1998). In certain instances, possibly influenced by environmental signals, trypanosome development is confined to the midgut. Is it just coincidence that the teneral midgut expresses the EP

repeat protein and the first trypanosome surface marker to appear and to persist during procyclic differentiation and midgut establishment are the EP procyclins (Vassella *et al.*, 2001)? Does this relationship imply a role for the EP proteins, for example in parasite-vector signaling? Are the EP repeat sequences simply serving to extend the polypeptides away from cell membranes and thus play more of a physical than physiological role? Only *in vivo* experiments will be able to resolve some of these questions and perhaps the answers will lead to a new vector-based approach to the control African sleeping sickness.

General Discussion

The original objective of my thesis research was to isolate, identify and characterize molecules within the midgut of the tsetse fly. My research led to the identification of several prokaryote molecules from both the primary and secondary midgut bacterial symbionts and to the identification of one tsetse protein.

The first molecule identified and characterized was a 60 kDa molecular chaperone, the most predominant soluble protein in the tsetse midgut. It was localized to the tsetse bacteriome and was synthesized by the obligate bacterial endosymbiont, *Wigglesworthia glossinidia.* Why this symbiont produces such enormous quantities of a chaperonin and how the bacterium or the tsetse utilize this protein, if at all, remains a mystery.

Since the midgut bacterial symbiont *Sodalis glossinidius* was previously shown to be linked to the susceptibility of tsetse to trypanosome infections, I set out to grow this microbe, to clone it and to identify molecules that may aid in understanding parasite-tsetse interactions. It was again surprising that a major protein identified was a 60 kDa chaperone. This protein was also released, in large quantities, into the culture supernatant. Again, the role of this chaperone in tsetse is unknown.

Finally I identified an EP repeat-containing protein from midguts of adult teneral *G. m. morsitans* and *G. p. palpalis*. The EP repeat proteins were predicted to be soluble, species-specific and immunogenic molecules. However, a role for these molecules within the midgut has not been defined.

Research presented in this thesis has thus revealed several predominant molecules, both insect and bacterial, that are found within the midgut of the tsetse.

Although the functional relevance of these proteins was not resolved, they deserve further study as they may play central roles in tsetse viability or trypanosome transmission. More specifically, it is possible that *Sodalis* alters the tsetse midgut into a more hospitable environment for parasite establishment and may thus be critical for parasite transmission. The monoclonal antibodies produced to the *Sodalis* chaperone and to the outer membrane porin will be useful tools that allow more precise quantitation and localization of *Sodalis* within tsetse tissues than was possible in the past. In addition, membrane feeding experiments using anti-symbiont and tsetse-protein specific monoclonal antibodies and their target proteins may aid in understanding the molecular interactions between vector and parasite that may eventually lead to advancements in the control of African trypanosomiasis.

References cited

Acosta-Serrano, A., Cole, R.N. and Englund, P.T. (2000) Killing of *Trypanosoma brucei* by concanavalin A: structural basis of resistance in glycosylation mutants. *Journal of Molecular Biology.* 304: 633-644.

Acosta-Serrano, A., Vassella, E., Liniger, M., Kunz Renggli, C., Brun, R., Roditi, I. and Englund, P.T. (2001) The surface coat of procyclic *Trypanosoma brucei*: programmed expression and proteolytic cleavage of procyclin in the tsetse fly. *Proceedings of the National Academy of Sciences of the United States of America.* 98: 1513-1518.

Aebersold, R. and Goodlett, D.R. (2001) Mass spectrometry in proteomics. *Chemical Reviews.* 101: 269-295.

Akman, L. and Aksoy, S. (2001) A novel application of gene arrays: *Escherichia coli* array provides insight into the biology of the obligate endosymbiont of tsetse flies. *Proceedings of the National Academy of Sciences of the United States of America.* 98: 7546-7551.

Akman, L., Rio, R.V.M., Beard, C.B. and Aksoy, S. (2001) Genome size determination and coding capacity of *Sodalis glossinidius*, an enteric symbiont of tsetse flies, as revealed by hybridization of *Escherichia coli* gene arrays. *Journal of Bacteriology.* 183: 4517-4525.

Aksoy, S. (1995a) Molecular analysis of the endosymbionts of tsetse flies: 16S rDNA locus and over-expression of a chaperonin. *Insect Molecular Biology.* 4: 23-29.

Aksoy, S. (1995b) *Wigglesworthia gen. nov.* and *Wigglesworthia glossinidia sp. nov.*, taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. *International Journal of Systematic Bacteriology.* 45: 848-851.

Aksoy, S. (2000) Tsetse--A haven for microorganisms. *Parasitology Today.* 16: 114-118.

Aksoy, S., Chen, X. and Hypsa, V. (1997) Phylogeny and potential transmission routes of midgut-associated endosymbionts of tsetse (Diptera:Glossinidae). *Insect Molecular Biology.* 6: 183-190.

Aksoy, S., Pourhosseini, A.A. and Chow, A. (1995) Mycetome endosymbionts of tsetse flies constitute a distinct lineage related to Enterobacteriaceae. *Insect Molecular Biology.* 4: 15-22.

Allsopp, R. (2001) Options for vector control against trypanosomiasis in Africa. *Trends in Parasitology.* 17: 15-19.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research.* 25: 3389-3402.

Anderson, L. and Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis.* 18: 533-537.

Anderson, N.G. and Anderson, N.L. (1978a) Analytical techniques for cell fractions. XXI. Two-dimensional analysis of serum and tissue proteins: multiple isoelectric focusing. *Analytical Biochemistry.* 85: 331-340.

Anderson, N.L. and Anderson, N.G. (1978b) Analytical techniques for cell fractions. XXII. Two-dimensional analysis of serum and tissue proteins: multiple gradient-slab gel electrophoresis. *Analytical Biochemistry.* 85: 341-354.

Anderson, N.L., Parish, N.M., Richardson, J.P. and Pearson, T.W. (1985) Comparison of African trypanosomes of different antigenic phenotypes, subspecies and life cycle stages by two-dimensional gel electrophoresis. *Molecular and Biochemical Parasitology.* 16: 299-314.

