FUNCTIONAL GENOMICS IN INSECT IMMUNOLOGY: IDENTIFICATION AND CHARACTERIZATION OF *RHODNIUS PROLIXUS* IMMUNE GENES.

by

Raul URSIC BEDOYA

Maitrise,	Université Joseph Fourier	2001
	University of Wisconsin – Madison	
Licence,	Université Joseph Fourier	2000

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

In the
Department of Biological Sciences

SIMON FRASER UNIVERSITY

Summer 2008

APPROVAL

Name: Raul Ursic Bedoya

Degree: Doctor of Philosophy

Title of Thesis:

Functional genomics in insect immunology: Identification and characterization of *Rhodnius prolixus* immune genes

Examining Committee:

Chair: Dr. D. Green, Assistant Professor

Dr. C. Lowenberger, Associate Professor, Senior Supervisor
Department of Biological Sciences, S.F.U.

Dr. C. Beh, Assistant Professor
Department of Molecular Biology and Biochemistry, S.F.U.

Dr. N. Harden, Associate Professor
Department of Molecular Biology and Biochemistry, S.F.U.

Dr. N. Haunerland, Professor
Department of Biological Sciences, S.F.U.

Dr. M. Moore, Professor
Department of Biological Sciences, S.F.U.
Public Examiner

Dr. E. Huebner, Professor
Department of Biological Sciences, University of Manitoba
External Examiner

4 July 2008
Date Approved

ii

Declaration of Partial Copyright Licence

SFU SIMON FRASER UNIVERSITY
THINKING OF THE WORLD

STATEMENT OF
ETHICS APPROVAL

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

(a) Human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

(b) Advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

(c) as a co-investigator, in a research project approved in advance,

or

(d) as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Bennett Library
Simon Fraser University
Burnaby, BC, Canada

ABSTRACT

Rhodnius prolixus is a bloodfeeding hemipteran insect and a vector of *Trypanosoma cruzi*, the etiologic agent of Chagas disease. This disease affects over 18 million people in Latin America, and there are no vaccines or efficacious drugs. Because of the unique lifecycle of *T. cruzi* within the vector, and its avoidance of the hemocoel where the primary immune response occur, we investigated the role of the innate immune response of *R. prolixus* to pathogens and parasites, including *T. cruzi*. Insect immunity is based on three principles: pathogen recognition, induction of appropriate regulatory pathways; and the production of molecules to eliminate the pathogens. In this dissertation, I use functional genomics (Suppressive Subtractive Hybridization) to identify and describe genes involved in the immune response of *R. prolixus* and the implications for *T. cruzi*. Initially we identified expressed sequence tags (ESTs) corresponding to genes that are differentially expressed in response to parasites and bacteria, and assigned putative gene functions based on sequence similarities. This thesis focuses on four genes of interest including a transcription factor (Dorsal), which may function in the regulation of expression of immune peptides, and three antimicrobial peptides (Defensin, Lysozyme and Prolixin) that serve directly to eliminate pathogens. Transcriptional factor binding motifs (NF-κB) present in the promoters of two of these genes (lysozyme and defensin) suggest they are transcriptionally regulated by Dorsal, whereas prolixin is not. We evaluated the temporal and spatial expression profiles of the antimicrobial peptide genes, using real time quantitative PCR, to establish molecular relationships between parasite and vector. Subsequently, we expressed dorsal and prolixin in bacteria and tested their functions. We concluded that invasion of the hemocoel of *R. prolixus* activates components of the immune system and the production of compounds lethal to *T. cruzi*, but the pathogen survives by living

exclusively in the intestine avoiding vector responses. This study contributes to our overall knowledge of insect immunity, the arsenal of immune molecules available to different insects, and identifies novel and highly conserved immune molecules found in higher and lower insects.

To my family,

    Davor, Teresa, Mercedes, José y Angela.

Aunque a veces "la vida es triste y jodida", aun vale la pena luchar por ella y vivirla al máximo. Soy quien soy gracias a ustedes.
Je vous aime très fort.

ACKNOWLEDGEMENTS

The work presented in this dissertation would have not been possible it were not for the participation of several individuals whom I would like to thank.

First and foremost, I would like to thank Carl Lowenberger whom since the very first day we met in the fall of 2000 at the Animal Health and Biomedical Sciences department of the University of Wisconsin – Madison, believed in me and gave me his unconditional support. It truly has been a fantastic experience, where, I believe, we have both learned a lot. For me, from being a third year undergraduate with some theoretical background and hardly any hands on experience to finishing my PhD with several publications, it has been the most personally fulfilling achievement in my life so far. It would be impossible to overstate the importance of Carl's guidance and friendship to this time of my life and my future.

I would like to express my sincere gratitude as well to the rest of my doctoral committee: Nicholas Harden, Christopher Beh and Norbert Haunerland for their valuable feedback and friendly advice throughout my time in graduate school.

Several members of the Lowenberger laboratory have been instrumental to my degree. These people include Dawn Cooper and Hamed Nazzari, with who is has been a pleasure to work and laugh with. In addition, Kendra Foster, Jutta Buchhop, Jerry Ericsson, Richard Pluncket, Ranil Waliwitiya and Carolina Perez who helped to keep a healthy atmosphere in the laboratory and for their contribution to some of the work presented in this dissertation. Many undergraduate assistants have really contributed to my work as well, including: Dan Fornika, Stuart Murray, Angela Babuk and Sarah Fairhurst.

Some people helped making the unpleasant administrative life of graduate school much less stressful, including: Fiona Burrows, Marlene Nguyen and Amelia Siu.

Although a significant proportion of my time in the past six years was spent inside the laboratory, I am fortunate to have a phenomenal group of friends to remind me that there is life outside the lab. Many thanks to: Mathias Schuetz, the best roommate I have ever had (Go Croatia!), and his wife Ioanna who are like family to me. Carolina Ordoñez, Rishi Kwatra, Arthur Drozdzik and Alicia Barcenas for their complicity and the late, fun nights on Granville Street. The rest of the Ordoñez family (Ricardo, Rocio, Pablo and Pamela) for practically adopting me and making me feel closer to home. The rest of the Cooper family (Gordon, Iris and Garret) for showing me the beautiful B.C and for the many drinks and food we shared. Barbara Stewart for always buying me a birthday cake and putting up with the celebrations.

"A healthy mind in a healthy body" (Juvenal (55 AD - 127 AD)) they say, and I could not have achieved this if it was not for my friends in the Bio/Mbb/Rem intramurals soccer team (Fall 2007 champions!) which include some folk already mentioned but also: Bradley Davis, Michel Leroux, Jen Silver, Jen Barret, Jonathan Cummings, Courtney Mckay, Michael Healey, Nick Inglis and Paul Menard.

Several people of the neighboring labs were great friends and colleagues. These include: Mariana Oviedo (and Brent), Christiana Cheng, Mel Hart, Ryan Conder, Lisa Neame, Susana Patiño, Kat Gazdik and Andreas Steimel.

The time I spent on the Graduate Issues Committee and the Bisc graduate caucus was satisfying and enriching primarily because of the people involved. Most have been mentioned already, others include: Brian Jones, Nicole Turnbridge, Keith Tierney, Suzanne Gray, Andrea Pomeroy, Clea Moray and Mike Letourneau.

Outside the lower mainland, many people were invaluable. Despite the physical distance between us, they never forgot about me and always treated me

like I never left Bolivia. My best friends: Carlos Bonadona, Vania Loayza, Macarena Zannier and the rest of Instantes '96. My favorite uncle and aunt, Marcelo and Martha Claure for allowing me to crash at their home in Miami countless times.

Last but not least, I would like to thank my immediate family; my parents Davor and Teresa, my sisters Mercedes and Angela, and my brother José (the real Dr. Ursic!) for their love, understanding and unconditional support.

I hope I have acknowledged everyone who deserves it, but if I have omitted anyone it has been by pure forgetfulness and not by will.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

AMP   Antimicrobial peptide.

Contig   A physical map of contiguous genomic DNA assembled using overlapping cloned segments.

EST   Expressed Sequence Tags are small pieces of DNA sequence (usually 200 to 500 nucleotides long) that are generated by sequencing either one or both ends of an expressed gene.

Hemimetabolous Refers to insects that undergo incomplete metamorphosis to reach the adult stage. They are characterized by three different life cycle stages: egg, nymph and adult. Development from nymph to adult takes place through gradual changes.

Hemolymph   Insects have an open circulatory system. The hemolymph is the blood-like fluid that bathes tissues and is circulated throughout the body cavity by the dorsal vessel, a functional equivalent of the heart.

Holometabolous Refers to insects that require complete metamorphosis to reach the adult stage. They are characterized by four different life cycle stages: embryo, larva, pupa and adult.

Homolog   A gene related to a second gene by descent from a common ancestral DNA sequence. The term, homolog, may apply to the relationship between genes separated by the event of speciation or to the relationship between genes separated by the event of genetic duplication.

Melanization   The deposition of melanin at the site of injury as a result of the activation of a biochemical cascade involving a key enzyme, phenol oxidase, which is activated in response to septic injury. This activation releases toxic reactive oxygen species that may attack invading microbes.

Ortholog   Orthologs are genes in different species that evolved from a

common ancestral gene by speciation. Normally, orthologs retain the same function in the course of evolution. Identification of orthologs is critical for reliable prediction of gene function in newly sequenced genomes.

Paralog

Paralogs are genes related by duplication within a genome. Orthologs retain the same function in the course of evolution, whereas paralogs evolve new functions, even if these are related to the original one.

Proteome

The complete collection of proteins in a particular cell.

RNA interference

The silencing of gene expression by the introduction of double-stranded RNAs that trigger the specific degradation of a homologous target mRNA and often subsequently decrease production of the encoded protein.

Transcriptome

The complete collection of mRNAs in a particular cell.

CHAPTER 1: INTRODUCTION

URSIC BEDOYA Raul
Simon Fraser University
Department of Biological Sciences
Burnaby, British Columbia, V5A 1S6, Canada

Arthropod Borne diseases

Arthropod transmitted infectious and parasitic diseases cause severe human mortality and morbidity throughout the world particularly in developing countries where health conditions are poor. Malaria, yellow fever, dengue, African and American trypanosomiasis, encephalitis, leishmaniasis, filariasis, and onchocerciasis are all transmitted to humans by insects and continue to afflict a large proportion of the world's population and place enormous burdens on public health systems. According to the World Health Organization, all combined, these diseases cause the death of over two million individuals per year (WHO, 2002b), with malaria being the deadliest of them all.

Malaria, tuberculosis, and HIV/AIDS are the three most devastating diseases as determined by the World Health Organization, and only malaria is vector-borne. For more than 80 years, massive spaying of insecticides controlled insect borne diseases with spectacular results. The discovery of dichlorodiphenyltrichloroethane (DDT) as an insecticide in the 1930s helped drastically decrease the number of agricultural pests and disease transmitting insects (WHO, 1979). Unfortunately, the ban of DDT in the 1970s, the high cost of long term insecticide vector control, coupled with the increased emergence of insecticide resistance and parasite resistance to drugs have resulted in a re-emergence of vector-borne diseases worldwide.

Over the past decade, there has been a significant increase in the amount of public attention and funding given to tropical diseases, especially to malaria by the Gates foundation, and this has propelled advances in the field of vector research. In fact, US politicians and private foundations have determined and pronounced publically that they can, and will eradicate malaria. However, other diseases (termed neglected diseases) which almost exclusively affect people in developing countries who are too poor to pay for any kind of treatment have been ignored by pharmaceutical companies as they do not constitute a market that can

2

attract investment in drug research and development. One of the groups of neglected diseases is those caused by trypanosomes. These include African sleeping sickness and the American Chagas disease.

Trypanosomes and trypanosomiases

The trypanosomes are protozoans belonging to the order Kinetoplastida and the family Trypanosomatidae. This family includes genera: *Trypanosoma, Leishmania, Endotrypanum, Phytomonas, Leptomonas, Herpetomonas, Crithidia, Blastocrithidia* and *Rhynchoidomonas*. Members of the genus *Trypanosoma* (single-celled flagellates) are all parasitic, characterized by the presence of an organelle called the kinetoplast which contains mitochondrial DNA (Marquardt *et al.*, 2000). Several members of this genus infect mammals and are transmitted by arthropod vectors. Mammalian trypanosomes are subdivided in two groups depending where infective forms are produced in the arthropod vector:

- salivaria: infective forms are produced in the salivary glands and transmitted to the vertebrate host by feeding.
- stercoraria: infective forms are produced in the posterior region of the digestive tract, and released with the feces and urine.

These organisms are responsible for important veterinary and human diseases the trypanosomiases (Barrett *et al.*, 2003).

Among the salivaria, three subspecies of *Trypanosoma brucei* are responsible for trypanosomiases in the African continent:

- *T. b. brucei* causes animal trypanosomiases in wild and domestic animal species. The disease in cattle is named Nagana and represents a major obstacle to the economic development of impoverished rural areas affected. Animals can also host the human pathogen parasites.
- *T. b. gambiense* in West and Central Africa causes 90% of reported cases of Human African Trypanosomiasis (HAT), a chronic infection affecting the

central nervous system that can remain asymptomatic for an indefinite period of time.

- *T.b. rhodesiense* in East Africa causes the remaining 10% of reported cases of HAT but is characterized by an acute infection of the central nervous system which shows symptoms within weeks.

Trypanosoma brucei parasites are transmitted to mammals by both sexes of Tsetse flies (Diptera, Glossina spp.); large biting flies which are strictly hematophagous. Tsetse flies cover a large area of sub Saharan Africa and are closely associated with vegetation found by water sources, in gallery-forests and wooded savannah. Over 30 species and sub-species have been described but only 6 of them are recognized as vectors. Tsetse flies acquire *T. brucei* by blood feeding from an infected mammal. The parasite migrates to the insect's digestive tract along with the blood and differentiates into midgut procyclics capable of perforating the peritrophic matrix, which holds the blood meal. In the space between the peritrophic matrix and the gut epithelia, parasites divide by binary fission until they reach a population of 10^5-10^6 individuals. Subsequently, trypanosomes migrate to the proventriculus (anterior midgut), before penetrating through the peritrophic matrix again, and undergoing further differentiation into long migratory forms, which migrate and invade the salivary glands of the fly. In the salivary glands, epimastigote forms attach to the epithelium via the flagellum and multiply. In the final stage inside the arthropod vector, epimastigotes become infective metacyclics, which infect the vertebrate host with the saliva during blood feeding. The metacyclics multiply for a few days at the bite site, and then invade the bloodstream and lymphatics and subsequently the tissues, causing the most damage to the central nervous system (CNS) (Fenn and Matthews, 2007). In addition to vectorial transmission, Human African Trypanosomiasis (HAT) can be acquired through blood transfusion and from an infected expecting mother to her unborn child if the trypanosome crosses the placenta and infect the fetus.

HAT is fatal if left untreated. Chemotherapy to treat the disease exists and the appropriate drug to use depends on the parasite and the stage of the disease

(Croft *et al.*, 2005). Pentamidine or Suramin are used for the treatment of the first stage of *T.b. gambiense* or *T.b. rhodesiense* HAT respectively. Melarsoprol is used for the treatment of the second stage of the disease, and alternatively Eflornithine is used but only against *T.b. gambiense*. Unfortunately, these drugs, particularly Melarsoprol have significant toxic side effects for the patient. The World Health Organization estimates the number of cases is currently between 50,000 and 70,000.

In the Americas, a different type of parasite is found, *Trypanosoma cruzi* a stercorarian trypanosome which causes American trypanosomiases. *Trypanosoma cruzi* is also transmitted by hematophagous arthropods from a different insect order: the kissing bugs (Hemiptera). Both of the trypanosomiases are zoonoses; the infectious agent is naturally transmitted between animal reservoir hosts and humans. American trypanosomiasis is a complex disease in part because it involves a heterogeneous parasite population consisting of multiple strains separated into two major phylogenetic groups. Historically, *T. cruzi* strains were classified based on two different criteria (Momen, 1999):

- Zymodemes were classified based on different enzymatic profiles of parasite strains;
- Schizodemes were classified based on different electrophoretic mobility profiles of kinetoplast mitochondrial DNA.

An important iso-enzymic variability between parasite isolates exists, but three major groups or zymodemes (Z1, Z2 and Z3) were originally established. Zymodemes Z1 and Z3 have been associated with the sylvatic cycle and Z2 with the domestic cycle.

Subsequently, these three original zymodemes were further divided into 43 groups when more parasite stocks and a broader range of enzymes were studied. Phylogenetic reconstruction of these zymodemes allowed the identification of two major groups, each exhibiting considerable genetic

variability. The current hypothesis is that the population structure of *T. cruzi* is clonal instead of sexual and that the genetic and biological variability are the result of the independent evolution of clonal lines (Tibayrenc *et al.*, 1990, Tibayrenc, 1995).

Nuclear DNA markers such as the 24Sa ribosomal RNA (rRNA) gene and the intergenic region of the mini-exon genes support this dichotomy within the *T. cruzi* taxon. Currently, clustering of strains into these two principal groups are denominated *T. cruzi* I and II (Buscaglia and Di Noia, 2003). *T. cruzi* I is preferentially associated with the sylvatic transmission cycle, although it can also be found in the domestic one, whereas *T. cruzi* II is mainly associated with the domestic cycle.

Strain classification is important because there is significant difference in the incidence and severity of infections of each strain. Different strains have different tissue tropism in the vertebrate host. Some strains will preferentially infect gastrointestinal tissues whereas other will infect cardiac tissues, and mixed infections will have the most damaging effects for the infected individual (Dutra *et al.*, 2005). Furthermore, different strains and mixed infections have shown different responses to treatment with currently available drugs.

In 1994, The World Health Organization's Special Programme for Research and Training in Tropical Diseases (TDR) launched the *T. cruzi* genome project. Recently, the genome of the CL Brener strain of this parasite was completed and its annotation revealed the preponderance of mucins, surface glycoproteins containing a varied array of O-liked oligosaccharides. Close to 6% of all predicted genes encode such molecules highlighting their importance in the biology of the parasite (El-Sayed *et al.*, 2005). *Trypanosoma cruzi* mucins (TcMUC) are complex and heterogeneous (Buscaglia *et al.*, 2006), however they can be divided into two major types: those present in the insect stages and those present in the mammalian stages.

6

Inside the insect vector, TcSMUG, (*T. cruzi* small mucin-like gene family) mature products from epimastigotes and metacyclic trypomastigotes are 35–50kDa proteins that contain no repeats, no O-glycosylated Threonine runs, and no variable regions. Mucins from cell-derived trypomastigotes (tGPI-mucins) are significantly larger and more complex; their size ranges between 60 and 200kDa. Although up to 25% of the mucin genes in the *T. cruzi* genome are non-functional pseudo genes, the remarkable complexity and variety of these surface molecules makes them ideal molecules to defeat the host immune system.

American trypanosomiasis is commonly called Chagas disease in memory of Carlos Chagas, the Brazilian physician who first identified and described the agent and disease (Chagas, 1909). In 2000, it was estimated that 18 million people were infected with Chagas disease, 100 million were at risk of acquiring it and approximately 20,000 per year died because of it (WHO, 2002a). This disease is highly correlated with poverty and poor housing conditions and thus affects mostly impoverished rural populations in the Americas. In terms of morbidity, the disease represents the highest burden of disease among tropical diseases in the Americas representing close to 700,000 Disability Adjusted Life Years (DALY: the number of healthy years of life lost due to premature death and disability) (WHO, 2002a). Total direct and indirect costs, such as medical treatment and loss in production, of Chagas disease represent a major economic burden for Latin American countries.

Trypanosoma cruzi, the etiologic agent of Chagas disease infects a broad range of mammals and is transmitted to humans through blood transfusion, congenital transmission or by the blood sucking Reduviidae (Hemiptera, sub family Triatominae) insects. Currently, about 80% of transmission is vector borne, estimated to be 200,000 new cases annually. Triatome insects feed exclusively on vertebrate blood, and over 130 species have been described. The most important species of Chagas' disease vectors are *Triatoma infestans, Triatoma*

dimidiata, Triatoma brasiliensis, Triatoma maculata, Triatoma sordida, Rhodnius prolixus, Rhodnius neglectus, Rhodnius pallescens and *Panstrongylus megistus* covering a geographical area spanning from the southern United States to the south of Argentina (WHO, 2002a).

In contrast to many vector-borne diseases, transmission of *T. cruzi* does not occur as the vector injects saliva into its host during blood feeding. Instead, upon blood feeding, the insect defecates on the host's skin. The feces contain infective metacyclic trypomastigotes, which reach the host's peripheral blood system via contamination of conjunctiva, mucous membranes or micro abrasions produced after scratching of the bite site. In the chronic stage of the disease, essential organs are irreversibly damaged, thus it is a serious debilitating disease.

The human disease is clinically described in three phases: the acute phase lasts 1-2 months and is characterized by inflammation at and near the site of infection (Chagoma) and general malaise. During this stage, parasites remain in the peripheral blood system, thus being the phase which drugs are most effective. Fatality due to Chagas disease occurs in a minority of cases (2-8%) during the acute phase mostly affecting young children due to myocarditis or memingoencephalitic complications (WHO, 2002a).

Following the acute phase, a long asymptomatic phase, termed indeterminate, which typically lasts 10-20 years begins; the parasite infects tissues intracellularly and thus remains hidden from the host's immune system. The majority of infected individuals remain in this stage, however, depending on the endemic area; 20-30% will advance into the chronic phase. The latter phase is characterized by serious and irreversible cardiac, neurological or digestive tissue damage caused by intracellular development of amastigotes. Most of the morbidity and mortality due to Chagas disease occur during the chronic stage affecting adults aged 20-50 years old. About half of infected adults develop cardiac problems, such as arrhythmia or apical aneurysm, leading to heart failure and premature death. The other half develops digestive problems associated

with distension of parts of the alimentary tract. Even though the vast majority of acute infections are asymptomatic, 10 to 30% of these cases become chronic. Acute infections with Chagas disease affect the digestive tract, but the more common outcome is a chronic infection of the heart disabling and killing young adults at the peak of their working life. It is important to point out that there is no preventive vaccine or effective treatment to cure chronic Chagas disease. The development of a vaccine for the disease is unlikely because *T. cruzi* antigens can stimulate autoimmunity (WHO, 2002a).