Baker, R.D., Maudlin, I., Milligan, P.J., Molyneux, D.H. and Welburn, S.C. (1990) The possible role of Rickettsia-like organisms in trypanosomiasis epidemiology. *Parasitology.* 100 Pt 2: 209-217.

Barra, D., Simmaco, M. and Boman, H.G. (1998) Gene-encoded peptide antibiotics and innate immunity. Do 'animalcules' have defence budgets? *FEBS Letters.* 430: 130-134.

Barrett, M.P. (1999) The fall and rise of sleeping sickness. *Lancet.* 353: 1113-1114.

Beecroft, R.P., Roditi, I. and Pearson, T.W. (1993) Identification and characterization of an acidic major surface glycoprotein from procyclic stage *Trypanosoma congolense. Molecular and Biochemical Parasitology.* 61: 285-294.

Bondi, M., Messi, P., Sabia, C., Baccarani Contri, M. and Manicardi, G. (1999) Antimicrobial properties and morphological characteristics of two *Photorhabdus luminescens* strains. *New Microbiologica.* 22: 117-127.

Bruce, D. (1915) The Croonian lectures on trypanosomes. *Lancet.* 2: 55-63.

Bruce, D., Hamerton, A.E., Bateman, H.R. and Mackie, F.P. (1909) The development of *Trypanosoma gambiense* in *Glossina palpalis. Proceedings of the Royal Society of London.* B81: 405-414.

Budd, L.T. (1999) DFID-Funded Tsetse and Trypanosomiasis Research and Development Since 1980 (Economic Analysis). Department for International Development

Cappello, M., Li, S., Chen, X., Li, C.B., Harrison, L., Narashimhan, S., Beard, C.B. and Aksoy, S. (1998) Tsetse thrombin inhibitor: bloodmeal-induced expression of an anticoagulant in salivary glands and gut tissue of *Glossina morsitans morsitans. Proceedings of the National Academy of Sciences of the United States of America.* 95: 14290-14295.

Cattand, P. (1987) Advances in the field diagnosis of human African trypanosomiasis. *La Medicina Tropicale nella Cooperazione allo Sviluppo.* 3: 105-108.

Charles, H., Heddi, A., Guillaud, J., Nardon, C. and Nardon, P. (1997) A molecular aspect of symbiotic interactions between the weevil *Sitophilus oryzae* and its endosymbiotic bacteria: over-expression of a chaperonin. *Biochemical and Biophysical Research and Communications.* 239: 769-774.

Cheeseman, M. and Gooding, R. (1985) Proteolytic enzymes from tsetse flies, *Glossina morsitans* and *Glossina palpalis* (Diptera: Glossinidae). *Insect Biochemistry.* 15: 677-680.

Cheng, Q. and Aksoy, S. (1999) Tissue tropism, transmission and expression of foreign genes *in vivo* in midgut symbionts of tsetse flies. *Insect Molecular Biology.* 8: 125-132.

Chou, P.Y. and Fasman, G.D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in Enzymology and Related Areas of Molecular Biology.* 47: 45-148.

Dale, C. and Maudlin, I. (1999) *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans. International Journal of Systematic Bacteriology.* 49 Pt 1: 267-275.

Dale, C. and Welburn, S.C. (2001) The endosymbionts of tsetse flies: manipulating host-parasite interactions. *International Journal for Parasitology.* 31: 627-630.

Dale, C., Young, S.A., Haydon, D.T. and Welburn, S.C. (2001) The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. *Proceedings of the National Academy of Sciences of the United States of America.* 98: 1883-1888.

Dean, G.J., Wilson, F. and Wortham, S. (1968) Some factors affecting eclosion of *Glossina morsitans* Westw. from pupae. *Bulletin of Entomological Research.* 57: 367-377.

Denlinger, D.L. and Ma, W.C. (1974) Dynamics of the pregnancy cycle in the tsetse *Glossina morsitans. Journal of Insect Physiology.* 20: 1015-1026.

Deziel, E., Comeau, Y. and Villemur, R. (2001) Initiation of biofilm formation by *Pseudomonas aeruginosa* 57RP correlates with emergence of hyperpiliated and highly adherent phenotypic variants deficient in swimming, swarming, and twitching motilities. *Journal of Bacteriology.* 183: 1195-1204.

Doran, J.L., Collinson, S.K., Kay, C.M., Banser, P.A., Burian, J., Munro, C.K., Lee, S.H., Somers, J.M., Todd, E.C. and Kay, W.W. (1994) fimA and tctC based DNA diagnostics for *Salmonella. Molecular and Cellular Probes.* 8: 291-310.

Dunavan, C.P. (2000) Vital signs. *Discover.* 21: 38-41.

East, I.J., Fitzgerald, C.J., Pearson, R.D., Donaldson, R.A., Vuocolo, T., Cadogan, L.C., Tellam, R.L. and Eisemann, C.H. (1993) *Lucilia cuprina*: inhibition of larval growth induced by immunization of host sheep with extracts of larval peritrophic membrane. *International Journal for Parasitology.* 23: 221-229.

Emini, E.A., Hughes, J.V., Perlow, D.S. and Boger, J. (1985) Induction of hepatitis A virus-neutralizing antibody by a virus- specific synthetic peptide. *Journal of Virology.* 55: 836-839.

Endege, W.O., Lonsdale-Eccles, J.D., Olembo, N.K., Moloo, S.K. and ole-MoiYoi, O.K. (1989) Purification and characterization of two fibrinolysins from the midgut of adult female *Glossina morsitans centralis. Comparative Biochemistry and Physiology B- Biochemistry and Molecular Biology.* 92: 25-34.

Ensgraber, M. and Loos, M. (1992) A 66-kilodalton heat shock protein of *Salmonella typhimurium* is responsible for binding of the bacterium to intestinal mucus. *Infection and Immunity.* 60: 3072-3078.

Evans, D.A. and Ellis, D.S. (1983) Recent observations on the behaviour of certain trypanosomes within their insect hosts. *Advances in Parasitology.* 22: 1-42.