Chagas chemotherapy is based on two nitroheterocyclic compounds nifurtimox and benznidazole. The results obtained with both drugs vary according to the phase of Chagas disease, the period of treatment and the dose, the age and geographical origin of the patients. Unfortunately, they have a very limited efficacy in the prevalent chronic stage, and toxic side effects (Rodriques Coura and de Castro, 2002). Nifurtimox's side effects include anorexia, weight loss, psychic alterations, excitability or sleepiness and intestinal disturbances, such as nausea, vomit or diarrhea. The most serious side effects induced by benznidazole are agranulocytosis, sore throat, fever, septicemia, and mucosal bleeding. Good results have been achieved in the acute phase, particularly in children under 12 years of age. Chemotherapy dosage varies with the stage of the disease but it can last up to 90 consecutive days (Croft *et al.* 2005). The prohibitive cost, toxic side effect and prolonged treatment often lead to treatment cessation.

Several diagnostic tools are available: symptomatic (Romaña's sign), parasitological (xenodiagnosis; blood feeding of uninfected vectors with patient's blood, followed by microscopic analysis of the insect's feces to detect the parasite); serological (ELISA, Indirect immunofluorescence and indirect haemagglutinnation) and molecular (PCR). PCR is likely the most sensitive technique as it is possible to detect DNA from one parasite in 20 mL of blood (Avila *et al.*, 1991), but it is difficult to implement its use in the field.

However, in large part because Chagas disease affects neglected impoverished rural populations, it often goes undiagnosed and thus untreated (Moncayo, 1999, WHO, 2002a, Dutra *et al.*, 2005). Chagas is a disease of the poor, mainly because poor house construction creates habitat for the insect vectors. As such, it comes as no surprise that Bolivia as one of the poorest countries in South America where in 1999, 63% of the households were living in poverty (mostly indigenous rural populations) faces a major challenge in defeating this disease.

Vector biology and control

Vectors of *T. cruzi* are insect belonging to a well characterized subfamily of the Reduviidae: the *Triatominae*. This subfamily is further divided in 6 tribes (*Triatomini, Rhodniini, Cavernicoloni, Bolboderini, Alberproseniini, and Linschosteini*), containing 19 genera and 138 species (Galvão *et al.*, 2003). Two tribes only, *Triatomini* and *Rhodniini* are of medical importance. Most species are found between the parallels 45°S and 40°N, and at altitudes up to 2000 meters above sea level. They are prevalent in areas between the tropics (WHO, 2002a).

A hallmark of these insects is that they are obligately hematophagous. As such, all nymphal stages and both sexes of the adults require blood meals for development and reproduction. In its natural habitat, triatomes acquire their blood meal from surrounding vertebrates. Depending on the ecology of the species, three main patterns are distinguishable:

- sylvatic species, which feed on wild animals;
- domestic species, which feed on humans and domesticated animals; and
- sylvatic species in transition to becoming domestic.

From an epidemiological perspective, domestic species are the most relevant due to their close association with humans and because they are highly anthropophilic. These include *Triatoma infestans, Triatoma brasiliensis* and

Panstrongylus megistus in South America and *Rhodnius prolixus* and *Triatoma dimidiata* in Central America. *R. prolixus* is the main vector of Chagas disease in Venezuela, Colombia and limited areas of Central America, whereas *T. infestans* is the most widespread vector of Chagas disease in South America.

Triatoma infestans is a strictly domesticated species or, exceptionally, found in wild ecotopes. This species is a native of Bolivia, the only country where its existence in the sylvatic environment has been proven (Noireau *et al.*, 2005, Cortez *et al.*, 2007).

Rhodnius prolixus is morphologically very similar to *R. robustus*; but *R. robustus* appears to be entirely sylvatic, while *R. prolixus* is a domestic species and is rarely found in the wild. *Rhodnius prolixus* has become the model system in insect physiology despite having a longer life cycle and lower reproductive rate than other medically-relevant arthropods (Wigglesworth, 1984). Depending on blood meal frequency and ambient temperature, *R. prolixus* generally develops from egg to adult in 3-4 months whereas other species can take up to 12 months (Pennington *et al.*, 2005).

Given the major socioeconomic burden of Chagas and the omnipresence of its many vector species, Latin-American governments have recognized the importance of vector control (Ramsey and Schofield, 2003). Initially, in 1991, seven South American countries (Argentina, Bolivia, Brazil, Chile, Paraguay, Uruguay, and later Peru) formed the Southern Cone Initiative to combat Chagas disease. The rationale for this initiative was the idea that an international agreement would encourage the consistency and continuity of national programs for disease control (Dias *et al.* 2002). The main objectives of this initiative were:

- to eliminate vectorial transmission by eliminating vectors from domiciliary and peri-domiciliary areas.
- to reduce disease transmission through blood transfusion by a systematic screening of blood donors.

In 1997, Central American and the Andean Pact countries agreed to launch their own Chagas control programs. Vector control efforts in all of these programs focus on the application of insecticides; because of their cost, house improvements to decrease bug infestations are used minimally. Blood bank screening has had a very positive impact in limiting transmission of the disease through blood transfusion in the disease-endemic region (Schofield and Dujardin, 1997).

Vector control implementation is based on three main strategies: use of chemical insecticides, community health education and rural housing improvements, but insecticide spraying is by far the most commonly used method to combat vectorial transmission.

Since the early stages of vector control programs, scientists and health authorities have tried to use different methods to clear infested houses (spraying kerosene or boiling water over the walls and even a limited trial of military flame-throwers). A major breakthrough for vector control took place in the 1940s with the discovery of synthetic insecticides. Although DDT was very ineffective in combating Chagas disease vectors, organochlorides gave satisfactory results at relatively high doses. Despite leaving an unpleasant smell, spraying organochlorides over the walls of infested houses became the best method of vector control. In the early 1980s synthetic pyrethroid insecticides were introduced marking another improvement in vector control. These insecticides need fewer doses and less frequent applications, thus making them more cost effective. From that perspective, they constitute the preferred tool to eliminate triatominae insects (WHO, 2002a).

From a strategic standpoint, Chagas disease control programs still rely heavily on massive spraying of pyrethroid insecticides. While this constitutes an important first step to reduce insect vector populations, it is not a good long-term strategy.

Scientists recognize that constant exposure to the insecticide will eventually create resistance. This has been the case in *Anopheles gambiae* (malaria vectors) populations in Sub Saharan Africa, and it is probably just a matter of time before this phenomenon occurs in Latin American triatominae insects.

Additionally, in order to maintain its activity, spraying with pyrethroid insecticides must be done on average every six months. For poorer countries (i.e. Bolivia) , this constitutes a serious problem, as it is impossible to spray the estimated 700,000 households located on endemic areas on such a frequent basis because the country is not in the position to assure the financial continuity of such a strategy, which highlights another major challenge for vector control programs.

From an ecological stand point; insecticide spraying may fail to eliminate peridomestic species but definitely does not affect sylvatic species. This is a major concern as re-infestation occurs soon after spraying. For instance, in Central America where the main vector, *T. dimidiata,* retains sylvatic and peridomestic populations in many areas, the vectors gradually re-colonize houses from which the original domestic population has been eliminated. Similar predictions have been made throughout the endemic areas.

Chagas disease control programs in the Americas have had very positive results since their inception and implementation, but it would be premature to assume the disease has been defeated. Public health agencies and governments in the Americas must maintain political pressure and commit financial resources to long term programs. Furthermore, research into novel strategies of vectorial control, surveillance and treatment should be encouraged.

Triatome - *T. cruzi* interactions

The transmission of *T. cruzi,* and most other vector-borne parasites, is based on the evolution of blood feeding as a strategy to obtain nutrients. Hematophagy also is a tremendous opportunity for parasitic organisms such as *T. cruzi* to be transmitted to a large variety of potential vertebrate hosts, which is

beneficial to its own survival and dispersal; the parasite depends completely on the insect vector for its transmission and dissemination.

Trypanosoma cruzi's life cycle in the invertebrate host starts with the ingestion of trypomastigotes, found in the blood of the vertebrate host. In the stomach, the long trypomastigotes transform into rounded flagellates known as spheromastigotes. These migrate to the intestine and transform into short and long epimastigotes. Epimastigotes multiply in the intestine, and then transform into infective trypomastigotes. Trypomastigotes are infective forms that are released in the feces and urine of the insect during acquisition of a blood meal from vertebrates (Figure 1-1). Compared to other insect borne pathogens, *T. cruzi* uses an inefficient way of transmission to its vertebrate host, as blood feeding is not enough to infect the vertebrate. After being deposited with the insect's feces, *T. cruzi* must still find a way to penetrate the skin, which it is naturally unable to do and thus requires the participation of the vertebrate host. By scratching the bite site the host can allow the parasite to come into direct contact with the blood stream and ensure continuation of its life cycle. This is in sharp contrast with *Plasmodium* spp., yellow and dengue viruses, and *T. brucei*, all of which are transmitted via the salivary glands upon blood feeding of their respective insect vectors. Another trypanosome, *Trypanosoma rangeli* is transmitted via saliva when Triatomes blood-feed. *T. rangeli* is capable of crossing the midgut epithelium, surviving in the hemolymph and infecting the salivary glands to be transmitted during the next blood meal (Azambuja and Garcia, 2005). Interestingly, *T. rangeli* is not a human pathogen, but it is pathogenic to the insect vector. One can't help wondering what are the genetic differences that account for such different life cycles and why *T. cruzi* has not evolved a salivary transmission?

Understanding the fine details of trypanosome-vector dynamics that determine host susceptibility is a challenging task. Many factors are involved at the biochemical and physiological level; including the nutritional state of the vector,

digestive enzymes, bacterial symbionts and the immune system. Upon blood feeding, the insect stores the blood bolus in the stomach (anterior midgut) and then slowly passes fractions into the intestine (posterior midgut) where acid proteolysis via different cathepsins occurs to digest blood. This is a potentially very hostile environment for the parasite; however addition of pepstatin (acid proteinase inhibitor) had no effect on *T. cruzi* infection rates (Garcia and Gilliam, 1980). Additionally, it appears that different strains of *T. cruzi* exhibit differential survival rates; a crop lytic factor can lyse strain Y but not Dm28c or CL which agglutinate and seem to be protected (Mello *et al.*, 1996).

In terms of the role the immune system of the host plays, artificial injection of *T. cruzi* into the hemocoel of potential vectors, showed clearing of the parasite within a few days, and lysozyme like and phenoloxidase activity were detected (Mello *et al.*, 1995, Azambuja *et al.*, 1999). More recently, a spatial and temporal modulation of nitric oxide and nitric oxide synthase activity was reported in response to injection of *T. cruzi*, *T. rangeli* and LPS (Whitten *et al.*, 2007). During a normal infection via a blood meal; physical interaction between the parasite and the insect midgut epithelia has been documented (Alves *et al.*, 2007) and are required for the normal development of the parasite. A local immune response is likely to occur although no clear evidence of this has been shown.

Insect immunity

The field of insect immunology was pioneered by Louis Pasteur himself in the mid-nineteenth century when he was commissioned by the French government to examine "pébrine", a disease decimating silkworms in Southern France. From 1853 to 1865, France's silk production plummeted from 26,000 tons per year to only 4,000 because of this insect disease caused by, at the time unknown, microsporidian protozoan pathogen *Nosema bombysis*. Infected silkworm larvae died before spinning their cocoons and were covered by black spots throughout their bodies. Pasteur discovered that he could also induce the black spots by damaging the cuticle thus discovering melanization. Pasteur's

work on the silkworm disease established the scientific mindset for the studies later on carried out by his collaborators and successors at the Pasteur Institute in Paris in invertebrate and vertebrate immunology (Brey, 1988). Most of the work done there focused mainly on cellular and vertebrate adaptive immunity.

From a molecular standpoint, insect immunology has received a great deal of attention in the past 25 years both from basic as well as applied researchers. Comparative studies in immunology between invertebrates and vertebrates have revealed the lack of an antibody mediated response in insects, a hallmark of higher organisms. Thus, invertebrates only have an innate immune response which still is very efficient in eliminating foreign bodies (non-self). Invading organisms generally must overcome its host innate immune system at three different levels: initially, the physical barriers of the integument or gut epithelia; secondly, cellular defense mechanisms mediated by hemocytes; and finally humoral components typified by antimicrobial peptides (AMPs).

The first level of defense is normally encountered at the gut epithelial level via interaction of the pathogen with extracellular adhesion molecules and other mechanisms not well understood. Such challenge tends to trigger a localized, relatively weak, epithelial response where AMPs are secreted into the gut lumen. Once the invading microorganism reaches the insect's hemocoel, pathogen specific molecular patterns are bound by soluble pathogen-recognition molecules present in the hemolymph which labels them as non-self.

The activation of the immune system relies on the very basic ability of recognizing and discriminating self from non-self; this takes place when a set of highly conserved molecular patterns produced by microorganisms which are absent from host cells are recognized as non-self by a set of host receptors (Medzhitov and Janeway, 2002). Microbe-associated molecular patterns (MAMPs), also known as pathogen-associated molecular patterns (PAMPs), are molecular motifs characteristic of certain organisms which are recognized by the

16

insect immune response through a series of soluble pattern recognition receptors (PRRs) which are constitutively expressed and omnipresent in the hemolymph or as transmembrane cellular receptors. These not only allow the host to recognize the foreign microorganism but also to discriminate among different microorganisms. PAMPs include bacterial lipopolysaccharide (LPS), peptidoglycan (PGN) and lipoteichoic acid (LTA); fungal β1,3-glucans; and in mammals: nucleic acids (unmethylated CpG, double stranded RNA). In invertebrates, microbial septic infection is detected via PRRs from two distinct families:

- Gram-negative binding proteins (GNBPs)/beta-glucan recognition proteins (ßGRPs); characterized by two distinct domains, an N-terminal glucan-binding domain and a C-terminal domain similar to that of bacterial glucanases. GNBPs are present in most invertebrates, but have not been found in vertebrates (Kim *et al.,* 2000).
- Peptidoglycan recognition proteins (PGRPs); comprised of 13 members in *Drosophila melanogaster*, recognize the inner layer of PGN characteristic of Gram negative bacteria and contain a conserved domain related to the bacteriophage II amidase; activity that some members still retain (Werner *et al.,* 2000).

Non-self recognition by PRRs rapidly results in a cellular and humoral response. Studies in cellular mediated immune response were pioneered by Élie Metchnikoff (1845-1916), a Russian zoologist and microbiologist who worked at the Institut Pasteur. He was awarded in 1908 the Nobel Prize in physiology and medicine (along with Paul Ehrlich) for his discovery of phagocytes as the basis of cell-mediated immunity.

An insect's cellular immunity is assumed by hemocytes, the equivalent of our blood cells, which engage in different defense mechanisms depending on their lineage (Strand, 2008). The literature is heterogeneous in the nomenclature describing these cells, and most work has been done in *Drosophila*

melanogaster, some Lepidoptera and mosquitoes. In *D. melanogaster*, three lineages have been characterized based on their morphology, function and molecular markers: plasmatocytes (phagocytosis), lamellocytes (encapsulation) and crystal cells (melanization). Lepidopteran hemocytes are morphologically and functionally similar but named granulocytes, plasmatocytes and oenocytoids respectively.

Plasmatocytes are specialized for phagocytosis, the uptake of particulate materials by engulfment, and are the most abundant type of hemocytes (>90%). Crystal cells are much less abundant (5%) hemocytes, responsible of initiating the phenoloxidase (PO) cascade which leads to melanization. Melanization refers to the deposition of melanin, a dark pigment resulting from the polymerization of polyphenols at the site of injury and around the pathogen to limit its movement and proliferation. Activation of the PO cascade also releases toxic reactive oxygen species (ROS) that may harm invading microbes. Lamellocytes are large, flat, adhesive hemocytes (to date found only in *Drosophila* larvae) specialized for encapsulation of parasitoids and other large foreign microorganisms which are unable to be eliminated by the previously described mechanisms (Strand, 2008). In addition to their role in cellular immunity, hemocytes are also known to participate in the humoral response as they are capable to produce and secrete antimicrobial peptides (AMPs) into the hemolymph.

Humoral defenses are characterized by a rapid, *de novo* production of a large number of potent antimicrobial peptides, reactive intermediates of nitrogen or oxygen and a complex enzymatic cascade yielding in clotting or hemolymph melanization (Hoffmann, 1995, Lowenberger, 2001). Insects have an open circulatory system, where tissues are bathed in a nutrient-rich liquid termed hemolymph thus maintaining it free of foreign organism is of foremost importance for the insect. The induction of the humoral response in *D. melanogaster* is regulated by two well characterized cellular pathways, termed Toll and IMD.

Induction of these two pathways leads to the transcriptional activation of AMPs (Figure 1-2).

AMPs are evolutionary conserved members of the immune system and they are found in a diversity of organisms including prokaryotes, invertebrates, vertebrates and plants. Because of their importance to insect immunity and their omnipresence in organisms, AMPs will be a major subject in this dissertation. The first inducible AMP to be described was isolated from hemolymph of the lepidopteran *Hyalophora cecropia* after bacterial challenge (Steiner *et al.*, 1981). Since that discovery, over 500 different AMPs have been described, most of which tend to have similar characteristics. AMPs are peptides or polypeptides under 100 amino acid residues long, with sizes ranging from 2 to 25kDa. Their secondary structure follow four different conformations: i) linear ii) α-helical, iii) β-hairpin or loop (single disulfide bond) and iv) β-stranded (presence of two or more disulfide bonds) (Boman, 2003, Bulet and Stocklin, 2005). Despite their numbers and size and structure polymorphism, AMPs have been categorized in only a small number of families. In humans, only three families exist, defensins, cathelicidins and histatins (De Smet and Contreras, 2005). Surprisingly, insects have a larger variety of AMPs. In *Drosophila*, AMPs have been classified into eight different classes, depending on their structure and the organism against which they are active. AMPs have mostly been shown to be active against three types of organisms, Gram positive and negative bacteria and yeast (Imler and Bulet, 2005) but have also been shown to be active against larger protozoan parasites (Boulanger *et al.*, 2006). These peptides are produced by immune tissues, primarily the fat body but also by the midgut epithelium and hemocytes. In the hemolymph, their concentrations can range from 1-100μM (antifungal Drosomycin) (reviewed in Lemaitre and Hoffmann, 2007)

In addition to AMPs, reactive intermediates of oxygen (ROI) and nitrogen (RNI) are also present in the hemolymph and contribute to the immune response. Evidence from various vertebrate and invertebrate models indicates that certain

ROI and RNI, in particular radical O_2-, H_2O_2, and radical NO, are cytotoxic molecules induced upon immune challenge, but also important cell-signaling molecules (Nappi and Christensen, 2005). These molecules have also been shown to have detrimental effects against parasites; in *Anopheles* mosquitoes in response to *Plasmodium* infections and in trypanosome-infected *R. prolixus* (Whitten *et al.*, 2001).

An emerging field of study in insect immunology and vector competence is the role of programmed cell death (apoptosis) in response to viral and parasitic infections. Apoptosis is a keystone cellular signaling pathway involved in tissue homeostasis and development, and DNA damage responses, but also in a variety of pathological processes. In lepidopteran insects, apoptosis has been established as a component of the innate immune response against baculovirus infections (Clem, 2005). The immune role of apoptosis in insects is strengthened by evidence of cross-talk between innate immunity pathways and apoptotic pathways. In mosquitoes, apoptosis occurs during *Plasmodium* and arbovirus infection in the midgut, suggesting that apoptosis plays a role in mosquito innate immunity as well; and research is underway to identify and characterize the different genes involved in this major vector of disease (Cooper *et al.*, 2007a, Cooper *et al.*, 2007b, Bryant *et al.*, 2008).

Although our knowledge of invertebrate immunity has greatly increased in the past 25 years, insect innate immunity research has mainly focused on a relatively narrow, higher order of insects: mostly Diptera (*Drosophila* and mosquitoes) and Lepidoptera. Studies in more ancient insects such as hemimetabolous hemipterans (true bugs) are comparatively scarce. Based on the current phylogenetic hypothesis for arthropods (Wheeler *et al.*, 2001), hemipterans are the most primitive hematophagous insects. Given the heterogeneity of invertebrates, identifying components of the innate immune response in hemipteran insects provides an invaluable evolutionary view of immunity. Triatome insects and their trypanosomatid parasites provide, despite limited

genetic tools available, an excellent case study because of the fundamental differences in their life cycles and their impact for the immune system of the insect.

As vector biologists, we hope that studying the vector's immune system and vector–parasite interactions will make possible the discovery of the biological determinants that permit such interactions, with the ultimate goal of using this knowledge to limit, or ideally stop the transmission of human pathogens. Several research groups worldwide study the molecular interactions between human pathogens and their insect hosts with the hope of developing vaccines that block parasite development and transmission or molecules that affect vector immunity, or to genetically modify vector competence and capacity. When the work presented in this thesis started, studies between the interactions of *T. cruzi* and triatominae were scarce compared to other insect/parasite pairs. If *T. cruzi* is injected into the hemocoel of *R. prolixus,* the parasite dies and cannot be recovered after 3 days (Azambuja *et al.,* 1999). Only one previous study attempted to reveal the interactions occurring in vivo between *T. cruzi* and the intestine of the vectors using electron microscopy (Kollien *et al.,* 1998) and which disclosed probable ultra-structural interactions. Different regions of the intestine of insects vary in their nutritional, surface conditions and the genes that are expressed there which may affect *T. cruzi* development. Other studies on a closely related organism, *Trypanosoma rangeli* (which is able to cross the midgut epithelia and survive in the hemolymph), suggested that this parasite has the capacity of disabling the pro-phenoloxidase pathway that normally leads to melanization, and avoids the immune system by infecting hemocytes (Azambuja and Garcia, 2005).

Goal and approach

The objective of my doctoral studies is to test the hypothesis that the protozoan parasite, *T. cruzi*, maintains a restricted life cycle in the midgut and hindgut of *Rhodnius prolixus* in order to limit its exposure to components of the

host innate immune system. To address this question, we have identified and described the genes involved in the innate immune response of hemipteran *R. prolixus,* an important vector of Chagas disease.

Of the thousands of genes in a genome, only a small subset are expressed at any given time and in any given cell type. Furthermore, the cellular concentration of a given protein is determined by a delicate balance of several cellular processes:

1. Transcription : synthesis of the primary RNA transcript.
2. Posttranscriptional modification of mRNA
3. Messenger RNA degradation
4. Translation: protein synthesis
5. Posttranslational modification of proteins
6. Protein targeting and transport
7. Protein degradation

In eukaryotes, gene regulation primarily takes place at the transcriptional level (although new discoveries made in the field of small RNAs could change that belief). The ability of selectively controlling the induction of genes is termed differential gene expression which controls a wide rage of essential biological processes incluiding the immune response (Harshman and James, 1998). Using differential gene expression, we have determined the repertoire of immune peptides used by this insect vector, determined the factors that induce and regulate the immune response, and evaluated the role of these molecules during normal infection of the insect by *T. cruzi.* The results were discussed in connection with studies on innate immunity in three other major orders of insects: the Lepidoptera, Diptera, and Hymenoptera.

This doctoral dissertation presents the findings of a functional genomics approach using Suppressive Subtractive Hybridization (SSH) to identify up-

regulated genes in response to different immune challenges in two different tissues (gut and fat body) of *R. prolixus*. Following the initial molecular screening, we molecularly characterized individual genes, with a strong emphasis on effector molecules (i.e. antimicrobial peptides) and a transcriptional regulator.

The present dissertation is organized in seven stand alone chapters, written in publishable format. I start with a general introduction, followed by the molecular screenings performed, continued by the molecular characterization of an evolutionary conserved transcription factor of the Rel/NFκB family of proteins. Subsequently, three different antimicrobial peptides are characterized. A general conclusion summarizes the main results and discusses their implications.

Figures
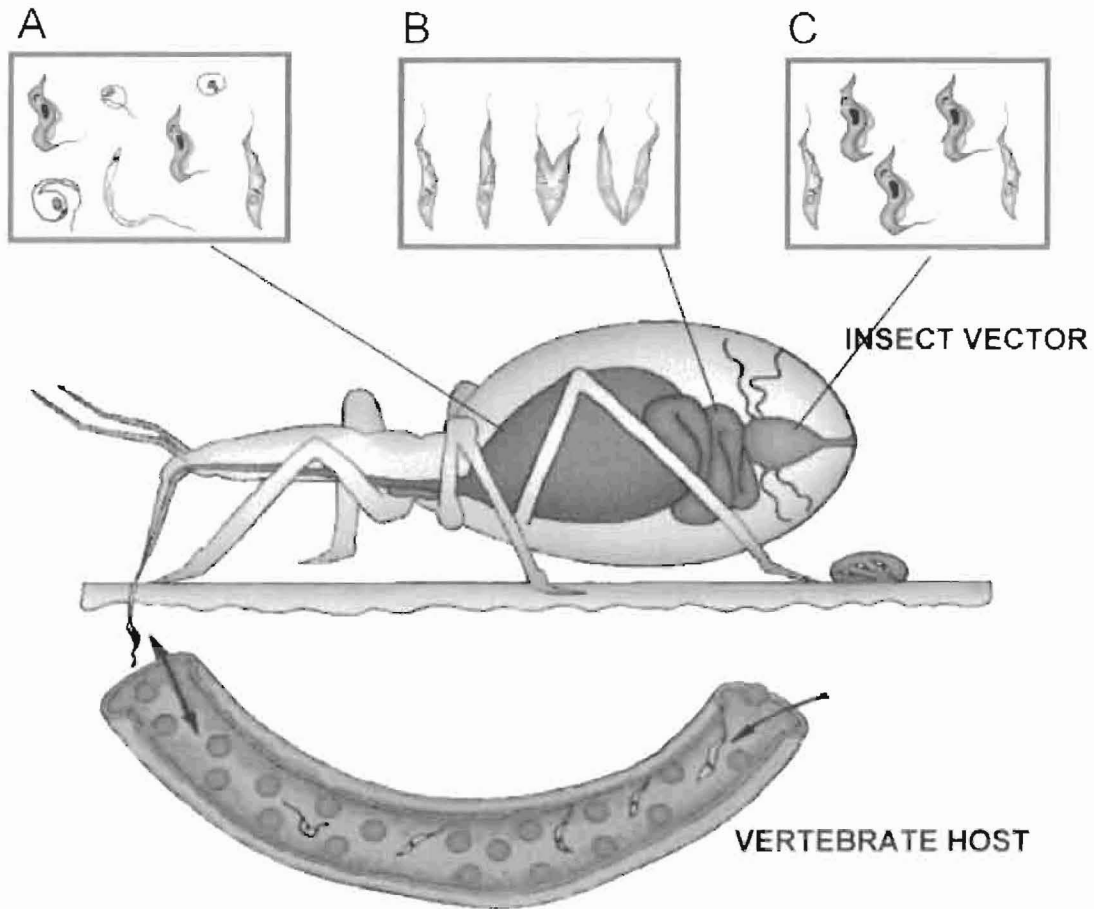
Figure 1-1 *Trypanosoma cruzi* life cycle in the invertebrate vector.

The insect feeds on blood infected with trypomastigote forms which transform into epimastigotes and some spheromastigotes in the stomach (A). In the intestine, the epimastigotes multiply (B) increasing the population of parasites. In the rectum, the epimastigotes transform into metacyclic trypomastigotes (C) which are eliminated with the feces and urine. Reprinted, with permission, from Journal of Insect Physiology, Volume 53 ©2007 by Elsevier Limited (Garcia *et al.*, 2007).
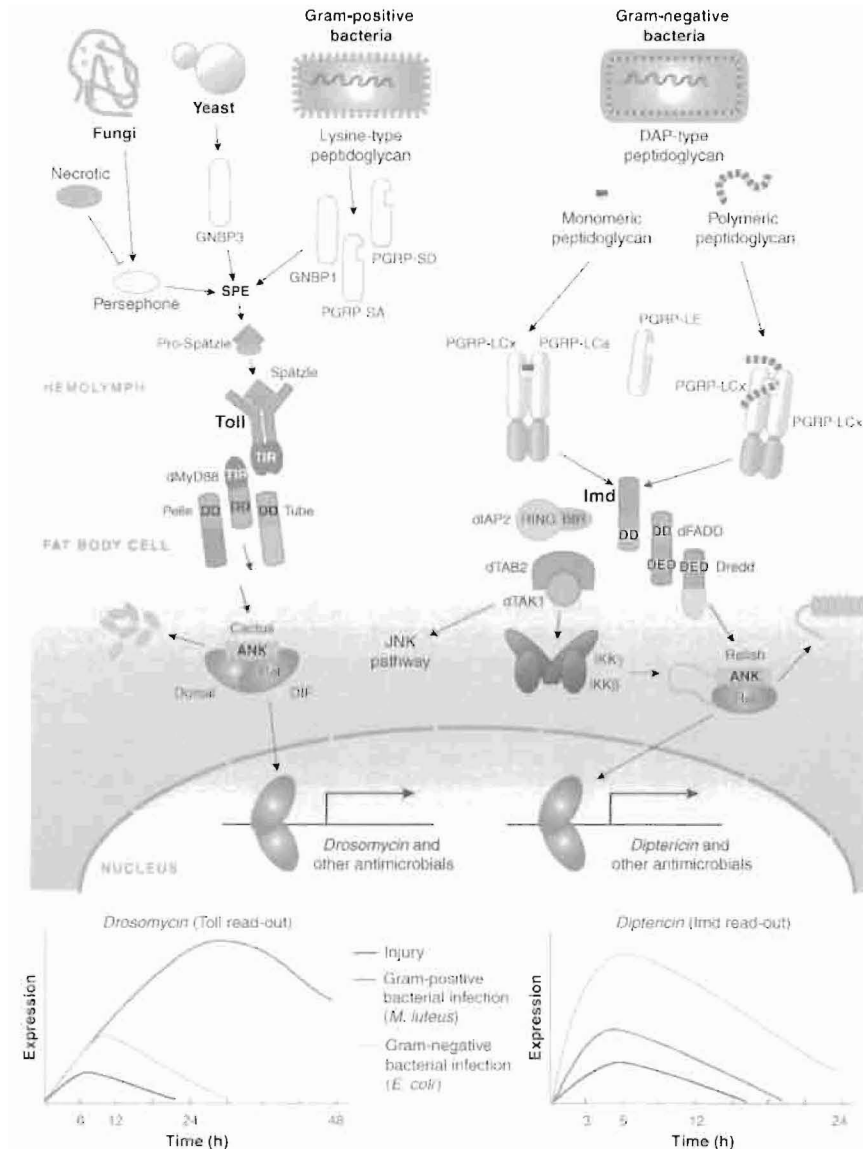
Figure 1-2 *Drosophila melanogaster*'s Toll and Imd immune pathway activation.

Antimicrobial peptide genes are regulated by a balance between two signaling pathways: the Toll pathway that is largely activated by fungi and Gram-positive bacteria, and the Imd pathway that is mainly activated by Gram-negative bacteria. According to the κB sites present in their promoters, antimicrobial peptide genes are more sensitive to either the Toll cascade (e.g., Drosomycin) or the Imd cascade (e.g., Diptericin) or are co-regulated. Reprinted, with permission, from the Annual Review of Immunology, Volume 25 ©2007 by Annual Reviews (Lemaitre and Hoffmann, 2007).

References

Alves, C.R., Albuquerque-Cunha, J.M., Mello, C.B., Garcia, E.S., Nogueira, N.F., Bourguingnon, S.C., De Souza, W., Azambuja, P. & Gonzalez, M.S., 2007. *Trypanosoma cruzi*: attachment to perimicrovillar membrane glycoproteins of *Rhodnius prolixus*. *Experimental parasitology,* 116, 44-52.

Avila, H.A., Sigman, D.S., Cohen, L.M., Millikan, R.C. & Simpson, L., 1991. Polymerase chain reaction amplification of *Trypanosoma cruzi* kinetoplast minicircle DNA isolated from whole blood lysates: diagnosis of chronic Chagas' disease. *Molecular and biochemical parasitology,* 48, 211-21.

Azambuja, P., Feder, D., Mello, C., Gomes, S. & Garcia, E., 1999. Immunity in *Rhodnius prolixus*: trypanosomatid-vector interactions. *Memorias do Instituto Oswaldo Cruz,* 94 Suppl 1, 219-22.

Azambuja, P. & Garcia, E.S., 2005. *Trypanosoma rangeli* interactions within the vector *Rhodnius prolixus*: a mini review. *Memorias do Instituto Oswaldo Cruz,* 100, 567-72.

Barrett, M.P., Burchmore, R.J., Stich, A., Lazzari, J.O., Frasch, A.C., Cazzulo, J.J. & Krishna, S., 2003. The trypanosomiases. *Lancet,* 362, 1469-80.

Boman, H.G., 2003. Antibacterial peptides: basic facts and emerging concepts. *Journal of internal medicine,* 254, 197-215.

Boulanger, N., Bulet, P. & Lowenberger, C., 2006. Antimicrobial peptides in the interactions between insects and flagellate parasites. *Trends in parasitology,* 22, 262-8.

Brey, P.T., 1988. The contributions of the Pasteur school of insect immunity. *In* P.T. Brey (ed.) *Molecular mechanisms of immune responses in insects.* 1st ed. London; New York, 1-39.

Bryant, B., Blair, C.D., Olson, K.E. & Clem, R.J., 2008. Annotation and expression profiling of apoptosis-related genes in the yellow fever mosquito, *Aedes aegypti. Insect biochemistry and molecular biology,* 38, 331-45.

Bulet, P. & Stocklin, R., 2005. Insect antimicrobial peptides: structures, properties and gene regulation. *Protein and peptide letters,* 12, 3-11.

Buscaglia, C.A., Campo, V.A., Frasch, A.C. & Di Noia, J.M., 2006. *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. *Nature reviews,* 4, 229-36.

Buscaglia, C.A. & Di Noia, J.M., 2003. *Trypanosoma cruzi* clonal diversity and the epidemiology of Chagas' disease. *Microbes and infection,* 5, 419-27.

Chagas, C., 1909. Nova tripanozomíase humana; estudo sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi* n. gen. n. sp., agente etiológico de nova entidade mórbida do homem. *Memorias do Instituto Oswaldo Cruz,* 1, 159-218.

Clem, R.J., 2005. The role of apoptosis in defense against baculovirus infection in insects. *Current topics in microbiology and immunology,* 289, 113-29.

Cooper, D.M., Pio, F., Thi, E.P., Theilmann, D. & Lowenberger, C., 2007a. Characterization of *Aedes* Dredd: a novel initiator caspase from the yellow fever mosquito, *Aedes aegypti. Insect biochemistry and molecular biology,* 37, 559-69.

Cooper, D.M., Thi, E.P., Chamberlain, C.M., Pio, F. & Lowenberger, C., 2007b. Aedes Dronc: a novel ecdysone-inducible caspase in the yellow fever mosquito, *Aedes aegypti. Insect molecular biology,* 16, 563-72.

Cortez, M.R., Emperaire, L., Piccinali, R.V., Gurtler, R.E., Torrico, F., Jansen, A.M. & Noireau, F., 2007. Sylvatic *Triatoma infestans* (Reduviidae, Triatominae) in the Andean valleys of Bolivia. *Acta tropica,* 102, 47-54.

Croft, S.L., Barrett, M.P. & Urbina, J.A., 2005. Chemotherapy of trypanosomiases and leishmaniasis. *Trends in parasitology,* 21, 508-12.

De Smet, K. & Contreras, R., 2005. Human antimicrobial peptides: defensins, cathelicidins and histatins. *Biotechnology letters,* 27, 1337-47.

Dias, J.C., Silveira, A.C. & Schofield, C.J., 2002. The impact of Chagas disease control in Latin America: a review. *Memorias do Instituto Oswaldo Cruz,* 97, 603-12.

Dutra, W.O., Rocha, M.O. & Teixeira, M.M., 2005. The clinical immunology of human Chagas disease. *Trends in parasitology,* 21, 581-7.

El-Sayed, N.M., Myler, P.J., Bartholomeu, D.C., Nilsson, D., Aggarwal, G., Tran, A.N., Ghedin, E., Worthey, E.A., Delcher, A.L., Blandin, G., Westenberger, S.J., Caler, E., Cerqueira, G.C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D.A., Carrington, M., Crabtree, J., Darban, H., Da Silveira, J.F., De Jong, P., Edwards, K., Englund, P.T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A.C., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M.J., Lorenzi, H., Louie, T., Machado, C.R., Mcculloch, R., Mckenna, A., Mizuno, Y., Mottram, J.C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J.L., Rinta, J., Robertson, L., Salzberg, S.L., Sanchez, D.O., Seyler, A., Sharma, R., Shetty, J., Simpson, A.J., Sisk, E., Tammi, M.T., Tarleton, R., Teixeira, S., Van Aken, S., Vogt, C., Ward, P.N., Wickstead, B., Wortman, J., White, O., Fraser, C.M., Stuart, K.D. & Andersson, B., 2005. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science*, 309, 409-15.

Fenn, K. & Matthews, K.R., 2007. The cell biology of *Trypanosoma brucei* differentiation. *Current opinion in microbiology*, 10, 539-46.

Galvão, C., Carcavallo, R., Rocha, D.S. & Jubert, J., 2003. A checklist of the current valid species of the subfamily Triatominae Jeannel, 1919 (Hemíptera, Rediviidae) and their geographical distribution, with nomenclatural and taxonomic notes. *Zootaxa*, 1-36.

Garcia, E.S. & Gilliam, F.C., 1980. *Trypanosoma cruzi* development is independent of protein digestion in the gut of *Rhodnius prolixus*. *The Journal of parasitology*, 66, 1052-3.

Garcia, E.S., Ratcliffe, N.A., Whitten, M.M., Gonzalez, M.S. & Azambuja, P., 2007. Exploring the role of insect host factors in the dynamics of *Trypanosoma cruzi-Rhodnius prolixus* interactions. *Journal of insect physiology*, 53, 11-21.

Harshman, L.G. & James, A.A., 1998. Differential gene expression in insects: transcriptional control. *Annual review of entomology*, 43, 671-700.

Hoffmann, J.A., 1995. Innate immunity of insects. *Current opinion in immunology*, 7, 4-10.

Imler, J.L. & Bulet, P., 2005. Antimicrobial peptides in *Drosophila*: structures, activities and gene regulation. *Chemical immunology and allergy*, 86, 1-21.

Kim, Y.S., Ryu, J.H., Han, S.J., Choi, K.H., Nam, K.B., Jang, I.H., Lemaitre, B., Brey, P.T. & Lee, W.J., 2000. Gram-negative bacteria-binding protein, a pattern recognition receptor for lipopolysaccharide and beta-1,3-glucan that mediates the signaling for the induction of innate immune genes in *Drosophila melanogaster* cells. *The Journal of biological chemistry*, 275, 32721-7.

Kollien, A.H., Schmidt, J. & Schaub, G.A., 1998. Modes of association of *Trypanosoma cruzi* with the intestinal tract of the vector *Triatoma infestans. Acta tropica,* 70, 127-41.

Lemaitre, B. & Hoffmann, J., 2007. The host defense of *Drosophila melanogaster. Annual review of immunology*, 25, 697-743.

Lowenberger, C.A., 2001. Form, function and phylogenetic relationships of mosquito immune peptides. *Advances in experimental medicine and biology,* 484, 113-29.

Marquardt, W.C., Grieve, R.B., Demaree, R.S. & Marquardt, W.C., 2000. *Parasitology and vector biology,* 2nd ed. San Diego, Calif.; London: Harcourt Academic.

Medzhitov, R. & Janeway, C.A., Jr., 2002. Decoding the patterns of self and nonself by the innate immune system. *Science*, 296, 298-300.

Mello, C.B., Azambuja, P., Garcia, E.S. & Ratcliffe, N.A., 1996. Differential in vitro and in vivo behavior of three strains of *Trypanosoma cruzi* in the gut and hemolymph of *Rhodnius prolixus. Experimental parasitology,* 82, 112-21.

Mello, C.B., Garcia, E.S., Ratcliffe, N.A. & Azambuja, P., 1995. *Trypanosoma cruzi* and *Trypanosoma rangeli:* interplay with hemolymph components of *Rhodnius prolixus. Journal of invertebrate pathology,* 65, 261-8.

Momen, H., 1999. Taxonomy of *Trypanosoma cruzi:* a commentary on characterization and nomenclature. *Memorias do Instituto Oswaldo Cruz,* 94 Suppl 1, 181-4.

Moncayo, A., 1999. Progress towards interruption of transmission of Chagas disease. *Memorias do Instituto Oswaldo Cruz,* 94 Suppl 1, 401-4.

Nappi, A.J. & Christensen, B.M., 2005. Melanogenesis and associated cytotoxic reactions: applications to insect innate immunity. *Insect biochemistry and molecular biology,* 35, 443-59.

Noireau, F., Cortez, M.G., Monteiro, F.A., Jansen, A.M. & Torrico, F., 2005. Can wild *Triatoma infestans* foci in Bolivia jeopardize Chagas disease control efforts? *Trends in parasitology,* 21, 7-10.

Pennington, P.M., Beard, C.B. & Anderson, J., 2005. Care, maintenance, and handling of infected Triatomines. *In* W.C. Marquardt (ed.) *Biology of Disease Vectors.* 2nd ed. Amsterdam: Elsevier Academic Press, 717-725.

Ramsey, J.M. & Schofield, C.J., 2003. Control of Chagas disease vectors. *Salud publica de Mexico,* 45, 123-8.

Rodriques Coura, J. & De Castro, S.L., 2002. A critical review on Chagas disease chemotherapy. *Memorias do Instituto Oswaldo Cruz, 97,* 3-24.

Schofield, C.J. & Dujardin, J.P., 1997. Chagas disease vector control in Central America. *Parasitology today,* 13, 141-4.

Steiner, H., Hultmark, D., Engstrom, A., Bennich, H. & Boman, H.G., 1981. Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature,* 292, 246-8.

Strand, M.R., 2008. The insect cellular immune response. *Insect Science,* 15, 1-14.

Tibayrenc, M., 1995. Population genetics of parasitic protozoa and other microorganisms. *Advances in parasitology,* 36, 47-115.

Tibayrenc, M., Kjellberg, F. & Ayala, F.J., 1990. A clonal theory of parasitic protozoa: the population structures of *Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas,* and *Trypanosoma* and their medical and taxonomical consequences. *Proceedings of the National Academy of Sciences of the United States of America,* 87, 2414-8.

Werner, T., Liu, G., Kang, D., Ekengren, S., Steiner, H. & Hultmark, D., 2000. A family of peptidoglycan recognition proteins in the fruit fly *Drosophila melanogaster. Proceedings of the National Academy of Sciences of the United States of America,* 97, 13772-7.

Wheeler, W.C., Whiting, M., Wheeler, Q.D. & Carpenter, J.M., 2001. The Phylogeny of the Extant Hexapod Orders. *Cladistics,* 17, 113-169.

Whitten, M., Sun, F., Tew, I., Schaub, G., Soukou, C., Nappi, A. & Ratcliffe, N., 2007. Differential modulation of *Rhodnius prolixus* nitric oxide activities following challenge with *Trypanosoma rangeli, T. cruzi* and bacterial cell wall components. *Insect biochemistry and molecular biology,* 37, 440-52.

Whitten, M.M., Mello, C.B., Gomes, S.A., Nigam, Y., Azambuja, P., Garcia, E.S. & Ratcliffe, N.A., 2001. Role of superoxide and reactive nitrogen intermediates in *Rhodnius prolixus* (Reduviidae)/*Trypanosoma rangeli* interactions. *Experimental parasitology,* 98, 44-57.

Who, 1979. *DDT and its derivatives* Geneva: World Health Organization.

Who, 2002a. *Control of Chagas disease: second report of a WHO expert committee* Geneva: World Health Organization.

Who, 2002b. *The world health report 2002: Reducing risks, promoting healthy life.* Geneva: World Health Organization.

Wigglesworth, V.B., 1984. *Insect physiology,* 8th ed. London; New York: Chapman and Hall.

Connecting statement 1

In the introductory chapter, I presented background information for the study of *Rhodnius prolixus* and subsequently described different components of their immune response that insects use to eliminate parasites and pathogens that breach the cuticle or enter the intestinal tract and pass through to the hemocoel. We believe that these innate immune responses, described to date in greatest detail in modern insects such as *Drosophila melanogaster*, are equally important to more ancient organisms such as *Rhodnius prolixus*. Similarly, while many studies have used comparisons with *D. melanogaster* to study immune related genes and ascribe putative function in holometabolous insects such as the Diptera and Lepidoptera, there are few studies that have studied similar phenomena, or studied the global expression of immune related genes, in hemimetabolous insects such as the hemipterans. In the next chapter we address this issue by using suppression subtractive hybridization (SSH) to generate cDNA libraries enriched for differentially expressed genes in response to infection with bacteria or *Trypanosoma cruzi*.

CHAPTER 2: *RHODNIUS PROLIXUS*: IDENTIFICATION OF IMMUNE RELATED GENES UP REGULATED IN RESPONSE TO PATHOGENS AND PARASITES USING SUPPRESSIVE SUBTRACTIVE HYBRIDIZATION

URSIC BEDOYA Raul, LOWENBERGER Carl
Simon Fraser University
Department of Biological Sciences
Burnaby, British Columbia, V5A 1S6, Canada

Abstract

We report the identification of immune-related molecules from the fat body and intestine of *Rhodnius prolixus*, an important vector of Chagas disease. Insects were challenged by introducing pathogens or *Trypanosoma cruzi,* the parasite that causes Chagas disease, into the hemocoel. RNA from intestines, or fat body was isolated 24 h after stimulation. We used suppressive subtractive hybridization to identify immune-related genes; generated three subtracted libraries, sequenced the clones and assembled the sequences. The functional annotation revealed expressed sequence tags (ESTs) generated in response to various stimuli in all tissues, and included pathogen recognition molecules, regulatory molecules, and effector molecules.