Evans, D.J., Jr., Evans, D.G., Engstrand, L. and Graham, D.Y. (1992) Urease-associated heat shock protein of *Helicobacter pylori. Infection and Immunity.* 60: 2125-2127.

Fayet, O., Ziegelhoffer, T. and Georgopoulos, C. (1989) The groES and groEL heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *Journal of Bacteriology.* 171: 1379-1385.

Feutrier, J., Kay, W.W. and Trust, T.J. (1986) Purification and characterization of fimbriae from *Salmonella enteritidis. Journal of Bacteriology.* 168: 221-227.

Filichkin, S.A., Brumfield, S., Filichkin, T.P. and Young, M.J. (1997) *In vitro* interactions of the aphid endosymbiotic SymL chaperonin with barley yellow dwarf virus. *Journal of Virology.* 71: 569-577.

Frisk, A., Ison, C.A. and Lagergard, T. (1998) GroEL heat shock protein of *Haemophilus ducreyi:* association with cell surface and capacity to bind to eukaryotic cells. *Infection and Immunity.* 66: 1252-1257.

Garduno, R.A., Garduno, E. and Hoffman, P.S. (1998) Surface-associated hsp60 chaperonin of *Legionella pneumophila* mediates invasion in a HeLa cell model. *Infection and Immunity.* 66: 4602-4610.

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

Gooding, R.H. and Jordan, A.M. (1986) Genetics of *Glossina morsitans morsitans* (Diptera: Glossinidae). XII. Comparison of field-collected and laboratory-reared flies. *Canadian Journal of Genetics and Cytology.* 28: 1016-1021.

Gupta, R., Birch, H., Rapacki, K., Brunak, S. and Hansen, J.E. (1999) O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins. *Nucleic Acids Research.* 27: 370-372.

Hale, J.E., Butler, J.P., Knierman, M.D. and Becker, G.W. (2000) Increased sensitivity of tryptic peptide detection by MALDI-TOF mass spectrometry is achieved by conversion of lysine to homoarginine. *Analytical Biochemistry.* 287: 110-117.

Hansen, J.E., Lund, O., Tolstrup, N., Gooley, A.A., Williams, K.L. and Brunak, S. (1998) NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconjugate Journal.* 15: 115-130.

Hao, Z., Kasumba, I., Lehane, M.J., Gibson, W.C., Kwon, J. and Aksoy, S. (2001) Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis. *Proceedings of the National Academy of Sciences of the United States of America.* 98: 12648-12653.

Henderson, I.R., Owen, P. and Nataro, J.P. (1999) Molecular switches--the ON and OFF of bacterial phase variation. *Molecular Microbiology.* 33: 919-932.

Hennequin, C., Porcheray, F., Waligora-Dupriet, A., Collignon, A., Barc, M., Bourlioux, P. and Karjalainen, T. (2001) GroEL (Hsp60) of *Clostridium difficile* is involved in cell adherence. *Microbiology.* 147: 87-96.

Hoffman, R. (1954) Zur Fortpflanzungsbiologie und zur intrauterinen Entwicklung von *Glossina palpalis*. *Acta Tropica.* 11: 1-57.

Hofmann, K., Bucher, P., Falquet, L. and Bairoch, A. (1999) The PROSITE database, its status in 1999. *Nucleic Acids Research.* 27: 215-219.

Hopp, T.P. and Woods, K.R. (1983) A computer program for predicting protein antigenic determinants. *Molecular Immunology.* 20: 483-489.

Hounsell, E.F., Davies, M.J. and Renouf, D.V. (1996) O-linked protein glycosylation structure and function. *Glycoconjugate Journal.* 13: 19-26.

Hursey, B.S. (2001) The programme against African trypanosomiasis: aims, objectives and achievements. *Trends in Parasitology.* 17: 2-3.

Hurst, L.D. and Randerson, J.P. (2002) Parasitic sex puppeteers. *Scientific American.* 286: 56-61.

Imbuga, M.O., Osir, E.O., Labongo, V.L., Darji, N. and Otieno, L.H. (1992) Studies on tsetse midgut factors that induce differentiation of blood- stream *Trypanosoma brucei brucei in vitro. Parasitology Research.* 78: 10-15.

Karplus, P.A. and Schulz, G.E. (1987) Refined structure of glutathione reductase at 1.54 A resolution. *Journal of Molecular Biology.* 195: 701-729.

Kim, W.S., Park, J.H., Ren, J., Su, P. and Dunn, N.W. (2001) Survival Response and Rearrangement of Plasmid DNA of *Lactococcus lactis* during Long-Term Starvation. *Applied and Environmental Microbiology.* 67: 4594-4602.

Kinter, M. and Sherman, N.E. (2000) *Protein sequencing and identification using tandem mass spectrometry.* New York: John Wiley & Sons, Inc., pp. 147-165.

Koch, R. (1904) Remarks on trypanosome diseases. *British Medical Journal.* 2: 1445-1449.

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology.* 305: 567-580.

Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology.* 157: 105-132.

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

Lal, A.A., Patterson, P.S., Sacci, J.B., Vaughan, J.A., Paul, C., Collins, W.E., Wirtz, R.A. and Azad, A.F. (2001) Anti-mosquito midgut antibodies block development of *Plasmodium falciparum* and *Plasmodium vivax* in multiple species of *Anopheles* mosquitoes and reduce vector fecundity and survivorship. *Proceedings of the National Academy of Sciences of the United States of America.* 98: 5228-5233.

Leak, S.A.G. (1999) *Tsetse biology and ecology: their role in the epidemiology and control of trypanosomiasis.* New York: CABI Publishing, pp. 568.

Lehane, M.J., Allingham, P.G. and Weglicki, P. (1996) Composition of the peritrophic matrix of the tsetse fly, *Glossina morsitans morsitans. Cell and Tissue Research.* 283: 375-384.

Lehane, M.J. and Msangi, A.R. (1991) 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. *Medical and Veterinary Entomology.* 5: 495-501.