Introduction

Rhodnius prolixus (family: Reduviidae) is an important vector of *Trypanosoma cruzi*, a protozoan parasite and etiological agent of American trypanosomiasis (Chagas disease) in Northern-South and Central America. Chagas disease affects an estimated 13 million people in the Americas causing significant morbidity; most acute infections are asymptomatic, yet 25 to 30% of these become chronic, leading to approximately 14,000 deaths annually (Mattock *et al.*, 2005). Currently, there is neither a preventive vaccine nor an effective treatment to cure chronic Chagas disease as the drugs used, based on nitro heterocyclic compounds, have a very limited efficacy in the chronic stage and toxic side effects often lead to treatment cessation.

Transmission of *T. cruzi* is atypical and shares very little with other major insect-borne diseases in which the parasites invade the salivary glands and are injected into the vertebrate as it takes a blood meal. *Trypanosoma cruzi* resides in the intestine/rectum of triatome insects. As the insect engorges, the insect defecates and droplets containing the parasites are deposited on the host's skin and may enter via the bite site or a mucosal membrane. This transmission strategy is inefficient, and we have hypothesized previously that by remaining exclusively in the gut, *T. cruzi* is not exposed directly to the hemolymph which contains the most potent components of the insects' immune response (Lopez *et al.*, 2003). The immune response of insects is innate, lacks the acquired component of vertebrates yet still is very efficient in eliminating pathogens using a combination of humoral and/or cellular defense responses.

The first step in the immune response requires the recognition of parasites as non-self. Insects recognize unique pathogen-associated molecular patterns (PAMPs) that are characteristic of microbial organisms (Nurnberger *et al.*, 2004) using host pattern recognition receptors (PRRs) (Medzhitov and Janeway, 2002). The two major PRRs in insects are the peptidoglycan recognition proteins

(PGRPs) and the Gram-negative bacteria binding proteins (GNBPs) (Osta *et al.*, 2004). Once specific PRRs are activated by the appropriate PAMP, signaling cascades are initiated. Surface molecules present on Gram-negative bacteria are PAMPs recognized by the receptors in the IMD pathway which results in the nuclear translocation of Relish (an NF-κB-like transcription factor), and the induction of antimicrobial peptides (AMPs) such as Cecropin, Drosocin, Defensin and Diptericin (Hoffmann and Reichhart, 2002, Meister *et al.*, 2005). In *Drosophila melanogaster*, challenge with fungi and Gram-positive bacteria activates the Toll pathway, which results in the NF-κB-like transcription factor, Dif, being translocated to induce expression of Drosomycin. This activation process also triggers various other proteolytic cascades, including melanization and coagulation, in which serine proteases and serpins are involved (Osta *et al.*, 2004) and cellular-mediated mechanisms including phagocytosis, nodulation, and encapsulation by hemocytes (Lavine and Strand, 2002). This insect immune system is very efficient and large numbers of bacteria can be removed within minutes of entry into the hemocoel (Hillyer *et al.*, 2003). In addition, the humoral response can contribute to the release of reactive intermediates of nitrogen or oxygen (Nappi *et al.*, 2000) all of which can contribute to the removal of parasites.

Insect innate immunity against protozoan parasites has been studied mostly in mosquitoes given their importance as vectors of major human diseases (Michel and Kafatos, 2005). Approximately two weeks after acquisition of an infected blood meal, *Plasmodium* sporozoites are released into the hemocoel and face both humoral and cellular immune responses. Despite massive parasite mortality, malaria parasites infect the salivary glands and subsequently are transmitted to the vertebrate host during a blood meal. It has been demonstrated *in vitro* that parasite mortality in mosquitoes is mediated by phagocytosis and the anti-plasmodial activity of AMPs (Vizioli *et al.*, 2001, Hillyer *et al.*, 2003). The exact molecular mechanisms by which eukaryotic parasites are recognized and killed are not well characterized and are an active research area.

Studies on the molecular interactions between *T. cruzi* and its triatome vectors are scarce compared with other insect/parasite combinations. Ultra-structural studies have revealed potential interactions occurring *in vivo* between *T. cruzi* and the intestinal epithelium of the vectors (Kollien *et al.*, 1998), but because different regions of the intestine vary in their nutritional potential and surface characteristics, we do not know how these differences affect local gene expression that may affect *T. cruzi* development. If the parasite is injected into the hemolymph of *R. prolixus*, lysozyme, pro-phenoloxidase (proPO), and agglutination are activated (Mello *et al.*, 1995), and the parasite is killed and cannot be recovered (Azambuja *et al.*, 1999). However, *T. cruzi* normally does not enter the hemocoel. In vitro studies have demonstrated the susceptibility of *T. cruzi* to insect immune peptides (Jaynes *et al.*, 1988, Barr *et al.*, 1995), and *in vivo* studies have generated insects refractory to the parasite by engineering the bacterial gut symbionts to express a potent AMP in the midgut (Durvasula *et al.*, 1997). Studies on a closely related organism, *Trypanosoma rangeli,* which crosses the midgut epithelia and survives in the hemolymph, suggest that this parasite avoids the humoral immune system by infecting hemocytes and has the capacity to disable the proPO pathway that normally leads to melanization (Gregorio and Ratcliffe, 1991, Gomes *et al.*, 1999). Subsequent studies (Azambuja *et al.*, 2005) have demonstrated host immune responses in which lectins bind to carbohydrate moieties on the surface of *T. rangeli* thereby preventing their attachment to midgut and salivary glands. Identifying the specific pool of genes involved in host-parasite interactions could provide an insight into molecular mechanisms involved in parasite development and the specificity of these interactions.

The expression of insect and mammalian immune factors is pathogen specific; insects such as *D. melanogaster* discriminate between fungal and bacterial infections and use two main pathways, the Toll and the Imd pathways, to express specific molecules involved in their defense (Hoffmann, 2003). In this report, we

have identified similar pathogen-specific responses in *R. prolixus* to bacteria and *T. cruzi* using Suppressive Subtractive Hybridization. This technique selectively identifies differentially expressed genes in response to a particular stimulus rather than a general transcriptome analysis. We report here the generation and functional annotation of pathogen-specific expressed sequence tags (ESTs) from three subtracted libraries constructed from fat body and intestinal tissues of *R. prolixus* after exposure to bacterial pathogens and the parasite *T. cruzi*.

Material and Methods

Insect colony maintenance.

A *Rhodnius prolixus* colony was maintained at Simon Fraser University at room temperature with a 12h light/dark cycle. The colony was blood fed approximately every three weeks on guinea pigs.

Immune activation and tissue dissection.

Bacteria (*Escherichia coli* and *Micrococcus luteus*) were grown in liquid LB culture over night at 37°C with vigorous shaking and 0.75mL of each bacterial culture were mixed together and pelleted by centrifugation for five minutes at 5,000g in a tabletop centrifuge. A sterile minuten pin was dipped in the bacterial pellet and injected into *R. prolixus* adults or 5th instar nymphs thoraxes (Lopez *et al.*, 2003). Naïve (non-challenged) insects were used as controls. *Trypanosoma cruzi* was obtained from the feces of infected *Triatoma infestans*. The parasites were washed with PBS, and centrifuged at 4°C for 5 min at 5000g, and re-suspended in liver infusion tryptose (LIT) media and counted. Five microliters containing approximately 2500 parasites were inoculated into adult insects with a sterile syringe. Control insects were inoculated with 5μl of sterile LIT media. Twenty-four hours after immune challenge (bacteria, *T. cruzi* or LIT), fat bodies and intestinal tissues were dissected and thoroughly rinsed in ice cold PBS to wash any contaminating faeces and/or blood meal. Tissues were stored in RNAlater (Ambion, USA) or directly used for subsequent RNA isolation.

Total RNA and mRNA isolation.

Tissues stored in RNAlater were centrifuged at 14,000g for 5 min at 4°C. The supernatant was removed and tissues were washed with 1mL of DEPC-treated water and pelleted once again to remove the liquid supernatant. Total RNA extraction was performed using Triazol (Invitrogen, Canada) according to manufacturer's specifications. mRNA was isolated using Purist poly A micro spin columns (Ambion, USA); 1µg of poly A RNA was used in the construction of each subtracted library.

Subtractive library construction.

We generated three subtractive libraries: a midgut library in response to bacterial (*Escherichia coli* and *Micrococcus luteus*) injection; a fat body library in response to bacterial injection and a fat body library in response to *T. cruzi* injection. All three subtracted libraries were built using PCR-Select cDNA Subtraction kit according to manufacturers' recommendations (Clontech, Canada). Suppressive Subtractive Hybridization (SSH) permitted the enrichment of differentially expressed sequences by hybridizing a TESTER (pool of cDNAs from which differentially expressed genes were identified) to a DRIVER (control cDNAs used to remove common sequences) (Diatchenko *et al.*, 1996, Hunt and Livesey, 2000). Ligation of specific adaptors to both ends of the cDNAs was performed prior to subtraction hybridization, followed by PCR amplification with specific primers to the adaptors. Amplification of hybrids corresponding to common sequences was suppressed, yielding a library enriched for differentially expressed sequences (Figure 2-1).

For gut and fat body subtracted libraries in response to bacterial challenge, TESTERS cDNAs were constructed with mRNA from bacteria inoculated samples and DRIVERS with mRNA from naïve (non-inoculated) insects.

The fat body – *T. cruzi* subtracted library (forward) was built using mRNA from immune activated fat bodies as TESTER and mRNA from sterile media (LIT) inoculated insects as DRIVER. Reverse subtracted libraries were built for fat body tissue libraries for subsequent differential screening, where TESTER and

DRIVER designations are inversed. Forward subtracted libraries were ligated overnight at 4°C into 2µg of pGemTeasy plasmid vector (Promega, USA) using 3µL of the secondary PCR products from each library, and transformed by heat shock into *E. coli* JM109 ultra-competent cells (Promega, USA). The resulting EST library was plated on LB agar supplemented with 100µg/mL ampicillin, 80µg/mL Xgal, 0.5mM IPTG and incubated overnight at 37°C.

Subtractive efficiency analysis and differential screening.

The efficiency of the subtraction of all three libraries was estimated using PCR by comparing the abundance of known cDNAs before and after subtraction. β-actin was selected as a non-differentially expressed gene. Internal primers were used to amplify a portion of this gene (qActF:5'AATCAAGATCATTGCTCCACCAG3'; ActR:5'TTAGAAGCATTTGCGGTGGAC3') under the following conditions: 94°C for 1min followed by 33 cycles of 94°C for 20s, 60°C for 20s and 72°C for 30s. Five microliters aliquots were removed from each reaction after 18, 23, 28 and 33 cycles and examined by electrophoresis on a 2% agarose gel and stained with ethidium bromide to confirm subtraction success.

Fat body subtracted libraries were screened for differentially expressed ESTs following manufacturer's instructions using the PCR-select cDNA subtraction screening kit (Clontech, Canada). Clones from bacteria and *T. cruzi* libraries (95 and 194 respectively) were selected randomly and grown in 50µl of LB-ampicillin (100µg/mL) for six hours at 37°C with moderate shaking in 96 well plates. Two microliters of bacterial culture were spotted in duplicate on Hybond+ membranes (Amersham Biosciences, Canada); allowed to grow for 2 hours at 37°C on a LB agar plate, denatured in 0.5M NaOH; 1.5M NaCl for 4 min, neutralized in 1.5M NaCl; 0.5M Tris/HCl pH 7.5 for another 4 min and allowed to dry for 30 min at room temperature. Nucleic acids were fixed to the membrane by using a UV crosslinker XL 1000 (Spectronics corporation, USA).

One hundred and fifty nanograms from the forward and reverse subtracted libraries were used to create a ^{32}P-labeled probe by random priming using PCR-Select differential screening kit (Clontech, USA) following manufacturer's instructions. Forward and reverse subtracted probes were hybridized in individual tubes with the DNA membrane at 65°C for 2.5 hours in a rotatory oven using Rapid-Hyb buffer (Amersham Biosciences, Canada). Following hybridization the membranes were washed with low stringency (2X SSC, 0.5%SDS; three times, 20 min each) and high stringency (0.2X SSC, 0.5% SSC; three times, twenty minutes each) buffers at 65°C to eliminate nonspecific binding due to excess probe. Membranes were exposed to a Kodak BioMax MS film (Eastman Kodak, USA) overnight at room temperature.

Plasmid isolation, DNA sequencing and database search.

Selected colonies (strong signal with the forward and low signal with the reverse subtracted probe) were grown overnight in 5mL of LB medium with 5µl of Ampicillin (100µg/µl) and purified using the Wizard Plus Miniprep DNA Purification System (Promega, USA). Sequencing reactions were performed using Big Dye v3.1 chemistry and run on an ABI PRISM 377 (Applied Biosystems, USA) at the DNA sequencing facility of the University of British Columbia. Analysis of the sequence data, detection of open reading frames and sequence alignment, were performed using Lasergene's modules Seqman, Megalign and Editseq (DNAstar, USA). Database search was performed using BLAST-X against non-redundant database at NCBI with default parameters. The best annotated hit from the similarity search was retained. For functional prediction of ESTs found in the database, we used an online gene ontology annotation tool: GoFigure (Khan *et al.*, 2003) and clustered the ESTs based on the biological process annotation when available. Novel ESTs were submitted to dbEST at the US National Center for Biotechnology Information (NCBI) and assigned accession numbers 37906674 to 37906768 (GenBank accession EB084319 to EB084413).

Results

Midgut subtracted library in response to bacteria.

In this study, we assessed the presence of immune related transcripts in midgut tissues in response to bacterial invasion of the hemocoel of *R. prolixus*. These genes represent components of a systemic immune response; genes induced in the midgut after a stimulation of the fat body. We randomly isolated and sequenced 90 independent clones from the midgut subtracted library in response to bacterial injection into the hemocoel. After sequencing, we precluded from our analysis redundant clones, sequences with inserts under 60bp in length and clones providing poor quality sequence.

Although, we attempted to minimize the presence of bacterial sequences by using polyA RNA in the construction of the subtracted library, four clones contained bacterial DNA, possibly originating from one or more of the midgut bacterial symbionts that triatome insects naturally harbor and these were excluded from subsequent analyses.

In total, 66 clones (73%) were analyzed which corresponded to different EST sequences (Table 2-1). Similarity search by comparison to public database at NCBI using BLAST-X resulted in 16 clones with no significant match, and 5 to hypothetical proteins deduced in silico from genome sequencing and annotation projects.

Housekeeping genes (ribosomal, mitochondrial), whose amplification during the suppressive subtractive hybridization is normally repressed, also were found in the library. This was likely because we isolated tissues (gut) from regions distant from those directly stimulated (fat body) which allowed for fewer overall differentially expressed genes and therefore more ESTs that corresponded to non-differentially expressed genes. The efficiency of the subtraction depends on

the number of genes differentially expressed; larger numbers of these mRNAs are found in areas of lowest background.

What is most interesting is the identification of seven ESTs corresponding to genes that have been shown to participate in different immune response mechanisms. Among this category, we found lysozyme, nitrophorin, transferrin, defensin and a mucin subunit, corresponding to effector, signaling and possibly recognition mechanisms. Other ESTs included putative transcriptional regulators such as transcription factors identified by the presence of DNA binding domains as well as peptidases (cathepsin B and an aminopeptidase) whose enzymatic activity can have an effect in the development and establishment of *T. cruzi*.

Subtracted library from fat body (bacteria inoculation).

Shotgun cloning of secondary PCR products (Figure 2-2) into pGemTeasy resulted in just over 100 recombinant clones. Ninety-five clones from this subtracted library from fat body (bacteria inoculation) were spotted and differentially screened (Figure 2-3) Clones producing a strong hybridization signal with the forward library probe and simultaneously producing a low hybridization signal with the reverse library probe have over a 95% probability of being differentially expressed transcripts in response to the immune challenge. Twenty randomly picked clones and seven highly up-regulated clones were sequenced, compared to NCBI database, and their putative function determined with GoFigure (Khan *et al.*, 2003) (Table 2-2).

Subtraction efficiency was assessed by amplifying a known housekeeping gene. A 120bp fragment of Beta actin was amplified by PCR using the subtracted library or the un-subtracted as templates. Aliquots were taken at different PCR cycles and analyzed on an agarose gel. The beta actin amplicon is almost undetectable in the subtracted sample but was detected after 23 amplification cycles on the unsubtracted sample indicating that the proportion of transcripts for this housekeeping gene is minimal (Figure 2-4). Subtraction efficiency analysis by PCR, and the results obtained after sequencing, demonstrated the high

quality of this library as housekeeping transcripts were barely detected and three out of seven up-regulated genes corresponded to all three defensin isoforms which we identified by an independent HPLC analysis of immune hemolymph described in chapter 5 (Lopez *et al.*, 2003).

Subtracted library from fat body (*T. cruzi* inoculation).

Following a secondary PCR (Figure 2-5) and cloning into pGemTeasy, one hundred and ninety clones randomly picked were spotted on a membrane and differentially screened by hybridization to either a forward or a reverse ^{32}P-labeled probe (Figure 2-6). Ten clones producing a strong hybridization signal with the forward library and simultaneously producing a low hybridization signal with the reverse library probe were identified and subsequently sequenced and compared to public databases. Subtraction efficiency analysis by PCR showed that the efficiency of the subtraction was lower than the other two libraries (Figure 2-7). This probably was due to the use of cDNAs obtained from LIT inoculated insects as the DRIVER (control). All genes up-regulated in response to wounding in both DRIVER and TESTER would have been removed leaving only the genes expressed specifically in response to the presence of *T. cruzi* in the hemocoel. Two genes, despite having a differential hybridization profile corresponded to false positives, encoding for ribosomal genes. Two clones isolated from this library were of particular interest. Clone C10 with a high similarity to a mucin/peritrophin receptor molecule and clone H9 corresponding to a partial sequence of a Dorsal/Rel homologue (Table 2-3).

When combined all three subtracted libraries contain 103 EST sequences (94 are novel) from the hemipteran *R. prolixus*. A large majority of the sequences here reported do not have any similarities with other sequences in the databases or have similarities with sequences encoding for hypothetical proteins obtained from genome sequencing projects whose functional role is yet unknown (Figure 2-8).

Discussion

Insect immunity has received a great deal of attention in the past 25 years both from a basic research as well as applied perspective. However, these studies have focused mainly on the higher orders of insects: Diptera, Hymenoptera and Lepidoptera. Studies in more ancient insects such as hemipterans are scarce. Given the heterogeneity of invertebrates and their immune systems (Loker *et al.*, 2004), identifying components of the innate immune response in hemipteran insects provides an invaluable evolutionary view of immunity. These insects and their trypanosomatid parasites provide, despite the limited genetic tools available, an excellent case study because of the fundamental differences in the host life cycles (incomplete versus complete metamorphosis of higher insects) and the impact of the parasite on the immune system of the insect. Moreover, two very closely related flagellate parasites may have very different life cycles in the same insect, eg. *T. rangeli* circumvents the immune system and survives in the hemolymph whereas *T. cruzi* is limited to the intestines of the insect, removed from most immune effector molecules. We do not know if this is a function of differential recognition or activation processes of the insect or evasion techniques by the parasite. Prior to testing these different possibilities we first must identify and characterize the different components of this vector's immune system. Lysozyme, pro-PO, and agglutination have been detected after natural infection with *T. rangeli* and artificial injection of *T. cruzi* into the hemolymph of *R. prolixus* (Mello *et al.*, 1995). Despite these initial findings, relatively little is known concerning the molecular mechanisms involved in the recognition, activation and effector molecules of the hemipteran immune response to parasites.

Many different approaches have been used to identify immune-related molecules involved in vector-parasite interactions. Complete transcriptome studies have sequenced large EST libraries from the tsetse fly midgut (Lehane *et al.*, 2003) and mosquito hemocytes (Bartholomay *et al.*, 2004), and microarrays were used

to identify transcripts of *D. melanogaster* expressed in response to viruses, bacteria, fungi and a protozoan parasite (Roxstrom-Lindquist *et al.*, 2004). We used SSH and differential screening to identify novel and known *R. prolixus* ESTs up-regulated in different tissues after two different challenges. Many studies have described the role of the fat body in producing potent immune molecules. Our data indicate that as higher insects, the midgut of hemipterans is also immuno-competent and produces antimicrobial peptides such as defensin and lysozyme and likely second messenger molecules (transferrin, nitric oxide). Our results are consistent with other studies and provide evidence for the presence of a systemic immune response in which communication molecules induce the expression of immune factors in cells/tissues distant from the initial point of infection (Boulanger *et al.*, 2006).

We describe here, in more detail, the identification of 6 molecules belonging to the three fundamental pillars of immunity: recognition, activation and effector mechanisms involved in the immune response of *R. prolixus*.

Transferrin: A protein involved in iron metabolism in both vertebrates and invertebrates and responsive to juvenile hormone (Nichol *et al.*, 2002), transferrin is also believed to be a component of the innate immune system by sequestering iron away from bacterial pathogens (Lowenberger, 2001). Transferrin has been shown to be up-regulated in vitro in A. *aegypti* cells treated with heat-killed bacterial cells and in termites following exposure to an entomopathogenic fungus (Yoshiga *et al.*, 1999, Thompson *et al.*, 2003). The transferrin gene was recently cloned and characterized in A. *aegypti*; its promoter region is rich in putative NF-κB binding sites, consistent with its postulated role in insect innate immunity (Harizanova *et al.*, 2005). Although the exact role of transferrin in insect innate immunity has not yet been clearly elucidated, a hint of its possible role comes from studies in the goldfish *Carassius auratus* where it functions as an immune stimulatory signal, when enzymatically cleaved, by activating macrophages (Stafford *et al.*, 2001, Stafford and Belosevic, 2003, Stafford *et al.*, 2004).

Nitrophorins: Nitric oxide is a multifunctional molecule; its role in innate immunity has been reported extensively against bacteria and eukaryotic parasites (Nappi *et al.*, 2000, Ascenzi and Gradoni, 2002, Foley and O'Farrell, 2003). Six nitrophorins have been identified in *R. prolixus* (Champagne *et al.*, 1995, Moreira *et al.*, 2003) mainly from salivary gland tissue after blood feeding where NO is transported to function as a vasodilator and facilitates the blood meal acquisition. Upon *T. rangeli* infection of the hemolymph, nitric oxide activity has been detected and high levels of superoxide seem to limit the H14 strain of *T. rangeli*, which fails to complete its life cycle in *R. prolixus* (Whitten *et al.*, 2001). We isolated an EST with high sequence similarity to nitrophorin 3 in midgut tissue after bacterial infection (Table 2-1) and we can only hypothesize that the presence of NO in the gut may be used to regulate the growth of bacterial flora and also may have a negative effect of *T. cruzi* development.