Lehane, M.J., Wu, D. and Lehane, S.M. (1997) Midgut-specific immune molecules are produced by the blood-sucking insect *Stomoxys calcitrans. Proceedings of the National Academy of Sciences of the United States of America.* 94: 11502-11507.

Li, Z., Clarke, A.J. and Beveridge, T.J. (1998) Gram-negative bacteria produce membrane vesicles which are capable of killing other bacteria. *Journal of Bacteriology.* 180: 5478-83.

Liang, J.F. and Kim, S.C. (1999) Not only the nature of peptide but also the characteristics of cell membrane determine the antimicrobial mechanism of a peptide. *Journal of Peptide Research.* 53: 518-522.

Lockey, T.D. and Ourth, D.D. (1996) Formation of pores in *Escherichia coli* cell membranes by a cecropin isolated from hemolymph of *Heliothis virescens* larvae. *European Journal of Biochemistry.* 236: 263-271.

Maloy, S.R. (1990) *Experimental techniques in bacterial genetics.* Boston: Jones and Bartlett Publishers, pp. 161-164.

Maudlin, I. (1991) Transmission of African trypanosomiasis: interactions among tsetse immune system, symbionts and parasites. *Advances in Disease Vector Research.* 7: 117-148.

Maudlin, I. and Ellis, D.S. (1985) Association between intracellular rickettsial-like infections of midgut cells and susceptibility to trypanosome infection in *Glossina* spp. *Zeitschrift fur Parasitenkunde.* 71: 683-687.

Maudlin, I. and Welburn, S.C. (2001) Genome - which genome? *Trends in Parasitology.* 17: 50-52.

Maudlin, I., Welburn, S.C. and Mehlitz, D. (1990) The relationship between rickettsia-like-organisms and trypanosome infections in natural populations of tsetse in Liberia. *Tropical Medicine and Parasitology.* 41: 265-267.

Merril, C.R., Goldman, D., Sedman, S.A. and Ebert, M.H. (1981) Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. *Science.* 211: 1437-1438.

Moller, S., Croning, M.D. and Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics.* 17: 646-653.

Moloo, S.K., Okumu, I.O. and Kuria, N.M. (1998) Comparative susceptibility of *Glossina longipennis* and *G. brevipalpis* to pathogenic species of *Trypanosoma. Medical and Veterinary Entomology.* 12: 211-214.

Moloo, S.K. and Shaw, M.K. (1989) Rickettsial infections of midgut cells are not associated with susceptibility of *Glossina morsitans centralis* to *Trypanosoma congolense* infection. *Acta Tropica.* 46: 223-227.

Moore, A., Richer, M., Enrile, M., Losio, E., Roberts, J. and Levy, D. (1999) Resurgence of sleeping sickness in Tambura County, Sudan. *American Journal of Tropical Medicine and Hygiene.* 61: 315-318.

Moore, D.A., Edwards, M., Escombe, R., Agranoff, D., Bailey, J.W., Squire, S.B. and Chiodini, P.L. (2002) African trypanosomiasis in travelers returning to the United kingdom. *Emerging Infectious Diseases.* 8: 74-76.

Morin, S., Ghanim, M., Sobol, I. and Czosnek, H. (2000) The GroEL protein of the whitefly *Bemisia tabaci* interacts with the coat protein of transmissible and nontransmissible begomoviruses in the yeast two-hybrid system. *Virology.* 276: 404-416.

Morin, S., Ghanim, M., Zeidan, M., Czosnek, H., Verbeek, M. and van den Heuvel, J.F. (1999) A GroEL homologue from endosymbiotic bacteria of the whitefly *Bemisia tabaci* is implicated in the circulative transmission of tomato yellow leaf curl virus. *Virology.* 256: 75-84.

Neuhoff, V., Arold, N., Taube, D. and Ehrhardt, W. (1988) 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.* 9: 255-262.

Nguu, E.K., Osir, E.O., Imbuga, M.O. and Olembo, N.K. (1996) The effect of host blood in the *in vitro* transformation of bloodstream trypanosomes by tsetse midgut homogenates. *Medical and Veterinary Entomology.* 10: 317-322.

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

Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering.* 10: 1-6.

Nogge, G. (1981) Significance of symbionts for the maintenance of an optimal nutritional state for successful reproduction in hematophagous arthropods. *Parasitology.* 82: 299-304.

Okedi, L.M.A. (1995) Chapter 10: Least specific sucker of vertebrate blood. In *University of Florida Book of Insect Records (2001).* http://recbk.ifas.ufl.edu.

O'Neill, S.L., Gooding, R.H. and Aksoy, S. (1993) Phylogenetically distant symbiotic microorganisms reside in *Glossina* midgut and ovary tissues. *Medical and Veterinary Entomology.* 7: 377-383.

Otieno, L.H., Darji, N., Onyango, P. and Mpanga, E. (1983) Some observations on factors associated with the development of *Trypanosoma brucei brucei* infections in *Glossina morsitans morsitans. Acta Tropica.* 40: 113-120.

Otvos, L., Jr., O, I., Rogers, M.E., Consolvo, P.J., Condie, B.A., Lovas, S., Bulet, P. and Blaszczyk-Thurin, M. (2000) Interaction between heat shock proteins and antimicrobial peptides. *Biochemistry.* 39: 14150-14159.

Pays, E. and Nolan, D.P. (1998) Expression and function of surface proteins in *Trypanosoma brucei. Molecular and Biochemical Parasitology.* 91: 3-36.

Pearson, T. and Anderson, L. (1980) Analytical techniques for cell fractions. XXVIII. Dissection of complex antigenic mixtures using monoclonal antibodies and two-dimensional gel electrophoresis. *Analytical Biochemistry.* 101: 377-386.

Pearson, T.W. (2001) Procyclins, proteases and proteomics: dissecting trypanosomes in the tsetse fly. *Trends in Microbiology.* 9: 299-301.

Pelle, R. and Murphy, N.B. (1993) Northern hybridization: rapid and simple electrophoretic conditions. *Nucleic Acids Research.* 21: 2783-2784.

Pinnock, D.E. and Hess, R.T. (1974) The occurrence of intracellular rickettsia-like organisms in the tsetse flies, *Glossina morsitans, G. fuscipes, G. brevipalpis and G. pallidipes. Acta Tropica.* 31: 70-79.

Ranford, J.C., Coates, A.R.M. and Henderson, B. (2000) Chaperonins are cell-signalling proteins: the unfolding biology of molecular chaperones. *Expert Reviews in Molecular Medicine.* http://www-ermm.cbcu.cam.ac.uk/00002015h.htm: 1-17.

Reinhardt, C., Steiger, R. and Hecker, H. (1972) Ultrastructural study of the midgut mycetome-bacteroids of the tsetse flies *Glossina morsitans, G. fuscipes*, and *G. brevipalpis* (Diptera, Brachycera). *Acta Tropica.* 29: 280-288.

Richardson, J.P., Beecroft, R.P., Tolson, D.L., Liu, M.K. and Pearson, T.W. (1988) Procyclin: an unusual immunodominant glycoprotein surface antigen from the procyclic stage of African trypanosomes. *Molecular and Biochemical Parasitology.* 31: 203-216.

Richardson, J.P., Jenni, L., Beecroft, R.P. and Pearson, T.W. (1986) Procyclic tsetse fly midgut forms and culture forms of African trypanosomes share stage- and species-specific surface antigens identified by monoclonal antibodies. *Journal of Immunology.* 136: 2259-2264.

Roditi, I. and Pearson, T.W. (1990) The procyclin coat of African trypanosomes. *Parasitology Today.* 6: 79-81.

Ross, R. and Thomson, D. (1910) A case of sleeping sickness studied by precise enumerative methods: regular periodical increase of the parasites disclosed. *Annals of Tropical Medicine and Parasitology.* 4: 261-265.

Rothbard, J.B. and Taylor, W.R. (1988) A sequence pattern common to T cell epitopes. *The European Molecular Biology Laboratory Journal.* 7: 93-100.

Ruepp, S., Furger, A., Kurath, U., Renggli, C.K., Hemphill, A., Brun, R. and Roditi, I. (1997) Survival of *Trypanosoma brucei* in the tsetse fly is enhanced by the expression of specific forms of procyclin. *Journal of Cell Biology.* 137: 1369-1379.

Saunders, D.S. (1962) Age and determination for female tsetse flies and the age compositions of samples of *Glossina pallidipes* Aust., *G. palpalis fuscipes* Newst., and *G. brevipalpis* Newst. *Bulletin of Entomological Research.* 53: 579-595.

Sbicego, S., Vassella, E., Kurath, U., Blum, B. and Roditi, I. (1999) The use of transgenic *Trypanosoma brucei* to identify compounds inducing the differentiation of bloodstream forms to procyclic forms. *Molecular and Biochemical Parasitology.* 104: 311-322.

Smith, D.H., Pepin, J. and Stich, A.H. (1998) Human African trypanosomiasis: an emerging public health crisis. *British Medical Bulletin.* 54: 341-355.

Sonnhammer, E.L., von Heijne, G. and Krogh, A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings of the International Conference on Intelligent Systems for Molecular Biology* . 6: 175-182.

Tetley, L., Turner, C.M., Barry, J.D., Crowe, J.S. and Vickerman, K. (1987) Onset of expression of the variant surface glycoproteins of Trypanosoma brucei in the tsetse fly studied using immunoelectron microscopy. *Journal of Cell Science.* 87: 363-372.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research.* 22: 4673-4680.

Tolson, D.L., Turco, S.J., Beecroft, R.P. and Pearson, T.W. (1989) The immunochemical structure and surface arrangement of *Leishmania donovani* lipophosphoglycan determined using monoclonal antibodies. *Molecular and Biochemical Parasitology.* 35: 109-118.

Uttenweiler-Joseph, S., Moniatte, M., Lagueux, M., Van Dorsselaer, A., Hoffmann, J.A. and Bulet, P. (1998) Differential display of peptides induced during the immune response of *Drosophila*: a matrix-assisted laser desorption ionization time-of- flight mass spectrometry study. *Proceedings of the National Academy of Sciences of the United States of America.* 95: 11342-11347.

Van den Steen, P., Rudd, P.M., Dwek, R.A. and Opdenakker, G. (1998) Concepts and principles of O-linked glycosylation. *Critical Reviews in Biochemistry and Molecular Biology.* 33: 151-208.

Vanet, A. and Labigne, A. (1998) Evidence for specific secretion rather than autolysis in the release of some *Helicobacter pylori* proteins. *Infection and Immunity.* 66: 1023-1027.

Vassella, E., Acosta-Serrano, A., Studer, E., Lee, S.H., Englund, P.T. and Roditi, I. (2001) Multiple procyclin isoforms are expressed differentially during the development of insect forms of *Trypanosoma brucei. Journal of Molecular Biology.* 312: 597-607.

Vickerman, K. (1969) The fine structure of *Trypanosoma congolense* in its bloodstream phase. *Journal of Protozoology.* 16: 54-69.

Vickerman, K. (1985) Developmental cycles and biology of pathogenic trypanosomes. *British Medical Bulletin.* 41: 105-114.

Vickerman, K., Myler, P.J. and Stuart, K.D. (1993). African trypanosomiasis. In *Immunology and Molecular Biology of Parasitic Infections.* Warren, K.S. (eds). Oxford, England: Blackwell Scientific Publications, pp. 170-212.

Vickerman, K., Tetley, L., Hendry, K.A. and Turner, C.M. (1988) Biology of African trypanosomes in the tsetse fly. *Biology of the Cell.* 64: 109-119.

Weiser, J.N. (1993) Relationship between colony morphology and the life cycle of *Haemophilus influenzae*: the contribution of lipopolysaccharide phase variation to pathogenesis. *Journal of Infectious Diseases.* 168: 672-680.

Weitz, B. (1963) The feeding habits of *Glossina. Bulletin of the World Health Organization.* 15: 473.

Weitz, B. and Glasgow, J.P. (1956) The natural hosts of some species of *Glossina* in East Africa. *Transactions of the Royal Soceity of Tropical Medicine and Hygiene.* 50: 593-613.