β1-3 glucan recognition protein (β-GRP): The innate immune system recognizes microorganisms through a series of pattern recognition receptors that are highly conserved in evolution. Beta-glucan recognition proteins (β-GRPs) are pattern recognition molecules that are conserved from insects to mammals and recognize foreign organisms and their unique cell wall components. The first GRP was isolated from the hemolymph of *B. mori* (Yoshida *et al.*, 1996), and subsequently GRPs were identified as pathogen recognition molecules in *M. sexta, A. subalbatus* and *D. melanogaster* cells. These molecules activate the phenoloxidase cascade leading to pathogen encapsulation (Kim *et al.*, 2000, Ma and Kanost, 2000, Wang *et al.*, 2005).

Hemolymph proteinase (HP): We identified an EST with high homology to a *M. sexta* serine proteinase found in the hemolymph (Jiang *et al.*, 2005). Serine proteinase pathways play a pivotal role in controlling immune processes in insects. Hemolymph proteinases, secreted into the hemolymph from the fat body or hemocytes, are responsible for initiating the complex biochemical cascade

involved in pro-phenoloxidase (proPO) cleavage and activation. Phenoloxidase (PO), activated from proPO through proteolysis by proPO-activating proteinase (PAP), is a key enzyme implicated in several defense mechanisms in invertebrates. Other proteinases were identified (cathepsin B and a leucine aminopeptidase), albeit from midgut tissue where these molecules aid in blood meal digestion. Cathepsin D, however, has been linked to the cleavage of immune related molecules in fish (Cho *et al.*, 2002) and in *R. prolixus* infected with *T. cruzi,* its activity is decreased (Borges *et al.*, 2006). Initial data on Cathepsin B transcript levels indicate its gut specificity but no differential expression when compared to non-infected controls (data not shown).

Rel/Dorsal: This molecule belongs to a super family of nuclear factors. In *D. melanogaster* Dorsal plays a central role in the establishment of dorso-ventral polarity during early embryogenesis, whereas Relish plays a main role in the IMD pathway by activating the transcription of antimicrobial peptides (Dushay *et al.*, 1996, Hedengren *et al.*, 1999). Recently, Raikhel and colleagues identified its homologue in *A. aegypti*; Relish 1 (REL1) selectively binds to different NF-κB motifs from insect immune gene promoters and mediates a specific antifungal immune response against *B. bassiana* (Shin *et al.*, 2005). Using a transgenic approach, in combination with RNAi technique, they elucidated its role as a key downstream regulator of Toll immune pathway in *A. aegypti* (Bian *et al.*, 2005). The dual role of this molecule in such important, yet very different, processes such as development and immunity makes it a very interesting case study for its recruitment by one or the other process from an evolutionary perspective. Expression and functional studies of this molecule should shed light on its role in *R. prolixus* as a developmental and/or immune related transcription factor.

Mucin/Peritrophin like: Mucins are surface or free glycoproteins known to bind lectins (another group of surface glycoproteins). *Trypanosoma cruzi*'s genome encodes for large families of surface molecules, which include trans-sialidases, mucins, gp63s, and a large novel family (>1300 copies) of mucin-associated

surface protein (MASP) genes (Pollevick *et al.*, 2000, Acosta-Serrano *et al.*, 2001, El-Sayed *et al.*, 2005). Specific *R. prolixus* lectins are known to interact selectively with *T. cruzi* (Pereira *et al.*, 1981), including a hemolymph galactoside-binding lectin, which could play an important role in the development of *T. rangeli* in the hemocoel of the insect vector. This lectin markedly enhanced the activation of clump formation by *T. rangeli* in *R. prolixus* hemocyte monolayers, with an increase in clump size and hemocyte aggregation (Mello *et al.*, 1999). More recently, gp150, an ecdysone-regulated mucin found in *D. melanogaster* hemocytes, midgut, and salivary glands was shown to be released from larval hemocytes to become a component of the clot and participate in the entrapment of bacteria (Korayem *et al.*, 2004).

Our study did not identify any AMPs other than defensin and lysozyme despite the fact that more than 250 different AMPs have been described from different insect orders. This suggests that *R. prolixus* may have a different arsenal of AMPs (possibly comprising molecules we have designated as having no known function), that this insect has not developed a wide variety of defense molecules, or that the production of these molecules is not transcriptionally regulated. Interestingly, we found little overlap among the genes up-regulated in our three subtracted libraries. Defensin was the only AMP found in both fat body and midgut in response to bacterial injection. It is known that different pathogens elicit specific immune responses. In *D. melanogaster* for instance, Gram-positive bacteria and fungi trigger the Toll pathway, whereas Gram-negative bacteria trigger the IMD pathway (recently reviewed in Lemaitre and Hoffmann, 2007). The immune response in insects to large foreign organisms such as parasites is mediated by nodulation and encapsulation and not solely by AMPs. Therefore, it is not surprising that genes found in response to bacterial challenge differ from the genes found in response to a protozoan parasite such as *T. cruzi*. It is worthwhile to note that lysozyme was found only in response to bacterial challenge and not to *T. cruzi*, in contrast to a recent microarray-based study in *D.*

melanogaster where lysozymes were found to be the main response to the protozoan parasite *O. muscaedomesticae* (Roxstrom-Lindquist *et al.*, 2004).

Our results indicate the activation of several pathogen specific genes in response to bacterial or parasitic invasion of the hemocoel. Some of these are homologous to genes described in other insect-parasite systems but the large number of unidentified genes suggests the possibility of unique immune genes in hemimetabolous insects. Future studies will characterize these novel immune related genes in terms of biological activity and their effects on parasite development and transmission.

Figures
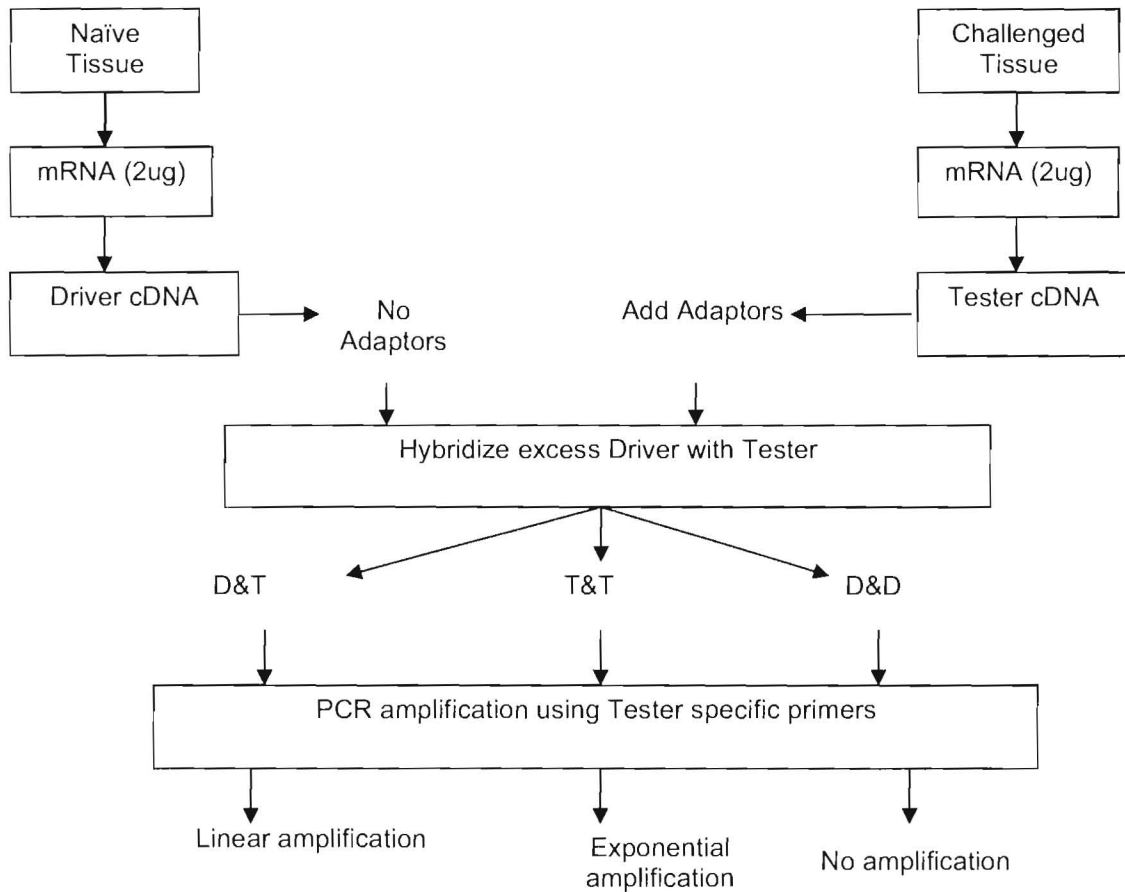
Figure 2-1 Suppressive Subtractive Hybridization overview.

Suppressive Subtractive Hybridization (SSH) permits the enrichment of differentially expressed sequences by hybridizing a TESTER (pool of cDNAs from which differentially expressed genes were identified) to a DRIVER (control cDNAs used to remove common sequences). Ligation of specific adaptors to both ends of the cDNAs was performed prior to subtraction hybridization, followed by PCR amplification with specific primers to the adaptors. Amplification of hybrids corresponding to common sequences was suppressed, yielding a library enriched for differentially expressed sequences.
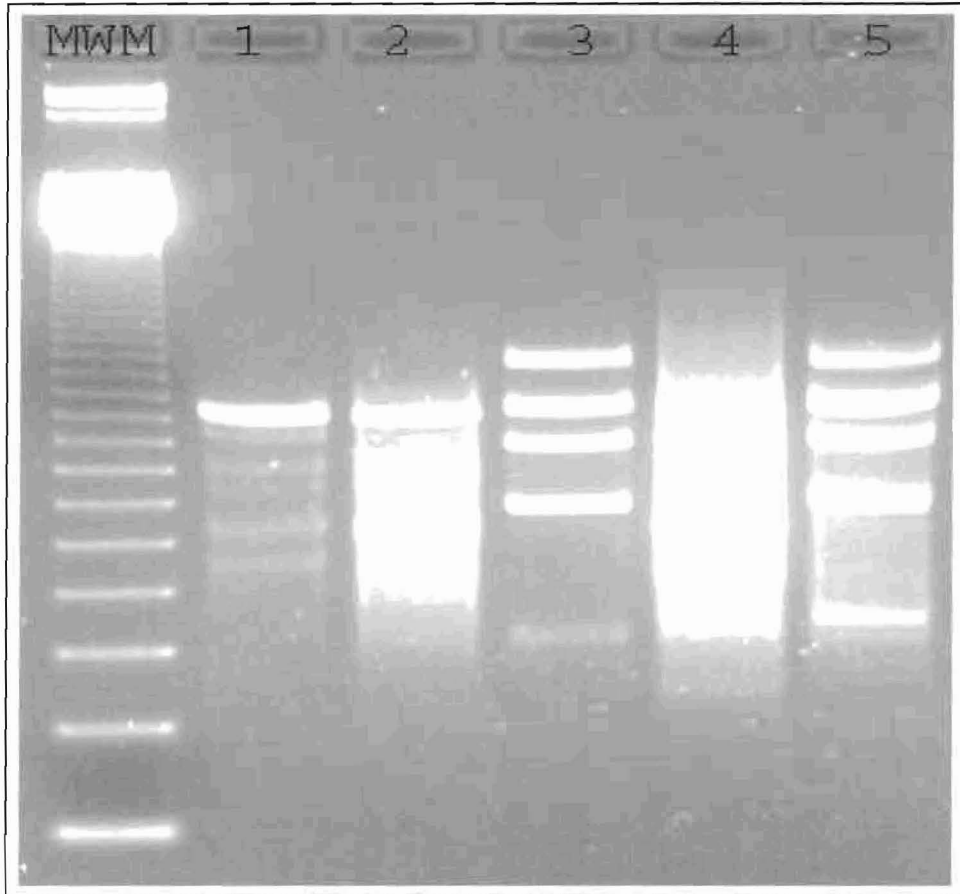
Figure 2-2 Fat body SSH library in response to bacterial injection.

Lanes 1&2: Subtracted and un-subtracted R. prolixus fat body 24h post bacterial inoculation. Lane 3: Control Φ174/HaeIII DNA. Lanes 4&5: Un-subtracted and subtracted human skeletal muscle with Φ174/HaeIII DNA. Lane 1 was shotgun cloned into pGemTeasy vector and transformed into *E. coli* JM109.
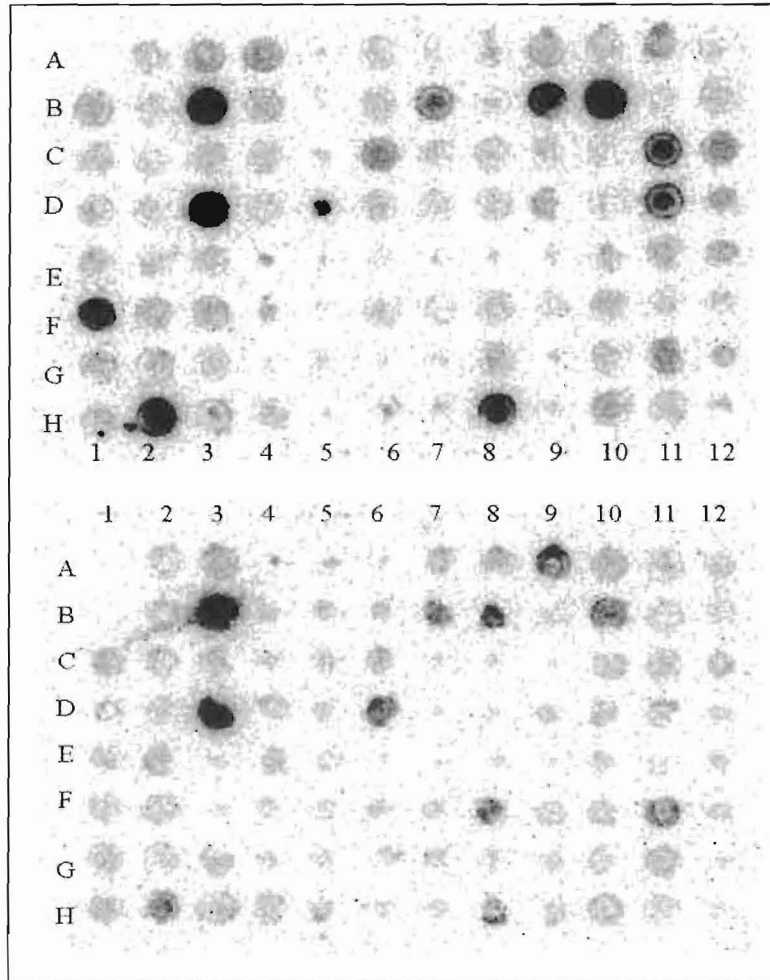
Figure 2-3 *R. prolixus* fat body bacteria injected subtracted library differential screening.

Top: 95 randomly picked clones hybridized to fwd subtracted library probe. Bottom: duplicate membrane hybridized to rev subtracted library probe. Colonies that produced a strong signal only in the forward subtracted probe were selected for sequencing.
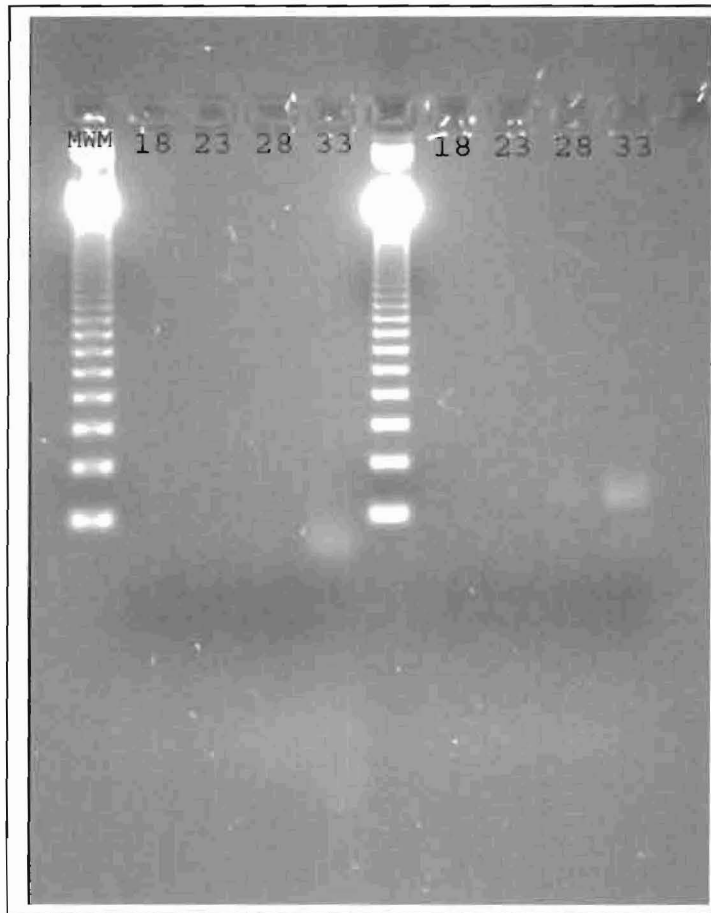
Figure 2-4 SSH subtraction efficiency test.

Subtraction efficiency was assessed by amplifying a known housekeeping gene. A 120bp fragment of Beta actin was amplified by PCR using the subtracted library (left) and the un-subtracted (right) as templates. 5uL aliquots were taken at different PCR cycles and separated on an agarose gel. The actin amplicon is almost undetectable in the subtracted sample yet it is detected after 23 cycles on the un-subtracted sample indicating that the proportion of transcripts for this housekeeping gene is minimal. MWM: 100bp molecular weight marker.
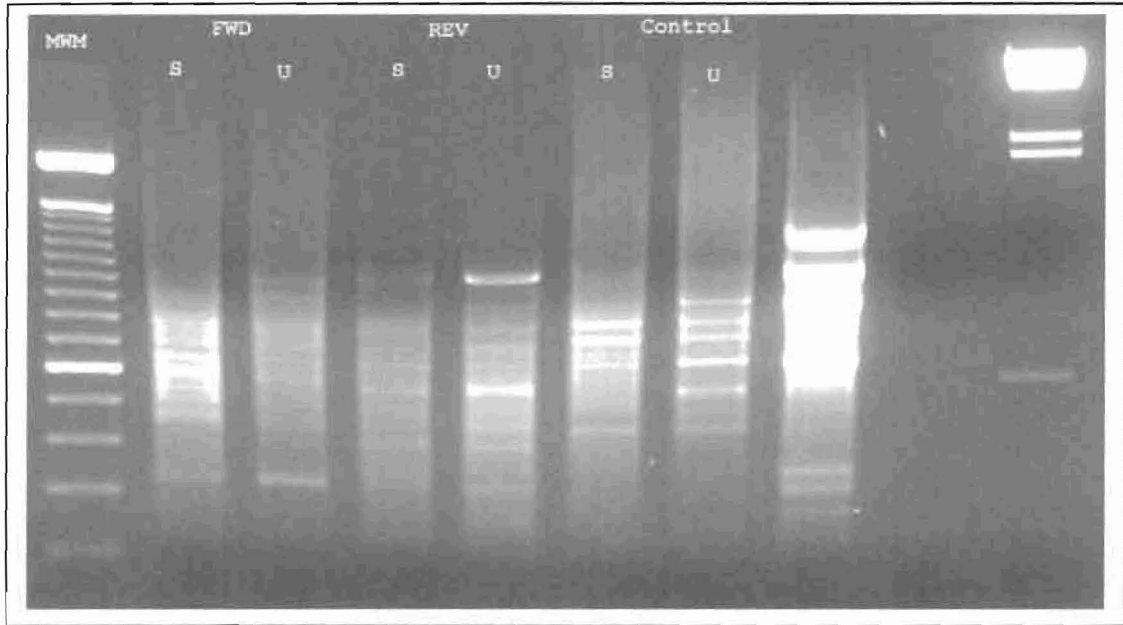
Figure 2-5 Fat body *T. cruzi* SSH library.

The subtracted library (forward) was built using mRNA from immune activated fat bodies as TESTER and mRNA from sterile media (LIT) inoculated insects as DRIVER. Reverse subtracted libraries were built for fat body tissue libraries for subsequent differential screening, where TESTER and DRIVER designations are inversed. FWD: forward subtracted library. REV; reverse subtracted library. Control: S: subtracted; U: un-subtracted. MWM: 100bp molecular weight marker.
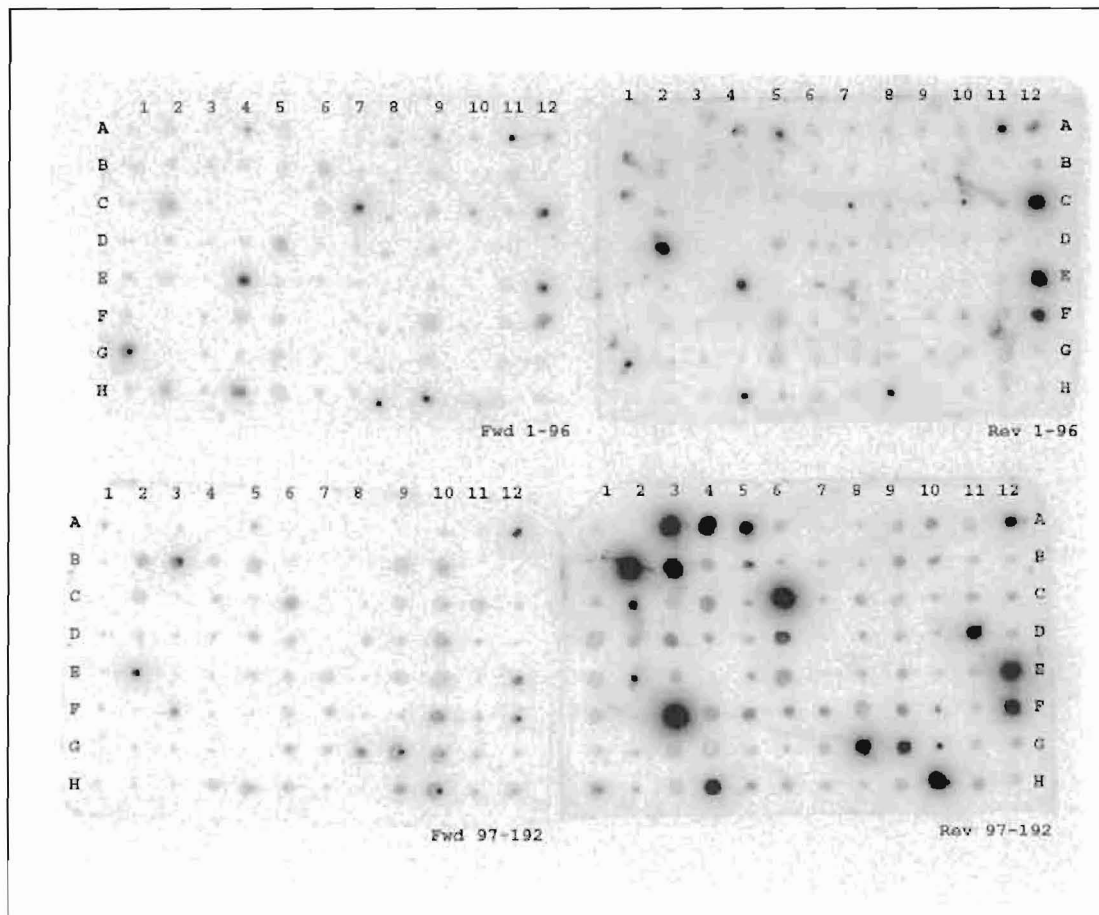
Figure 2-6 Differential screening of the *R. prolixus* fat body (*T. cruzi* injected) subtracted library.