Welburn, S.C., Arnold, K., Maudlin, I. and Gooday, G.W. (1993) Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse flies. *Parasitology.* 107: 141-145.

Welburn, S.C. and Dale, C. (1997). Isolation and culture of tsetse secondary endosymbionts. In *Molecular Biology of Insect Disease Vectors: A methods manual.* Crampton, J.M., Beard, C.B. and Louis, C. (eds). Birmingham: Chapman & Hall, pp. 547-554.

Welburn, S.C. and Maudlin, I. (1987) A simple *in vitro* method for infecting tsetse with trypanosomes. *Annals of Tropical Medicine and Parasitology.* 81: 453-455.

Welburn, S.C. and Maudlin, I. (1991) Rickettsia-like organisms, puparial temperature and susceptibility to trypanosome infection in *Glossina morsitans. Parasitology.* 102 Pt 2: 201-206.

Welburn, S.C., Maudlin, I. and Ellis, D.S. (1987) *In vitro* cultivation of rickettsia-like-organisms from *Glossina* spp. *Annals of Tropical Medicine and Parasitology.* 81: 331-335.

Welburn, S.C., Maudlin, I. and Ellis, D.S. (1989) Rate of trypanosome killing by lectins in midguts of different species and strains of *Glossina. Medical and Veterinary Entomology.* 3: 77-82.

White, A.C., Jr. (2000) The disappearing arsenal of antiparasitic drugs. *New England Journal of Medicine.* 343: 1273-1274.

Wilkins, S. and Billingsley, P.F. (2001) Partial characterization of oligosaccharides expressed on midgut microvillar glycoproteins of the mosquito, *Anopheles stephensi* Liston. *Insect Biochemistry and Molecular Biology.* 31: 937-948.

Yamaguchi, H., Osaki, T., Taguchi, H., Hanawa, T., Yamamoto, T. and Kamiya, S. (1996) Flow cytometric analysis of the heat shock protein 60 expressed on the cell surface of *Helicobacter pylori*. *Journal of Medical Microbiology.* 45: 270-277.

Yan, J., Cheng, Q., Li, C.B. and Aksoy, S. (2001a) Molecular characterization of three gut genes from *Glossina morsitans*: cathepsin B, zinc-metalloprotease and zinc-carboxypeptidase. *Insect Molecular Biology.* in press.

Yan, J., Cheng, Q., Li, C.B. and Aksoy, S. (2001b) Molecular characterization of two serine proteases expressed in gut tissue of the African trypanosome vector, *Glossina morsitans morsitans*. *Insect Molecular Biology.* 10: 47-56.

Yoshida, N., Oeda, K., Watanabe, E., Mikami, T., Fukita, Y., Nishimura, K., Komai, K. and Matsuda, K. (2001) Chaperonin turned insect toxin. *Nature.* 411: 44.

Zdarek, J. and Denlinger, D.L. (1993) Metamorphosis behaviour and regulation in tsetse flies (*Glossina* spp.) (Diptera: Glossinidae): a review. *Bulletin of Entomological Research.* 83: 447-461.

Appendix I. Abbreviations

1-D	one dimensional
2-D	two dimensional
A	alanine
BLASTP	Basic Local Alignment Search algorithm for proteins
bp	base pair
bpm	beats per minute
BSF	bloodstream forms
C	cysteine
°C	Celsius
CapLC system	capillary liquid chromatography system
CBB	colloidal Coomassie brilliant blue G-250
CDC	Centers for Disease Control
cDNA	complementary deoxyribonucleic acid
CDS	coding derived subsequence
CHES	2-[N-cyclohexylamino]ethanesulfonic acid
CI	cytoplasmic incompatibility
Ci	Curie
CSF	cerebro spinal fluid
D	aspartic acid
d	day
Da	Dalton
dATP	deoxyadenosinetriphosphate
DEPC	diethyl pyrocarbonate
DNA	deoxyribonucleic acid
dNTP	deoxynucleotidetriphosphate
ds	double stranded
dT	deoxythymidine
DTT	dithiothreitol
EBI	European Bioinformatics Institute
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme linked immunosorbent assay
EMBL	European Molecular Biology Laboratory
EP	glutamic acid:proline dipeptide
EtBr	ethidium bromide
EtOH	ethanol
F	phenylalanine
FACS	fluorescence activated cell sorter
FITC	fluorescein isothiocyanate
fm	femptomole

FW	free weight
G	glycine
G	guage
GC content	guanine/cytosine
Gmm	*Glossina morsitans morsitans*
GP01 F/R	*Sodalis* specific primers to plasmid DNA
Gp5.3	*Glossina palpalis* clone 5.3
Gpp	*Glossina palpalis palpalis*
H	histidine
h	hour(s)
HAT	human African trypanosomiasis
Hsp60 (GroEL)	heat shock protein 60
I	isoleucine
Ig	immunoglobulin
IPTG	isopropyl β-D-Thiogalactopyranoside
ISO-DALT	isoelectric focusing (first dimension) and separation by molecular weight (in Daltons; second dimension)
K	lysine
kb	kilobase
kDa	kilodaltons
KOH	potassium hydroxide
L	leucine
LB medium	Luria-Bertani medium
LRH	Lee Rafuse Haines
m	milli
μ	micro
M	molarity
M	methionine
mAb(s)	monoclonal antibodies
MALDI-TOF	matrix assisted laser desorption ionization time-of-flight
μCi	microCurie
mg	milligram
min	minute(s)
μL	microliter
mM	millimolar
MMI medium	Mitsuhashi and Maramorosch Insect medium
mol	mole
mRNA	messenger RNA
n	nano
N	asparagine
NCBI	National Center for Biotechnology Information

NEB	New England Biolabs
ng	nanogram
NIAID	National Institute of Allergy and Infectious Disease
NIH	National Institute of Health
nr	non-redundant
NSERC	National Sciences and Engineering Research Council of Canada
nt	nucleotide
OD	optical density
p	pico
P	proline
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	Brookhaven Protein Data Bank
pfu	plaque forming units
pI	isoelectric point
PIR	Protein Information Resource
pmol	picomole
pp	page(s)
PRF	Protein Research Foundation
PVDF	polyvinylidene difluoride
Q	glutamine
Q-TOF	quadruple time-of-flight mass spectrometer
R	arginine
RACE	rapid amplification of cDNA ends
RHG	Ron H. Gooding
RLO	rickettsia-like organisms
RLO1 F/R	rickettsia-like organisms specific forward and reverse primers
RNA	ribonucleic acid
RNAse	ribonuclease A
rpm	revolutions per minute
RT	reverse transcriptase
RT	room temperature
S	serine
SA	Serap Aksoy
SDS-PAGE	sodium dodecyl sulphate - polyacrylamide gel electrophoresis
sec	second(s)
Sg	*Sodalis glossinidius*
SIT	sterile insect technique
SOPE	*Sitophilus oryzae* principal endosymbiont
SSC buffer	saline-sodium citrate buffer
SwissProt	Swiss Institute of Bioinformatics Protein Database

T	threonine
T4 PNK	T4 polynucleotide kinase
TAE buffer	Tris-acetate/EDTA buffer
Taq DNA polymerase	*Thermus aquaticus* DNA polymerase
TIFF	tag image file format
TMHMM vs. 2.0	transmembrane detection- hidden Markov model algorithm
Tris	tris(hydroxymethyl)aminomethane
TWP	Terry W. Pearson
TYLC virus	tomato yellow leaf curl virus
UPM	universal primer mix
UV	ultra-violet
V	Volt
V	valine
v/v	volume per volume
Vh	volt-hour
VSG	variant surface glycoprotein
W	tryptophan
w/v	weight per volume
Wg	*Wigglesworthia glossinidia*
Y	tyrosine

Appendix II: List of Suppliers

Item	Supplier	Address
10 kDa molecular mass ladder	Gibco BRL®	Burlington, ON
15 ml polystyrene sterile screw cap tube	Sarstedt, Inc.	Newton, SC
Acrodisc™ syringe filter	Pall Gelman Laboratories	Ann Arbor, MI
α-cyano-4-hydroxycinnamic acid	Sigma-Aldrich Canada	Milwaukee, WI
adenosine 5' triphosphate	Amersham Pharmacia Biotech Inc.	Baie'd Urfe, PQ
agar agar	Difco	Detroit, MI
agarose LE (molecular biology grade)	Promega Corp.	Madison, WI
Alexa™ Fluor 488	Molecular Probes	Eugene, OR
Anti-Hsp60 SPA-807 mAb	StressGen Biotechnologies	Victoria, BC
BBL™ CampyPak Plus	Becton Dickinson	Sparks, MD
Bioanalyst Software	PE-SCIEX	Boston, MA
BioTrace™ (PVDF)	Pall Corporation	Ann Arbor, MI
CapLC system	Waters Corp.	Milford, MA
CHES	Sigma-Aldrich Canada	Oakville, ON
ClonaCell HY™ System	StemCell Technologies Inc.	Vancouver, BC
colloidal Coomassie brilliant blue G-250	EM Science	Gibbstown, NJ
designed oligonucleotide probes	Gibco BRL®	Burlington, ON
dNTP mix (10 mM each)	Amersham Pharmacia Biotech Inc.	Baie'd Urfe, PQ
DTT	Sigma-Aldrich Canada	Oakville, ON
Duralon UV noncharged nylon membranes	Stratagene	La Jolla, CA
EtBr	Gibco BRL®	Burlington, ON
FACSCalibur flow cytometer	Becton-Dickinson	San Jose, CA
Falcon 3915 PRO-BIND ELISA Plates	Becton Dickinson	Oxnard, CA
FITC conjugated mAbs	Caltag Laboratories	Burlingame, CA
GelCode™ Blue	Pierce Chemical Company	Rockford, IL
goat anti-mouse IgG/IgM-horseradish peroxidase	Caltag Laboratories	San Francisco, CA
Hybond N+ nylon membrane	Amersham Pharmacia Biotech Inc.	Baie'd Urfe, PQ
ImageQuant™ analysis software	Molecular Dynamics	Sunnyvale, CA
isotyping ELISA kit	American Qualex	La Mirada, CA
kanamycin A	Sigma Chemical Company	St. Louis, MO
Kodak Biomax MR film	Eastman Kodak Company	Rochester, NY
Labquake Shaker	Labindustries	Berkeley, CA
Lambda digest DNA standards	New England Biolabs	Beverly, MA
MALDI internal peptide standards	Sigma Chemical Company	St. Louis, MO
MALDI Plate (stainless steel)	Applied Biosystems	Foster City, CA
Mass Spectrometer DE-STR	Applied Biosystems	Foster City, CA
MicroPoly (A) Pure mRNA Isolation Kit	Ambion Inc.	Austin, TX