Top: clones 1-96 hybridized to fwd (left) or rev (right) subtracted probe. Bottom: clones 96-192 hybridized to fwd (left) or rev (right) subtracted probe.
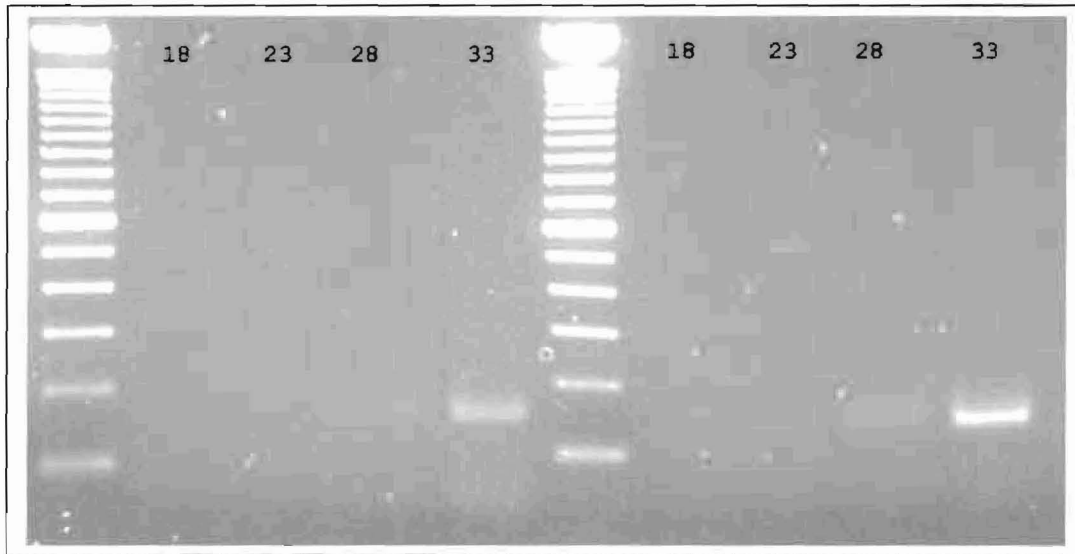
Figure 2-7 Fat body (*T. cruzi* injected) SSH library subtraction efficiency test.

Subtraction efficiency was assessed by amplifying a known housekeeping gene. A 120bp fragment of Beta actin was amplified by PCR using the subtracted library (left) and the un-subtracted (right) as templates. 5uL aliquots were taken at different PCR cycles and separated on an agarose gel. 100bp molecular weight marker.
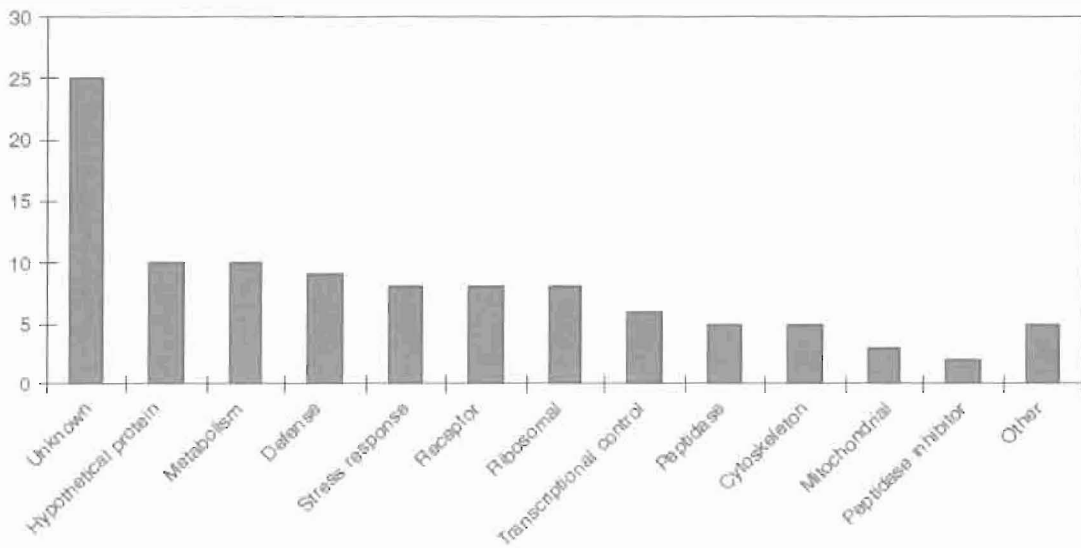
Figure 2-8 Functional prediction and classification of the generated ESTs.

Classification based on gene ontology using GoFigure. Novel ESTs were submitted to dbEST at the National Center for Biotechnology Information and assigned accession numbers 37906674–37906768 (GenBank accession EB084319–EB084413).

Tables

Table 2-1 *R. prolixus* midgut (bacteria-inoculated) subtraction library.

ESTs classified based on BLAST-X analysis against non-redundant database at NCBI. NSM: No significant match.

Clone	NCBI gi	Length (bp)	Blast-X match	Accession	E value	Putative gene function
2.59	37906735	325	Formin like protein	NP_035841.1	4e-23	Cytoskeleton
1.48	37906690	333	Kinesin like	gi41688591	4e-9	Cytoskeleton
2.61	37906737	325	WW domain binding protein 3	XP_235648.2	5e-26	Cytoskeleton
2.1	37906693	252	Alpha actinin	gi7441362	5e-36	Cytoskeleton
2.34	37906719	502	Lysozyme	AAN87265.1	1e-7	Defence
2.35	29335960	469	Defensin B	AAO74625.1	2e-50	Defence
2.22	37906708	317	Transferrin	AAW70172.1	1e-17	Defense
1.4	4204973	587	Nitrophorin 3	U61143.1	1e-84	Defense
2.2	37906694	278	Transferrin	AAA27820.1	5e-18	Defense
1.42	37906685	603	Hypothetical protein	CAH93767.1	0.057	Hypothetical protein
2.9	37906699	189	Hypothetical protein	XP_761391.1	0.064	Hypothetical protein
2.32	37906717	301	Hypothetical protein	XP_729786.1	0.4	Hypothetical protein
2.47	37906725	389	Hypothetical protein	CAD52327.1	0.69	Hypothetical protein
2.60	37906736	466	Hypothetical protein	BAB29490.1	2e-7	Hypothetical protein
2.50	37906728	241	Phosphomannose isomerase	AAK69388.1	1e-12	Metabolism
2.3	37906695	563	ATPase subunit 6	AAG31613.1	1e-37	Metabolism
1.3	37906676	291	polyamine oxidase	XP_508137.1	1e-8	Metabolism
2.29	37906714	460	ATP synthase β subunit	AAT06139.1	2e-65	Metabolism
2.15	37906702	238	Poly A binding protein	CAA40721.1	3e-29	Metabolism
2.36	37906720	1036	Sugar transporters	NP_568494.1	4e-16	Metabolism
2.49	37906727	687	Maltase precursor	CAA93821.1	6e-28	Metabolism
2.52	37906730	287	Polyamine oxidase	NP_997011.1	7e-14	Metabolism
1.16	20378665	468	Cytochrome oxydase 1	AAM20928.1	2e-60	Mitochondrial
2.40	37906721	439	mitochondrial thioredoxin	BAA13447.1	6e-28	Mitochondrial

Clone	NCBI gi	Length (bp)	Blast-X match	Accession	E value	Putative gene function
1.45	37906688	486	Cytochrome P450	BAA28946.1	9e-17	Mitochondrial
2.7	37906697	346	Mitochondrial peptidase	AB163419.1	1e-8	Peptidase
2.58	37906734	456	Aminopeptidase	Gi25311909	2e-6	Peptidase
1.52	37906692	336	Cathepsin B	AAT48985.1	5e-46	Peptidase
2.16	37906703	337	Leucine aminopeptidase	XP_786205.1	8e-8	Peptidase
2.33	37906718	552	Dipetalogastin	CAA10384.1	5e-45	Peptidase inhibitor
2.31	37906716	510	Thrombin inhibitor	AAK57342.1	6e-35	Peptidase inhibitor
1.14	37906682	154	CSP	CAB65177.1	1e-12	Receptor
2.28	37906713	484	Growth hormone inducible transmembrane protein	AAD44495.1	5e-21	Receptor
1.13	37906681	99	membrane-associated ring finger	NP_005876.2	7e-12	Receptor
2.10	37906700	365	Veph-A	XP_342257.1	8e-5	Receptor
2.27	37906712	635	Mucin subunit	AAA85523.1	9e-4	Receptor
2.46	37906724	592	NADH dehydrogenase	AAG31614.1	1e-62	Ribosomal
2.17	37906704	370	S24 ribosomal protein	AAS91555.1	2e-41	Ribosomal
2.48	37906726	245	Ribosomal protein L26	AAK92162.1	5e-17	Ribosomal
1.46	37906689	173	PRKA1	XP_790232.1	4e-9	Rna binding
2.23	37906709	394	GASZ	ABA90396.1	0.91	Signalling
1.5	37906677	328	Nin one	NP_001016830.1	1e-10	Stress response
1.44	37906687	618	UNR	CAD52327.1	1e-54	Stress response
1.6	37906678	322	Nin one	AAQ16153.1	4e-12	Stress response
2.19	37906706	140	HSP 70	AAP57537.3	4e-18	Stress response
2.55	37906731	377	HSP70	BAB92074.1	4e-42	Stress response
2.18	37906705	150	Chaperonin	NP_741154.1	5e-15	Stress response
1.43	37906686	140	HSP70	AAP57537.3	6e-18	Stress response
2.44	37906723	398	Zinc finger containing protein	CAD52327.1	0.69	Transcriptional control
1.7	37906679	618	UNR	gi137045	3e-54	Transcriptional control
1.1	37906674	149	NSM			Unknown
1.15	37906683	180	NSM			Unknown
1.2	37906675	185	NSM			Unknown

Clone	NCBI gi	Length (bp)	Blast-X match	Accession	E value	Putative gene function
1.40	37906684	297	NSM			Unknown
1.49	37906691	161	NSM			Unknown
1.8	37906680	197	NSM			Unknown
2.11	37906701	522	NSM			Unknown
2.21	37906707	171	NSM			Unknown
2.25	37906710	182	NSM			Unknown
2.30	37906715	304	NSM			Unknown
2.42	37906722	440	NSM			Unknown
2.5	37906696	276	NSM			Unknown
2.51	37906729	313	NSM			Unknown
2.56	37906732	190	NSM			Unknown
2.57	37906733	337	NSM			Unknown
2.8	37906698	449	NSM			Unknown

Table 2-2 *R. prolixus* fat body (bacteria-inoculated) subtraction library.

ESTs classified based on BLAST-X analysis against non-redundant database at NCBI. The first four clones were randomly picked whereas the rest were selected after differential screening. Rdm: Randomly picked clone. NSM: No significant match.

Clone	NCBI gi	Length (bp)	BLAST-X hit	Accession	E value	Putative gene function
Rdm 17	37906753	681	Kinesin like	Gi41688591	9e-62	ATP binding
B10	29335958	453	*R. prolixus* Defensin A	AAO74624.1	1e-45	Defence
H8	29335960	350	*R. prolixus* Defensin B	AAO74625.1	2e-42	Defence
H2	29335960	514	*R. prolixus* Defensin B	AAO74625.1	2e-51	Defence
Rdm 2	29335960	628	*R. prolixus* Defensin B	AAO74625.1	9e-51	Defence
Rdm 5	37906741	884	P. yoeli Hypothetical protein	EEA15590.1	0.062	Hypothetical protein
C11	37906760	465	Unknown protein	XP_379325.2	0.29	Hypothetical protein
F1	37906758	552	Hypothetical protein	CAG05504.1	0.82	Hypothetical protein
Rdm 20	37906756	600	H. sapiens hypothetical protein	EAL24336.1	3e-7	Hypothetical protein
Rdm 15	37906751	1012	A. mellifera ubiquitin ligase	XP_394362.2	6e-46	Ligase
Rdm 19	37906755	348	C.elegans pyruvate dehydrogenase	NP_500340.1	2e-38	Metabolism
Rdm 10	37906746	573	Dihydropteridine reductase	Gi442830	2e-49	Metabolism
B9	37906759	438	M. sexta Hemolymph proteinase	AAV91014.1	9e-30	Peptidase
Rdm 6	37906742	1033	G. gallus metalloprotease	XP_4185641.1	9e-54	Peptidase
Rdm 4	37906740	505	A. mellifera Metaxin like	XP_624291.1	1e-24	Protein transport
Rdm 18	37906754	458	B. clausii ABC transporter	BAD64657.1	0.8	Receptor
Rdm 8	37906744	735	β1-3 GRP	Gi52782700	1e-22	Receptor
Rdm 16	37906752	332	M.musculus proteasome 26S	AAH19112.1	4e-7	Ribosomal
Rdm 14	37906750	1022	H. sapiens ubiquitin	NP_066289.2	5e-160	Ribosomal
D5	37906757	278	Hypothetical transcription factor	AAX26421.1	0.36	Transcriptional control
Rdm 1	37906738	424	NSM			Unknown

Clone	NCBI gi	Length (bp)	BLAST-X hit	Accession	E value	Putative gene function
Rdm 11	37906747	678	NSM			Unknown
Rdm 13	37906749	423	NSM			Unknown
Rdm 3	37906739	1001	NSM			Unknown
Rdm 9	37906745	248	NSM			Unknown
Rdm 7	37906743	320	A. gambiae genomic clone	XP_312744.2	0.12	Unknown
Rdm 12	37906748	278	A. mellifera genomic clone	XP_394116.1	1e−21	Unknown

Table 2-3 *R. prolixus* fat body (*T. cruzi* inoculated) subtracted library.

ESTs classified based on BLAST-X analysis against non-redundant database at NCBI. NSM: no significant match.

Mb	Clone	Length (bp)	NCBI gi	BLAST-X hit	Accession	E value	Putative gene function
1–95	C2	1133	37906764	S. scrofa Flotillin	BAD08436.1	9e−96	Cytoskeleton
97–191	F7	507	37906767	B. Taurus Hypothetical protein	XP_583059.1	7e−14	Hypothetical protein
1–95	C10	536	37906765	C. felis Mucin-Peritrophin	AAM21357.1	6e−6	Receptor
1–95	C7	551	16117393	*R. prolixus* 16S ribosomal	AF324519.1	0.0	Ribosomal
97–191	B3	1022	2895883	*R. prolixus* Ribosomal RNA	AF045707.1	7e−180	Ribosomal
1–95	A8	459	37906762	A. mellifera genomic clone	XP_394615.2	5e−49	Stress response
1–95	H9	364	37906768	A. mellifera Dorsal B	AAP23056.1	1e−12	Transcriptional control
1–95	A4	481	37906761	P. troglodytes Formin-like	XP_522563.1	0.001	Transcriptional control
1–95	A9	488	37906763	NSM			Unknown
1–95	D5	1123	37906766	NSM			Unknown

References

Acosta-Serrano, A., Almeida, I.C., Freitas-Junior, L.H., Yoshida, N. & Schenkman, S., 2001. The mucin-like glycoprotein super-family of *Trypanosoma cruzi*: structure and biological roles. *Molecular and biochemical parasitology,* 114, 143-50.

Ascenzi, P. & Gradoni, L., 2002. Nitric oxide limits parasite development in vectors and in invertebrate intermediate hosts. *IUBMB life,* 53, 121-3.

Azambuja, P., Feder, D., Mello, C., Gomes, S. & Garcia, E., 1999. Immunity in *Rhodnius prolixus*: trypanosomatid-vector interactions. *Memorias do Instituto Oswaldo Cruz,* 94 Suppl 1, 219-22.

Azambuja, P., Ratcliffe, N.A. & Garcia, E.S., 2005. Towards an understanding of the interactions of *Trypanosoma cruzi* and *Trypanosoma rangeli* within the reduviid insect host *Rhodnius prolixus. Anais da Academia Brasileira de Ciencias,* 77, 397-404.

Barr, S.C., Rose, D. & Jaynes, J.M., 1995. Activity of lytic peptides against intracellular *Trypanosoma cruzi* amastigotes in vitro and parasitemias in mice. *The Journal of parasitology,* 81, 974-8.

Bartholomay, L.C., Cho, W.L., Rocheleau, T.A., Boyle, J.P., Beck, E.T., Fuchs, J.F., Liss, P., Rusch, M., Butler, K.M., Wu, R.C., Lin, S.P., Kuo, H.Y., Tsao, I.Y., Huang, C.Y., Liu, T.T., Hsiao, K.J., Tsai, S.F., Yang, U.C., Nappi, A.J., Perna, N.T., Chen, C.C. & Christensen, B.M., 2004. Description of the transcriptomes of immune response-activated hemocytes from the mosquito vectors *Aedes aegypti* and *Armigeres subalbatus. Infection and immunity,* 72, 4114-26.

Bian, G., Shin, S.W., Cheon, H.M., Kokoza, V. & Raikhel, A.S., 2005. Transgenic alteration of Toll immune pathway in the female mosquito *Aedes aegypti. Proceedings of the National Academy of Sciences of the United States of America,* 102, 13568-73.

Borges, E.C., Machado, E.M., Garcia, E.S. & Azambuja, P., 2006. *Trypanosoma cruzi*: effects of infection on cathepsin D activity in the midgut of *Rhodnius prolixus. Experimental parasitology,* 112, 130-3.

Boulanger, N., Bulet, P. & Lowenberger, C., 2006. Antimicrobial peptides in the interactions between insects and flagellate parasites. *Trends in parasitology,* 22, 262-8.

Champagne, D.E., Nussenzveig, R.H. & Ribeiro, J.M., 1995. Purification, partial characterization, and cloning of nitric oxide-carrying heme proteins (nitrophorins) from salivary glands of the blood-sucking insect *Rhodnius prolixus. The Journal of biological chemistry,* 270, 8691-5.

Cho, J.H., Park, I.Y., Kim, H.S., Lee, W.T., Kim, M.S. & Kim, S.C., 2002. Cathepsin D produces antimicrobial peptide parasin I from histone H2A in the skin mucosa of fish. *The FASEB journal,* 16, 429-31.

Diatchenko, L., Lau, Y.F., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. & Siebert, P.D., 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proceedings of the National Academy of Sciences of the United States of America,* 93, 6025-30.

Durvasula, R.V., Gumbs, A., Panackal, A., Kruglov, O., Aksoy, S., Merrifield, R.B., Richards, F.F. & Beard, C.B., 1997. Prevention of insect-borne disease: an approach using transgenic symbiotic bacteria. *Proceedings of the National Academy of Sciences of the United States of America,* 94, 3274-8.

Dushay, M.S., Asling, B. & Hultmark, D., 1996. Origins of immunity: Relish, a compound Rel-like gene in the antibacterial defense of *Drosophila. Proceedings of the National Academy of Sciences of the United States of America,* 93, 10343-7.

El-Sayed, N.M., Myler, P.J., Bartholomeu, D.C., Nilsson, D., Aggarwal, G., Tran, A.N., Ghedin, E., Worthey, E.A., Delcher, A.L., Blandin, G., Westenberger, S.J., Caler, E., Cerqueira, G.C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D.A., Carrington, M., Crabtree, J., Darban, H., Da Silveira, J.F., De Jong, P., Edwards, K., Englund, P.T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A.C., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M.J., Lorenzi, H., Louie, T., Machado, C.R., Mcculloch, R., Mckenna, A., Mizuno, Y., Mottram, J.C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J.L., Rinta, J., Robertson, L., Salzberg, S.L., Sanchez, D.O., Seyler, A., Sharma, R., Shetty, J., Simpson, A.J., Sisk, E., Tammi, M.T., Tarleton, R., Teixeira, S., Van Aken, S., Vogt, C., Ward, P.N., Wickstead, B., Wortman, J., White, O., Fraser, C.M., Stuart, K.D. & Andersson, B., 2005. The genome sequence of *Trypanosoma cruzi,* etiologic agent of Chagas disease. *Science,* 309, 409-15.

Foley, E. & O'farrell, P.H., 2003. Nitric oxide contributes to induction of innate immune responses to gram-negative bacteria in *Drosophila. Genes & development,* 17, 115-25.

Gomes, S.A., Feder, D., Thomas, N.E., Garcia, E.S. & Azambuja, P., 1999. *Rhodnius prolixus* infected with *Trypanosoma rangeli:* In vivo and in vitro experiments. *Journal of invertebrate pathology,* 73, 289-93.

Gregorio, E.A. & Ratcliffe, N.A., 1991. The prophenoloxidase system and in vitro interaction of *Trypanosoma rangeli* with *Rhodnius prolixus* and *Triatoma infestans* haemolymph. *Parasite immunology,* 13, 551-64.

Harizanova, N., Georgieva, T., Dunkov, B.C., Yoshiga, T. & Law, J.H., 2005. *Aedes aegypti* transferrin. Gene structure, expression pattern, and regulation. *Insect molecular biology,* 14, 79-88.

Hedengren, M., Asling, B., Dushay, M.S., Ando, I., Ekengren, S., Wihlborg, M. & Hultmark, D., 1999. Relish, a central factor in the control of humoral but not cellular immunity in *Drosophila. Molecular cell,* 4, 827-37.

Hillyer, J.F., Schmidt, S.L. & Christensen, B.M., 2003. Rapid phagocytosis and melanization of bacteria and *Plasmodium* sporozoites by hemocytes of the mosquito *Aedes aegypti. The Journal of parasitology,* 89, 62-9.

Hoffmann, J.A., 2003. The immune response of *Drosophila. Nature,* 426, 33-8.

Hoffmann, J.A. & Reichhart, J.M., 2002. *Drosophila* innate immunity: an evolutionary perspective. *Nature immunology,* 3, 121-6.

Hunt, S.P. & Livesey, F.J., 2000. *Functional genomics: a practical approach* Oxford; New York: Oxford University Press.

Jaynes, J.M., Burton, C.A., Barr, S.B., Jeffers, G.W., Julian, G.R., White, K.L., Enright, F.M., Klei, T.R. & Laine, R.A., 1988. In vitro cytocidal effect of novel lytic peptides on *Plasmodium falciparum* and *Trypanosoma cruzi. The FASEB journal,* 2, 2878-83.

Jiang, H., Wang, Y., Gu, Y., Guo, X., Zou, Z., Scholz, F., Trenczek, T.E. & Kanost, M.R., 2005. Molecular identification of a bevy of serine proteinases in *Manduca sexta* hemolymph. *Insect biochemistry and molecular biology,* 35, 931-43.

Khan, S., Situ, G., Decker, K. & Schmidt, C.J., 2003. GoFigure: automated Gene Ontology annotation. *Bioinformatics,* 19, 2484-5.

Kim, Y.S., Ryu, J.H., Han, S.J., Choi, K.H., Nam, K.B., Jang, I.H., Lemaitre, B., Brey, P.T. & Lee, W.J., 2000. Gram-negative bacteria-binding protein, a pattern recognition receptor for lipopolysaccharide and beta-1,3-glucan that mediates the signaling for the induction of innate immune genes in *Drosophila melanogaster* cells. *The Journal of biological chemistry,* 275, 32721-7.

Kollien, A.H., Schmidt, J. & Schaub, G.A., 1998. Modes of association of *Trypanosoma cruzi* with the intestinal tract of the vector *Triatoma infestans. Acta tropica,* 70, 127-41.

Korayem, A.M., Fabbri, M., Takahashi, K., Scherfer, C., Lindgren, M., Schmidt, O., Ueda, R., Dushay, M.S. & Theopold, U., 2004. A *Drosophila* salivary gland mucin is also expressed in immune tissues: evidence for a function in coagulation and the entrapment of bacteria. *Insect biochemistry and molecular biology,* 34, 1297-304.

Lavine, M.D. & Strand, M.R., 2002. Insect hemocytes and their role in immunity. *Insect biochemistry and molecular biology,* 32, 1295-309.

Lehane, M.J., Aksoy, S., Gibson, W., Kerhornou, A., Berriman, M., Hamilton, J., Soares, M.B., Bonaldo, M.F., Lehane, S. & Hall, N., 2003. Adult midgut expressed sequence tags from the tsetse fly *Glossina morsitans morsitans* and expression analysis of putative immune response genes. *Genome biology,* 4, R63.

Loker, E.S., Adema, C.M., Zhang, S.M. & Kepler, T.B., 2004. Invertebrate immune systems--not homogeneous, not simple, not well understood. *Immunological reviews,* 198, 10-24.

Lopez, L., Morales, G., Ursic, R., Wolff, M. & Lowenberger, C., 2003. Isolation and characterization of a novel insect defensin from *Rhodnius prolixus*, a vector of Chagas disease. *Insect biochemistry and molecular biology,* 33, 439-47.

Lowenberger, C., 2001. Innate immune response of *Aedes aegypti*. *Insect biochemistry and molecular biology,* 31, 219-29.

Ma, C. & Kanost, M.R., 2000. A beta1,3-glucan recognition protein from an insect, *Manduca sexta*, agglutinates microorganisms and activates the phenoloxidase cascade. *The Journal of biological chemistry,* 275, 7505-14.

Mattock, N., Pink, R. & Special Programme for Research and Training in Tropical Diseases., 2005. *Making health research work for poor people: progress 2003-2004: tropical disease research: seventeenth programme report* Geneva: TDR, World Health Organization.

Medzhitov, R. & Janeway, C.A., Jr., 2002. Decoding the patterns of self and nonself by the innate immune system. *Science,* 296, 298-300.

Meister, S., Kanzok, S.M., Zheng, X.L., Luna, C., Li, T.R., Hoa, N.T., Clayton, J.R., White, K.P., Kafatos, F.C., Christophides, G.K. & Zheng, L., 2005. Immune signaling pathways regulating bacterial and malaria parasite infection of the mosquito *Anopheles gambiae. Proceedings of the National Academy of Sciences of the United States of America,* 102, 11420-5.

Mello, C.B., Garcia, E.S., Ratcliffe, N.A. & Azambuja, P., 1995. *Trypanosoma cruzi* and *Trypanosoma rangeli*: interplay with hemolymph components of *Rhodnius prolixus. Journal of invertebrate pathology,* 65, 261-8.

Mello, C.B., Nigam, Y., Garcia, E.S., Azambuja, P., Newton, R.P. & Ratcliffe, N.A., 1999. Studies on a haemolymph lectin isolated from *Rhodnius prolixus* and its interaction with *Trypanosoma rangeli. Experimental parasitology,* 91, 289-96.

Michel, K. & Kafatos, F.C., 2005. Mosquito immunity against Plasmodium. *Insect biochemistry and molecular biology,* 35, 677-89.

Moreira, M.F., Coelho, H.S., Zingali, R.B., Oliveira, P.L. & Masuda, H., 2003. Changes in salivary nitrophorin profile during the life cycle of the blood-sucking bug *Rhodnius prolixus. Insect biochemistry and molecular biology,* 33, 23-8.

Nappi, A.J., Vass, E., Frey, F. & Carton, Y., 2000. Nitric oxide involvement in *Drosophila* immunity. *Nitric oxide,* 4, 423-30.

Nichol, H., Law, J.H. & Winzerling, J.J., 2002. Iron metabolism in insects. *Annual review of entomology,* 47, 535-59.

Nurnberger, T., Brunner, F., Kemmerling, B. & Piater, L., 2004. Innate immunity in plants and animals: striking similarities and obvious differences. *Immunological reviews,* 198, 249-66.

Osta, M.A., Christophides, G.K. & Kafatos, F.C., 2004. Effects of mosquito genes on *Plasmodium* development. *Science,* 303, 2030-2.

Pereira, M.E., Andrade, A.F. & Ribeiro, J.M., 1981. Lectins of distinct specificity in *Rhodnius prolixus* interact selectively with *Trypanosoma cruzi. Science,* 211, 597-600.

Pollevick, G.D., Di Noia, J.M., Salto, M.L., Lima, C., Leguizamon, M.S., De Lederkremer, R.M. & Frasch, A.C., 2000. *Trypanosoma cruzi* surface mucins with exposed variant epitopes. *The Journal of biological chemistry,* 275, 27671-80.

Roxstrom-Lindquist, K., Terenius, O. & Faye, I., 2004. Parasite-specific immune response in adult *Drosophila melanogaster*: a genomic study. *EMBO reports,* 5, 207-12.

Shin, S.W., Kokoza, V., Bian, G., Cheon, H.M., Kim, Y.J. & Raikhel, A.S., 2005. REL1, a homologue of Drosophila dorsal, regulates toll antifungal immune pathway in the female mosquito *Aedes aegypti. The Journal of biological chemistry,* 280, 16499-507.

Stafford, J.L. & Belosevic, M., 2003. Transferrin and the innate immune response of fish: identification of a novel mechanism of macrophage activation. *Developmental and comparative immunology,* 27, 539-54.

Stafford, J.L., Neumann, N.F. & Belosevic, M., 2001. Products of proteolytic cleavage of transferrin induce nitric oxide response of goldfish macrophages. *Developmental and comparative immunology,* 25, 101-15.

Stafford, J.L., Wilson, E.C. & Belosevic, M., 2004. Recombinant transferrin induces nitric oxide response in goldfish and murine macrophages. *Fish & shellfish immunology,* 17, 171-85.

Thompson, G.J., Crozier, Y.C. & Crozier, R.H., 2003. Isolation and characterization of a termite transferrin gene up-regulated on infection. *Insect molecular biology,* 12, 1-7.

Vizioli, J., Richman, A.M., Uttenweiler-Joseph, S., Blass, C. & Bulet, P., 2001. The defensin peptide of the malaria vector mosquito *Anopheles gambiae*: antimicrobial activities and expression in adult mosquitoes. *Insect biochemistry and molecular biology,* 31, 241-8.

Wang, X., Fuchs, J.F., Infanger, L.C., Rocheleau, T.A., Hillyer, J.F., Chen, C.C. & Christensen, B.M., 2005. Mosquito innate immunity: involvement of beta 1,3-glucan recognition protein in melanotic encapsulation immune responses in *Armigeres subalbatus*. *Molecular and biochemical parasitology,* 139, 65-73.

Whitten, M.M., Mello, C.B., Gomes, S.A., Nigam, Y., Azambuja, P., Garcia, E.S. & Ratcliffe, N.A., 2001. Role of superoxide and reactive nitrogen intermediates in *Rhodnius prolixus* (Reduviidae)/*Trypanosoma rangeli* interactions. *Experimental parasitology,* 98, 44-57.

Yoshida, H., Kinoshita, K. & Ashida, M., 1996. Purification of a peptidoglycan recognition protein from hemolymph of the silkworm, *Bombyx mori*. *The Journal of biological chemistry,* 271, 13854-60.

Yoshiga, T., Georgieva, T., Dunkov, B.C., Harizanova, N., Ralchev, K. & Law, J.H., 1999. *Drosophila melanogaster* transferrin. Cloning, deduced protein sequence, expression during the life cycle, gene localization and up-regulation on bacterial infection. *European journal of biochemistry / FEBS,* 260, 414-20.

Connecting statement 2

In Chapter two we used a functional genomics approach termed suppression subtractive hybridization to isolate expressed sequence tags (ESTs) up-regulated in response to immune challenge from *Rhodnius prolixus* fat body and intestinal tissues. Putative functional annotation of these ESTs based on similarity searches, revealed that the molecules they encode belong to one of the three different pillars of immunity: recognition, activation or effector mechanisms. While effectors are important in eliminating microbial pathogens, the expression must be tightly regulated. In the next chapter, we describe the identification and characterization of a transcription factor belonging to the Rel/NF-κB family of proteins, and we identify the potential effector target genes it regulates.

CHAPTER 3: MOLECULAR CLONING AND CHARACTERIZATION OF THREE *RHODNIUS PROLIXUS* DORSAL TRANSCRIPTION FACTORS, MEMBERS OF THE REL/NF-κB FAMILY

URSIC BEDOYA R., BUCHHOP J., LOWENBERGER C.
Simon Fraser University
Department of Biological Sciences
Burnaby, British Columbia, V5A 1S6, Canada

Abstract

Rhodnius prolixus is an ancient hematophagous hemipteran insect capable of mounting a powerful immune response. This response, in insects, is transcriptionally regulated in part by transcription factors of the Rel/NF-κB family. We have cloned and characterized three members of this transcription factor family in this insect. Dorsal 1A, and its differentially spliced product dorsal 1B, are expressed in early developmental stages and in a tissue specific manner in adults suggesting their dual role in development and immunity. In contrast, Dorsal 1C is encoded by a different transcript expressed only in the adult fat body in response to septic injury suggesting its exclusive role in immunity. Additionally, we identified putative NF-κB binding sites in the promoter regions of target genes known to be involved in the innate immune response of insects.

Introduction

Insects are coelomate metazoan hexapods that represent the largest and most varied group of animals on the planet with an estimated four million different species, of which only 950,000 have been described and classified (Chapman, 2005). Beetles (order: Coleoptera), butterflies and moths (order: Lepidoptera) and flies (order: Diptera) are the most numerous members of this class. Such remarkable diversity accounts for their evolutionary success, and has allowed them to conquer almost every ecological space on our planet. Insect evolutionary success cannot be attributed to a single feature of these animals but certainly their small size, ability to fly; relatively short generation time and potent immune system have had a positive impact on their survival.

In all of the diverse environments they inhabit, insects have been exposed to potential parasites and pathogens. Insects defend themselves from harmful organisms in three ways: first by behavioural avoidance, secondly by morphological physical defenses (cuticule and epithelia) and third, via a potent immune response. Over the past 25 years, insect immunology has been the subject of many mechanistic studies which have characterized the invertebrate immune response as innate (it lacks the antigen-antibody complex and the memory component characteristic of higher organisms). This response may be separated into two main components: cellular and humoral. The activation of the immune system relies on the very basic ability to recognize and discriminate self from non-self. In insects, this occurs by recognition of a set of highly-conserved molecular patterns produced by microorganisms which are absent from host cells by a set of host receptors. Microbe-associated molecular patterns (MAMPs), also known as pathogen-associated molecular patterns (PAMPs), are molecular motifs characteristic of certain organisms, for instance bacterial lipopolysaccharide (LPS), peptidoglycan (PGN) and lipoteichoic acid (LTA); or fungal $\beta1,3$-glucans. MAMPs are recognized by the insect immune response through a series of pattern recognition receptors (PRRs) which are constitutively

expressed and present in the hemolymph as soluble receptors or as cell surface transmembrane receptors. PRRs not only allow the host to recognize the foreign microorganism but also to discriminate among different types of microorganisms. Non-self recognition by PRRs rapidly results in a cellular and humoral response.

Cellular immunity is assumed by hemocytes, equivalent of our blood cells, which engage in different defence mechanisms depending on their lineage. In *Drosophila*, three lineages have been characterized: plasmatocytes which are specialized for phagocytosis, crystal cells, responsible for initiating the phenoloxidase (PO) cascade which leads to melanization, and lamellocytes, (found only in larvae) specialized for encapsulation of bacteria, parasitoids and other large foreign microorganisms (Strand, 2008).

Humoral defenses are characterized by a large number of antimicrobial peptides (AMPs) rapidly produced in the fat body, reactive intermediates of nitrogen or oxygen and a complex enzymatic cascade yielding in clotting or hemolymph melanization (Lemaitre and Hoffmann, 2007).

An insect's haemocoelic defence is mediated by a myriad of different tissues and free cells and biochemical cascades. Mounting such an immune response is costly and the host must regulate these mechanisms to avoid wasting valuable energetic resources. As is the case in other eukaryotic organisms, an insect cell or tissue only expresses a small subset of the genes encoded in its genome at all times, and is capable of turning on or off the expression of most other genes via complex signalling pathways. This differential gene expression governs a wide variety of biological processes including the immune response. In insects, gene regulation occurs primarily at the transcriptional level (Harshman and James, 1998). Transcription (the generation of mRNA using gene information from the genomic DNA as a template) is mainly regulated by two factors: a variety of *cis* regulatory DNA sequences found in the proximal 5' flanking sequence adjacent to the gene of interest; and by *trans* regulatory proteins named transcription

75

factors (TFs). These proteins maintain a modular structure characterized by a DNA binding domain and an activation domain. TFs bind their DNA target sequences as monomers, homo- or heterodimers to enhance mRNA generation mediated by RNA polymerase II.

A well characterized family of eukaryotic TFs is that of Nuclear Factor κB (NF-κB), named after the transcriptional activator of the immunoglobulin kappa light chain in human B-lymphocytes (Sen and Baltimore, 1986). A functional NF-κB transcription factor contains two members of the Rel family of proteins, characterized by a highly conserved 300aa N-terminal region. This portion contains the Rel Homology Domain (RHD) required for the formation of dimers, DNA binding, nuclear translocation and inhibitor binding. Rel proteins are constitutively expressed and are present in the cytoplasm as inactive zymogens bound to Inhibitor κB protein (IκB family) which masks the nuclear signal sequence. Upon release from IκB, Rel is translocated into the nucleus where it dimerizes and activates the transcription of its target genes by binding to a conserved nucleotide sequence element termed κB (Engstrom *et al.*, 1993). Members of the Rel/NF-κB family have been conserved through evolution from insects to mammals; in insects the first member of this family was described in *Drosophila melanogaster* and named Dorsal (Steward, 1987). Whereas Dorsal is involved in regulating the development of the dorsal-ventral axis of the Drosophila embryo, the two other members of this family of TFs, Dif and Relish are involved in activating the transcription of AMPs (Ip *et al.*, 1993, Dushay *et al.*, 1996). Rel/NF-κB members have been also identified in bees, mosquitoes, beetles and moths (Barillas-Mury *et al.*, 1996, Sagisaka *et al.*, 2004, Shin *et al.*, 2005, Tanaka *et al.*, 2005, Evans *et al.*, 2006), indicating their presence across insect orders.

In this study, we describe the molecular cloning of three different genes encoding members of the Rel/NF-κB family of transcription factors in *Rhodnius prolixus*, a hematophagous hemipteran insect and vector of Chagas disease. We also

evaluated their role in the immune response by evaluating their expression profile, identifying their putative target genes, confirming the recombinant protein activity and quantifying their activity in response to challenge with pathogens.

Material and Methods

Molecular cloning.

A 364bp EST from RhP-Dorsal was obtained initially by screening a fat body suppressive subtractive hybridization library in response to an artificial injection of *Trypanosoma cruzi* into the hemocoel of adult insects (Ursic-Bedoya and Lowenberger, 2007). Full length cDNAs were obtained using 5'-3' Rapid amplification of cDNA ends (RACE) with the Marathon cDNA synthesis kit (Clontech, USA) using three different PCR reactions with primers:

RpDoF1 (5'GACCATTGCAATCACGCGG3') – MgdT (5'CGGGCAGTGAGCGCAACGT$_{14}$ 3');

dgF-rhd(5'GRTTTCGSTACGAATGYGARGG3') –

RpDo-caR(5'AAGTTGTTCTAACTCTGACTGACCAC3') and

AP1 (5'CCATCCTAATACGACTCACTATAGGGC3') –

RpDo-R5 (5'GAGTTTTATGAATGAATCCGGTCCT3').

Reactions were performed with either Platinum Taq DNA polymerase (Invitrogen, USA) or iProof DNA polymerase (Bio-Rad, USA) under the following conditions: 95°C 60s; 95°C 10s; 60-65°C 15s; 72°C 40s. Annealing temperatures were modified according to the primer pair used. Subsequent cloning into pGem-Teasy vector, transformation into *E. coli* JM109 cells by heat shock, plasmid DNA isolation from recombinant clones using Wizard minipreps (Promega, USA), and DNA sequencing using BigDye v3.1 chemistry (Applied Biosystems, USA) was done as described previously (Ursic Bedoya *et al.*, 2005).

A similar approach was used to obtain the full-length cDNA sequence for β-GRP and Hemolymph proteinase using primer sets:

Rpβ-GRPF1: (5'AAGATTGAATGGACACCAGG3') – MgdT and

Rpβ-GRP -R3 (5'TGGGAATTCCACCAGACCTCCCAC3')–AP2.

RpHP F1: (5' TAGACAATCGTGACACGTTGCC 3') – MgdT to amplify the 3' end of the cDNA and RpHP-R1 (5' CGATATGCCTAGAAGTAACGAGC 3') – AP1 to amplify the 5'end.

The overlapping sequences were aligned using the SeqMan II module of Lasergene® v5 software (DNASTAR, USA) to generate the full cDNA sequence of all genes and to identify all putative open reading frames for all genes.

Transcriptional screening.

We had synthesized previously a panel of cDNAs from fat body and intestinal tissues in response to injection with bacteria or *T. cruzi* (Ursic-Bedoya *et al.*, 2008). We also extracted RNA and generated a cDNA from *R. prolixus* eggs using standard protocols as described previously (Ursic Bedoya *et al.*, 2005).

We screened these cDNAs with forward primers designed against the sequence in the 5'UTR of the three *R. prolixus* Dorsal-like molecules we found after sequence alignments of our clones. These primers were:

Do1C-F (5'TATTTCGGTGCGCACGTTGTGGTC3';

Do1A-F (5'ATGTGATGTGAGGATTAGTGTAGAG3'); and

Do1B-F (5'TAAGCTGTCGGTCGTAGGTTCC3').

Each of these forward primers was used with a common reverse primer DoStop (5'TTAGTTTACTTTTTGTTGTTCGACTG3') to screen the different cDNAs by PCR under the following conditions: 94°C: 1min, 30 cycles of 94°C: 20s, 65°C: 15s, 72°C: 2.5min in PCR reactions described above using Platinum Taq DNA polymerase (Invitrogen, USA). Products were visualized on a 1% agarose gel stained with ethidium bromide.

Genomic DNA extraction.

Genomic DNA was isolated from five starved adult insects. Insects were ground in a glass tissue grinder with 1.5mL of fresh DNA extraction buffer

(0.5%SDS, 0.2M NaCl, 25mM EDTA, 10mM Tris pH 8) and 1.5mL of phenol. After incubation at room temperature for 15 min, the homogenate was transferred to a 15mL Corex tube and centrifuged at 9200rpm for 20 min at 4°C. All centrifugation steps using 15mL Corex tubes were done on an Allegra 64R (Beckman Coulter, USA) centrifuge. The aqueous phase was transferred to a new tube to which an equal volume of phenol:chloroform (1:1) was added. The contents were thoroughly homogenized by vortexing followed by centrifugation at 8000rpm for 20 min at 4°C. The supernatant was again transferred to a new tube, mixed with an equal volume of chloroform, and centrifuged at 8000rpm for 15 min at 4°C. The resulting supernatant was transferred to a new tube and 1:10 volume of 4M ammonium acetate plus 2.5 volumes of 95% ethanol were added and then stored in a freezer at -20°C for 1h to precipitate the genomic DNA. After incubation, the DNA was pelleted by centrifugation at 8000rpm for 30 min at 4°C. The resulting pellet was dissolved in 100µL of EB buffer (10mM Tris-Cl pH 8.5). RNAse A treatment was done using 50 mg at 37°C for 30 min. After RNAse A treatment, the genomic DNA was further extracted with 100µL of phenol:chloroform:Iso-amyl alcohol (25:25:1); 100µL of chloroform and washed with 95% ethanol. The resulting pellet was resuspended in 100µL of 10mM Tris-Cl, pH 8.5 buffer and quantified using a Biophotometer (Eppendorf, Germany).

Target gene identification.

Inverse PCR (iPCR) (Triglia, 2000) was used to amplify upstream genomic regions of immune-related genes to identify potential NFκB binding sites. Restriction enzymes were chosen based on the cDNA or genomic DNA sequence of the target genes (β1-3 glucan recognition protein (β-GRP), Lysozyme, Defensin and hemolymph proteinase), so that the cut sites were within the first 500bp of the open reading frame or the available genomic clone. Restriction digest analysis was done using New England Biolab's NEBcutter v2.0 (http://tools.neb.com/NEBcutter2/index.php).