Mini-Protean II one-dimension gel apparatus	Bio-Rad Laboratories	Hercules, CA
MMI medium	Sigma-Aldrich Canada	Oakville, ON
Nanosep MF filters	Pall Filtron Corp.	Northborough, MA
oligo (dT) cellulose resin	Ambion Inc.	Austin, TX
OmniPur™ low melt agarose gel	EM Science	Gibbstown, NJ
Pellet Pestle	KONTES Glass Company	Vineland, NJ
Perkin-Elmer gas-phase sequencer	Applied Biosystems	Foster City, CA
PE-SCIEX Q-STRi	Applied Biosystems	Foster City, CA
Pharmalyte™ 3-10 ampholines	Amersham Pharmacia	Upsala, Sweden
Phosphor Screen cassettes	Molecular Dynamics	Sunnyvale, CA
Photoshop ™ 5.5 graphic software	Adobe Systems Inc.	San Jose, CA
poly-L-lysine	Sigma Chemical Company	St. Louis, MO
porcine trypsin (sequence grade)	Promega Corp.	Madison, WI
POROS R2 C18 Resin	PerSeptive Biosystems	Foster City, CA
Primer Premier ™	PREMIER Biosoft International	Palo Alto, CA
Protein A- agarose beads	EY Labs	San Mateo, CA
QIAEX II Gel Extraction Kit	Qiagen Inc.	Mississauga, ON
Qiaprep Spin MiniPrep Kit	Qiagen Inc.	Mississauga, ON
Q-TOF glass capillary needles	Protana Inc.	Staermosegaards, Denmark
Q-TOF mass spectrometer	Micromass	Beverly, MA
restriction endonucleases	New England Biolabs	Beverly, MA
RNase H	United States Biological Corp.	Cleveland, OH,
RNase H- reverse transcriptase	Canadian Life Technologies	Burlington, ON
Sephadex™ G-25F gel permeation column	Amersham Pharmacia Biotech Inc.	Baie'd Urfe, PQ
SEQUEST ™	Finnigan MAT	San Jose, CA
SERVA Blue G	SERVA Electrophoresis	Heidelberg, DRG
SMART RACE cDNA Amplification Kit	CLONTECH Laboratories Inc.	Palo Alto, CA
Speed Vac Concentrator	Savant	Hicksville, NY
Stratagene UV Stratalinker 2400	Stratagene	La Jolla, CA
SuperSignal Dura™ chemiluminescence substrate	Pierce Chemical Company	Rockford, IL
T25 sterile tissue culture flasks	Corning Costar Corp.	Cambridge, MA
T4 polynucleotide kinase	New England Biolabs	Beverly, MA
Taq DNA polymerase	University of Victoria: C. Upton	Victoria, BC
Taq DNA polymerase buffer	Canadian Life Technologies	Burlington, ON
Thermal Cycler (GeneAmp System 2400)	Perkin-Elmer	Norwalk, CN
TOPO™ cloning kit	Invitrogen	Carlsbad, CA
UMAX Astra 3400	UMAX Technologies, Inc.	Freemont, CA
VacuGene XL Vacuum blotting System	Pharmacia Biotech Inc.	Uppsala, Sweden
Zap Express cDNA Synthesis Kit	Stratagene	La Jolla, CA

Vita

Surname: Haines *Given Names*: Lee Rafuse

Place of Birth: Prince George, BC, Canada

Educational Institutions Attended:

Trinity Western University 1989-1994

Degrees Awarded:

B. Sc. (Major) May 1994

Honors and Awards:

UVic Fellowship for Graduate Studies	1999-2001
Alexander James Wood and Eileen Ford Wood Scholarship (UVic)	2000-2001
Graduate Teaching Research Fellowship (UVic)	2000-2002
Graduate Student Award for Excellence in Teaching (UVic)	Mar 2001

Publications:

Haines, L.R., Haddow, J. H., Gooding, R. H. and Pearson, T. W. (2002). Surface association and extracellular release of a chaperonin from *Sodalis glossinidius*, a symbiont of the tsetse. *Microbiology* (submitted Apr 2002).

Haines, L.R., Haddow, J. H., Aksoy, S., Gooding, R. H. and Pearson, T. W. (2002). The major protein in the midgut of teneral *Glossina morsitans morsitans* is a molecular chaperone from the endosymbiotic bacterium *Wigglesworthia glossinidia*. *Insect Biochemistry and Molecular Biology* (in press, accepted April 2002).

Haddow, J. D., Poulis, B. D., **Haines, L. R.**, Gooding, R. H., Aksoy, S. and Pearson, T. W. (2002). Identification of major soluble salivary gland proteins in teneral *Glossina morsitans morsitans*. *Insect Biochemistry and Molecular Biology* (in press, accepted Jan 2002).

Chase, J. C., Coates, J. A., Haddow, J. D., Stewart, M. H., **Haines, L. R.**, Whitaker, D. J., Kent, M. L., Olafson, R. W. and Pearson, T. W. (2001). Analysis of *Kudoa thyrsites* (Myxozoa: Myxosporea) spore antigens using monoclonal antibodies. *Diseases of Aquatic Organisms* 45: 121-129.

Chase, J. C., Coates, J. A., Haddow, J. D., Stewart, M. H., **Haines, L. R.**, Whitaker, D. J., Kent, M. L., Olafson, R. W. and Pearson, T. W. Soft flesh syndrome in Atlantic salmon: an immunodiagnostic test for *Kudoa thyrsites* (Myxozoa: Myxosporea) in the operculum and somatic muscle tissue of infected fish. *Diseases of Aquatic Organisms* (in preparation).

Haines, L.R. and Pearson, T. W. Characterization of a novel protein with extensive proline-glutamic acid repeats from the tsetse fly *Glossina palpalis palpalis*. *Insect Molecular Biology* (in preparation).

Research Presentations and Technical Reports

Haines, L. R., Chase, J. C., Jackson, A. and Pearson, T. W. Identification of *Theileria parva* macroschizont proteins using protein chemistry and mass spectrometry. Meeting with the DFID, ILRI, TIGR research group on vaccines for East Coast fever. Nairobi, Kenya. December 2001.

Haines, L. R. Development of a diagnostic test for occult infections with the filarial parasites *Onchocerca volvulus* and *Loa loa*. Med Biotech Laboratories. Kampala, Uganda. December 2001.

Haines, L. R., Gooding, R. W., Haddow, J. D., Hache, N., Stewart, M. H., Ruepp, S., Roditi, I. and Pearson, T.W. Molecular interactions between African trypanosomes and tsetse: a role in differentiation and cell death? Poster and Abstract for 25th Meeting of the International Scientific Council for Trypanosomiasis Research and Control. Mombasa, Kenya. October 1999.

Haddow, J. D., **Haines, L. R.**, Lippert, D. and Poulis, B. A. D. (1999). Proteomics techniques: a report describing optimization of 2-D gels, spot digestion and mass spectrometry developed at the University of Victoria.

University of Victoria Partial Copyright License

I hereby grant the right to lend my thesis to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the Library of any other university, or similar institution, on behalf or for one of its users. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis:

The identification and partial characterization of molecules found within the midgut of the tsetse fly.

Author: 

Lee Rafuse Haines

Date: April 26 2002