Inverse oriented primers were designed based on cDNA sequence from Lysozyme 1A, β-GRP, Defensin A identified from our suppressive subtractive

hybridization screening (Chapter 2). The genomic sequence for Hemolymph proteinase was obtained after sequencing 2kb of a 4.3kb amplicon obtained with primers RpHPMet 5'ATCATGATTAATCAATTATCC3' and RpHPstopR 5'GTACATCCTCCATAAGTTAGA3'designed to amplify the open reading frame of the gene.

One microgram of genomic DNA was digested with 10U of a single restriction enzyme in an air incubator at 37°C for 3h. Restriction enzymes used for each gene were: DpnI, EcoRV, RsaI for Lys1A; DpnI, EcoRI, EcoRV, RsaI for β-GRP; DpnI for Def A and BamHI, DpnI for Hemolymph proteinase. Restriction enzymes were heat inactivated whenever possible according to manufacturer's instructions (New England Biolabs, USA) or DNA was isolated by a phenol:chloroform extraction. Approximately 200ng of digested genomic DNA were self ligated with 12U of T4 Ligase (Promega, USA) in 100μL reactions at 16°C for 16h in a thermo cycler. Two microliters of the ligation reaction were used in a PCR reaction using iProof DNA polymerase (BioRad, USA). Primers pairs used in individual reactions are listed in Table 6-1. Amplicons obtained for each gene were cloned into pGemTeasy, transformed into *E. coli* JM109 cells, and sequenced or directly sequenced from the original PCR amplicon with BigDye v3.1 (Applied Biosystems, USA).

In addition to the molecular approach, we also used bioinformatic tools to confirm or obtain data iPCR was unable to provide. We searched for genomic clones in the recently released trace data archives of the *Rhodnius prolixus* genome sequencing project using Mega Blast searches (http://www.ncbi.nlm.nih.gov/blast/mmtrace.shtml) with the first 200-300 nucleotides of the open reading frame from every gene investigated. Contigs containing the identified genomic clones and the remaining of the open reading frames were constructed using the SeqMan II module of Lasergene® v5 software with loose assembling parameters to accommodate for large gaps corresponding to introns. Putative transcription binding sites were identified using Alibaba 2.1

software with lazy restriction parameters (available at: http://www.gene-regulation.com)

Recombinant protein expression.

RpDorsal 1C open reading frame or the N terminal end containing the rel homology domain was amplified using the primers:

LICadF (5'gacgacgacaagatgaaccaatctgttcggaga3') and

LICadR (5'gaggagaagcccggtttagtttacttttgttgttcg3') or

LIC-RHDadR (5'gaggagaagcccggttaagttaataattttgtattcg3')

respectively under the conditions 95°C: 1min, 30 cycles of 94°C: 20s, 60°C: 15s, 72°C: 45s. Amplicons were cloned into pET$_{32}$ expression vector by ligation independent cloning (LIC) as described in Novagen's pET System manual. Recombinant plasmid DNA was first transformed into non expression host *E. coli* NovaBlue cells, grown overnight in liquid LB broth containing Carbenicillin (50mg/mL), and then purified using the WizardPlus Miniprep DNA Purification System (Promega, USA). DNA sequencing of clones was done to confirm cloning into the correct reading frame prior to transformation into the bacterial expression host. Five nanograms of plasmid DNA were transformed into *E. coli* Origami 2(DE3) by heat shock following manufacturer's recommendations (Novagen, USA). The recombinant bacteria were plated on LB-Carbenicillin (50µg/mL) plates and incubated overnight at 37°C.

The next morning a single colony forming unit was used to inoculate 100mL of fresh LB-carbenicillin (50µg/mL) liquid media and grown at 37°C with vigorous shaking until OD$_{600}$≈ 0.6. Recombinant protein expression was induced by adding IPTG to a final concentration of 1mM and bacterial cultures were incubated at room temperature for 6h with shaking.

The resulting bacteria were isolated by centrifugation at 10,000g for 15 minutes and washed once with 20mM Tris-Cl, pH 7. The pellet was stored at -70°C. The recombinant protein was released from the bacterial cytoplasm after lysing the cells with 5 mL per gram of pellet of BugBuster reagent supplemented with 1KU

of rLysozyme and 0.1U of Benzonase per mL (all reagents from Novagen, USA) while shaking at room temperature for 20 min. After incubation the clear cell lysate was centrifuged 10,000g at 4°C for 20 min. The supernatant was transferred to a clean microcentrifuge or falcon tube while the pellet was resuspended in ½ supernatant volume of PBS; both samples were stored on ice. The protein content of each sample was estimated using Bradford reagent (BioRad, USA). Aliquots of all samples were resolved by SDS-PAGE to confirm presence of the recombinant protein in comparison to cell extracts from non induced recombinant bacterial cells.

In vitro activity.

Approximately 10μg of soluble cell lysate was used to assay the binding of the recombinant protein to the mammalian core NF-κB site: 5'GGGACTTTCC3'; using NoShift II Transcription Factor Assay System (Novagen, USA), a non-radioactive assay, alternative to standard electrophoretic mobility shift assays (EMSA). A biotin-labeled DNA probe consisting of a double-stranded consensus transcription factor binding site and a single-stranded capture region was incubated with the cell lysate containing the recombinant transcription factor. If functionally active, the transcription factor binds specifically to the double-stranded consensus sequence. Upon addition of a double-stranded DNA specific nuclease, the DNA probe bound to the transcription factor is protected from digestion whereas the unbound probe is degraded. The reactions then were transferred to a 96-well plate coated with the complementary strand to the capture region of the probe and the probe/transcription factor complex was captured on the plate. After one wash to remove unbound biotin and digested probe, streptavidin-alkaline phosphatase was added and allowed to bind to the biotinylated probe. A second wash step was followed by the addition of a chemiluminescent Alkaline Phosphatase substrate. Chemiluminescence detection was performed on a Victor 3V (Perkin Elmer, USA) microplate luminometer. The assay was replicated two times with every individual sample

run in triplicate. Assays were validated by ensuring that the percent digestion of the negative reagent control was above 90% as per manufacturer's suggestion.

Adult *R. prolixus* were inoculated with *Escherichia coli*, *Micrococcus luteus* or poked with a sterile needle as described (Ursic-Bedoya and Lowenberger, 2007). Two hours post inoculation; the fat bodies were dissected into approximately three volumes of NucBuster Extraction Reagent 1 (Novagen, USA). The tissue was homogenized with a sterile pestle, vortexed for 15s and incubated on ice for 5 minutes. Following a second vortex for 15s, the tissue was centrifuged at 16,000g, 4°C for 5 minutes. The supernatant was removed and the pellet was resuspended in 1/50 original tissue volume of 100x Protease Inhibitor Cocktail (Novagen, USA), 1/50 volume of 100mM DTT and 1.5 volumes NucBuster Extraction Reagent 2. The suspension was vortexed 15s, incubated 5 min. on ice, vortexed for another 15s and centrifuged at 16,000g, 4°C for 5 minutes. The supernatant, containing nuclear protein extracts was transferred to a new tube and the protein concentration was quantified by measuring light absorption at 280nm on a ND-1000 Spectrophotometer (NanoDrop Technologies, USA). All extractions were performed on ice. Three to eight micrograms of protein from the nuclear extractions were used in subsequent NoShift II NF-κB transcription factor assays that were performed as described above. Samples were assayed in triplicates and repeated twice independently.

Results

Recently, we isolated a 364bp EST from a suppressive subtractive hybridization (SSH) cDNA library made from *R. prolixus* fat body tissue 24h after injecting *T. cruzi* into the hemocoel (Ursic-Bedoya and Lowenberger, 2007). One of the putative reading frames of this EST shared significant amino acid identity (39%) with *Apis mellifera* Dorsal protein splice variant B (Blast X, E value: $1e^{-12}$). We used 5'-3' RACE to obtain the full length cDNA clones. DNA sequencing of fifteen independent clones revealed three different molecules with different 5' end sequences, particularly in the 5' UTR. Dorsal 1A mRNA is 2323bp long, and

encodes a putative 624 amino acid protein with an apparent molecular weight of 69.4kDa. Dorsal 1B is a differentially spliced version of Dorsal 1A, with a 75bp deletion within intron 1, which results in the use of an alternative start methionine codon. Dorsal 1B mRNA encodes for a putative 611 amino acid protein with an apparent molecular weight of 67.8kDa. Dorsal 1C mRNA has a different 5'UTR sequence from the other two mRNAs and uses a third methionine codon, common to all three mRNAs as its initial methionine start codon, and encodes a 579 amino acid protein of an apparent molecular weight of 64.6kDa (Figure 3-1). The *R. prolixus* dorsal nucleotide sequences reported in this paper have been submitted to the GenBank with accession numbers: EF634460, EF634461 and EF634462.

The deduced amino acid sequences of all three proteins contain conserved protein domains characteristic of members of the Rel/NF-κB family of transcription factors. Analysis with CDD (Marchler-Bauer *et al.*, 2005) and ScanProsite (Hulo *et al.*, 2006) identified three main protein domains common to all three proteins. The first is the Rel Homology Domain (RHD) (pfam 00554), characteristic of eukaryotic Rel/NF-κB transcription factors. The second conserved domain, C-terminal to RHD is shared by the immunoglobulins, plexins, and transcription factors and is termed the IPT domain (cd01177). This is an immunoglobulin-like fold domain, responsible for DNA binding. *R. Prolixus* Dorsal also contain a 19 amino acid lysine rich nuclear localization signal (NLS) responsible for targeting the free Dorsal molecules to the nucleus (Figure 3-2).

Of singular importance is the absence of C-terminal ankyrin (ANK) motifs, a 33 amino acid motif found in other members of the TF family such as Relish, consisting of two alpha helices separated by loops responsible for mediating protein-protein interactions. Multiple copies of this motif bind to RHD-IPT to inhibit its DNA binding activity and must be cleaved proteolytically to obtain an active TF (Stoven *et al.*, 2000).

To confirm the predicted protein activity of these molecules, we cloned and expressed Dorsal 1C in a bacterial expression system and used a non-

radioactive EMSA alternative assay to show binding to the consensus NF-κB sequence 5'GGGACTTTCC3'. Results (Figure 3-3) show luminescence levels resulting from digest protection of the target sequence by protein binding. Cell lysates obtained from bacteria expressing Dorsal 1C show luminescence similar to that of the positive control reagent provided by the manufacturer. Cell lysates from cells expressing just the RHD or from non-induced recombinant bacteria show very weak binding activity comparable to that of the negative control reagent. Differential splicing of some Rel/NF-κB TFs occur in the C terminal end before the RHD, suggesting this region is not critical for protein activity. Our results supports this statement as we expressed Dorsal 1C in pET_{32}, a bacterial expression vector which adds a 17kDa N terminal tag to the recombinant protein which enhances its solubility. The addition of this tag did not affect protein binding to its DNA target sequence, whereas expression of a recombinant protein missing the IPT domain lacked binding activity (RHD only). These results confirm the specific ability of *R. prolixus* Dorsal to bind to an NF-κB consensus site and suggest that the functional TF is a homodimer which requires the IPT domain for dimerization and/or DNA binding, in accordance with other members of the family.

As Rel/NF-κB transcription factors have been shown to be involved in different insect biological processes such as embryonic development and induction of the immune system, we screened a variety of cDNAs from different tissues and developmental stages to identify transcripts of these molecules. A different expression profile was discovered for each mRNA. Dorsal 1C was found to be expressed only in fat body tissues after immune challenge; suggesting its exclusive role in immunity. In contrast, Dorsal 1A and 1B seem to have a dual role as transcripts were found expressed constitutively in eggs and in midgut and fat body tissue, and also in response to non-self invasion of the hemocoel (Table 3-1).

To investigate the role of these Dorsal-like proteins in *R. prolixus* immune response, we searched for putative NF-κB binding sites in promoter regions of

several immune genes (β1-3 glucan recognition protein (β-GRP), Lysozyme, Defensin and hemolymph proteinase) identified in our SSH libraries (Ursic-Bedoya and Lowenberger, 2007). Inverse PCR gave variable results depending on the gene of interest (Figure 3-3). Whereas no amplicons were produced for the hemolymph proteinase molecule, β-GRP primers resulted in as many as five and as few as one amplicon per digested genomic DNA sample. However, sequence analysis and database comparison of the amplicons obtained did not identify any promoter sequence. The amplicons obtained for lysozyme and defensin were specific but yielded short (<100bp) 5' promoter sequence.

The genome of *R. prolixus* is being sequenced but some non-annotated sequence data is available on public databases at the NCBI trace data files. Genomic clones with high sequence identity for all our genes of interest were identified. When we assembled and annotated the available genomic sequences with our corresponding cDNA's into a single contig, we acquired an average of 500bp of DNA sequence upstream of the start ATG codon. These sequences were then used to search for putative TF binding sites. Bioinformatic analysis of the promoter regions using Alibaba 2.1 identified several putative TF binding sites. Of particular interest were NF-κB (5'-GGG$_{A/G}$AYYYYYY-3') and GATA (5'-$_{T/A}$GATAA-3') binding sites (Table 3-3). NF-κB sites were identified in the promoter regions of Lysozyme and Defensin. GATA sites were ubiquitous with the exception of the Defensin B gene. The orientation of these sites was sometimes reversed and the consensus sequence was found in the non-coding strand, suggesting a similar organizational feature to that described in *D. melanogaster* (Senger *et al.*, 2004). A third type of putative binding site, CCAAT/enhancer binding protein B (CEBP), was found abundantly in the promoter regions of all genes investigated (Data not shown). These sites have been shown to have immune roles in mammals and in mosquitoes the closely associated NF-κB and C/EBP sites function cooperatively to activate defensin genes (Meredith *et al.*, 2006) and possibly other genes.

Because AMPs with different activities had different putative binding sites in their promoter regions, we quantified the activity of NF-κB TF in the nucleus of fat body tissues in response to the injection of either *E. coli* (Gram -) or *M. luteus* (Gram +) by evaluating TF binding to the NF-κB target sequence. Results obtained with variable amounts of nuclear extracts are reported in Figure 3-5. The luminescence detected for nuclear extracts isolated from all our experimental samples were comparable to that of the negative control reagent which does not allow us to make any statement concerning the protein levels of NF-κB TF being up-regulated in response to bacterial infection. Such low activity levels could be explained partially by a low yield of nuclear extracts contaminated with cytosolic proteins, indicating a need to better optimize our extraction protocol, originally designed to be used with in vitro cultured cells.

Discussion

Differential splicing seems to be a common occurrence in TFs, including in the Rel/NF-κB family. In insects, differential splicing has been described for several members of the family belonging to different insect orders. *Drosophila melanogaster* has three different Rel/NF-κB TFs; Dorsal, Dif and Relish (Lemaitre and Hoffmann, 2007). Initially, the *Drosophila* transcription factor Dorsal, was described in playing a key role in the establishment of dorsoventral polarity in the early embryo. However, its alternatively spliced version, Dorsal-B has been suggested to be involved in the immune response as it is up-regulated in response to septic injury (Gross *et al.*, 1999). Dorsal and Dorsal-B are identical in the N-terminal region, however, Dorsal-B lacks the nuclear localization signal at the end of the RHD domain and the C-terminal end is completely different (Gross *et al.*, 1999). No differentially spliced versions of Dif and Relish have been described in *D. melanogaster*, however the *relish* gene encodes four transcripts, which originate from alternative start sites and result in different N terminal truncated proteins (Hedengren *et al.*, 1999).

In mosquitoes, no Dif orthologue exists, thus Dorsal is thought to substitute its immune related role along with differentially spliced versions of their respective Rel/NF-κB TFs. The malaria vector *Anopheles gambiae* has only two Rel/NF-κB genes, Rel1 and Rel2, which are homologues of *Drosophila*'s Dorsal and Relish respectively. Rel2 is differentially spliced into a shorter version lacking the ankyrin repeats and a death domain (Meister *et al.*, 2005). In *Aedes aegypti*, Relish has three alternatively spliced transcripts encoding different proteins. The predominant *Aedes* Relish protein contains both the RHD domains and the Inhibitor kappa-B like domain. The second version maintains the RHD domains but completely lacks the IκB-like domain. In the third transcript, a deletion replaces most of the N-terminal sequence and RHD however; the IκB-like domain is intact (Shin *et al.*, 2002). *Aedes aegypti* homologues to *D. melanogaster* Dorsal and *A. gambiae* REL1 have been identified and are encoded by two isoforms: AaREL1-A and -B. The transcriptional profile of both transcripts is similar, though their specificity for different NF-κB sites differs and results in the activation of different AMPs (Shin *et al.*, 2005).

Clearly, differential splicing in dipteran insects has a much more dramatic consequence for NF-κB TFs than those reported here for *R. prolixus* Dorsal 1A and 1B where only thirteen amino acids are lost, all in the 5'UTR region or in the initial coding region of the proteins, and no major protein domain is affected. However, a similar phenomenon has been recently described for *Bombyx mori* Rel proteins for which two isoforms RelA and B differ in the 5' sequence. Alternative splicing removes 241bp of transcript resulting in the loss of 52 amino acids of RelA (Tanaka *et al.*, 2005). What is noteworthy in these cases is that no functional domains seem to be removed by differential splicing. As is the case of *Drosophila*'s Dorsal and Dif and c-Rel, RelA, and RelB in vertebrates, *R. Prolixus* Dorsal proteins have N-terminal RHD and IPT domains while their C-terminal sequences contain transcriptional activation domains but no ankyrin repeats. This suggests that they do not require a proteolytic cleavage for activation.

The *in silico* approach used to identify promoters regions of target genes resulted much more efficient and yielded much longer promoter sequence than the *in vitro* approach using iPCR. The unsatisfactory results obtained for β-GRP may be related to primer design, having primers cross intron/exon junction, or a low stringency of PCR conditions

Analysis of immune genes involved in recognition (β-grp), activation (hemolymph proteinase) and effectors (defensin and lysozyme) revealed the presence of NF-κB sites only in the promoters of antimicrobial peptides (lysozyme and defensins), whereas GATA sites were present ubiquitously. NF-κB sites in *D. melanogaster* were discovered and shown early on to play a role in the inducibility of AMPs (Engstrom *et al.*, 1993). Now we know that these DNA motifs are critical for the inducibility of AMPs in both Toll and Imd immune signaling pathways in this and other insects. The role of GATA binding sites in insect immunity has always been described as being cooperative to proximal NF-κB sites (Kadalayil *et al.*, 1997) and required for tissue specificity (Petersen *et al.*, 1999, Senger *et al.*, 2006); thus finding them isolated in promoter regions of immune related genes, especially in genes lacking apparent NF-κB sites, is unusual and warrants further investigation.

The expression profile of *R. prolixus* Dorsal 1A and 1B was similar to that of *D. melanogaster* Dorsal B (Gross *et al.*, 1999) and *A. aegypti* Rel1 transcripts (Shin *et al.*, 2005) which are found in larvae and adults challenged with bacteria. In contrast, Dorsal 1C transcripts were only found in adults after bacterial insult.

However, differential expression of individual Rel/NF-κB TFs may not be as relevant for the insect's immune response. Reports in the literature paint a blurry picture concerning the specificity of response and activation of target genes by Rel/NF-κB TFs. Ectopic expression of Rel/NF-κB TFs to perform promoter analysis with reporter genes suggest that target gene regulation is based on the relative binding affinity of the TF for the NF-κB site. Our results indirectly agree with this as none of the NF-κB sites we identified matched exactly the core

consensus site sequence used in our functional assay suggesting for a certain degree of plasticity in this interaction. Moreover, a recent study based on ectopic expression of *D. melanogaster*'s Dorsal and Dif followed by microarray analysis suggests that some immune genes may be redundantly induced by different Rel/NF-κB TFs (Pal *et al.*, 2008).

The unraveling of immune gene regulation in *R. prolixus* is not as complete as for other insects, including the well characterized *D. melanogaster*, mosquito disease vectors, or model insects such as *Manduca sexta*. However, homology based studies have proven very valuable for the identification of related TFs and AMPs. The forthcoming release of the *R. prolixus* genome will greatly speed the discovery or the numerous missing pieces. As more genomes of pterygote, apterygote, holometabolous and hemimetabolous insects are annotated and released, we will have more opportunities to compare the evolution and regulation of immune-related genes and the role of transcription factors in immune gene regulation.

Figures

```
MNQSVRRTSESSGGGVASMPYIKIIEQPASKALRFRYECEGRSAGSIPGVNSTPE
NKTFPTIQIVGYRGRAVVVVSCVTKDSPYRPHPHNLVGKEGCKKGVCTVEINNET
MTAAFANLGIQCVKKKDIEEALRVREEIRVDPFRTGFSHKTQTSGIDLNSVRLCF
QAFLEGPQRGKFTNPLSPIVSEPIYDKKAMADLVICKLSHCSGSVAGGNEIILLC
EKVAKEDISVRFYEEKDGQVVWEGLGDFTPTQVHKQVAISFRTPRYKTLEIEQPV
QVLIQLRRPSDNATSEALPFQITPLDSGRPFFWSLRRSIGQKADYRTFSTILQTN
TKLLTNQERIDDDANNNTNNSNNNNEIKDVEVEDNVVSVADIPQPKVVVVSNQEE
EIKDEEENCGQSELEQLLVKDLENICNNNSDLGIYTSFQMALKNPCQDLLDHESY
EDIAPPRPTAAKPTLLQQPLSPVPHLDDEVLPPLPPKRAKRTGPDTKTLPSPPQK
KTGLFQKLFNTNKRKKNKEVQERPLQSRGSIKSLPAEMIISSNEVDVTEAEHYAL
YTSVAPHATASEFDEMSFYYSSVEQQKVN
```

Figure 3-1 Deduced amino acid sequence of _R. prolixus_ Dorsal 1C.

Dorsal 1C is the shortest isoform, 579 amino acids long with an estimated molecular weight of 64.6kDa. RHD domain in bold; IPT domain is shaded; an Asparagine rich region is in italics; Nuclear localization signal is underlined. No ankyrin motifs were found.

```
              10        20        30        40        50        60
    ---------+---------+---------+---------+---------+---------+
1   MNNLESFDFSLSRMDGQNIDESNSAINISDVIEVIETDCAAGGGLMNQSVRRTSESSGGG  RhP Do 1A
1   ------------MDGQNIDESNSAINISDVIEVIETDCAAGGGLMNQSVRRTSESSGGG  RhP Do 1B
1   --------------------------------------------MNQSVRRTSESSGGG  RhP Do 1C
```

Figure 3-2 *R. prolixus* Dorsal isoforms.

Only the first 60 amino acids are shown. Sequence information for all three molecules has been submitted to Genbank under accession numbers EF634460, EF634461 and EF634462.
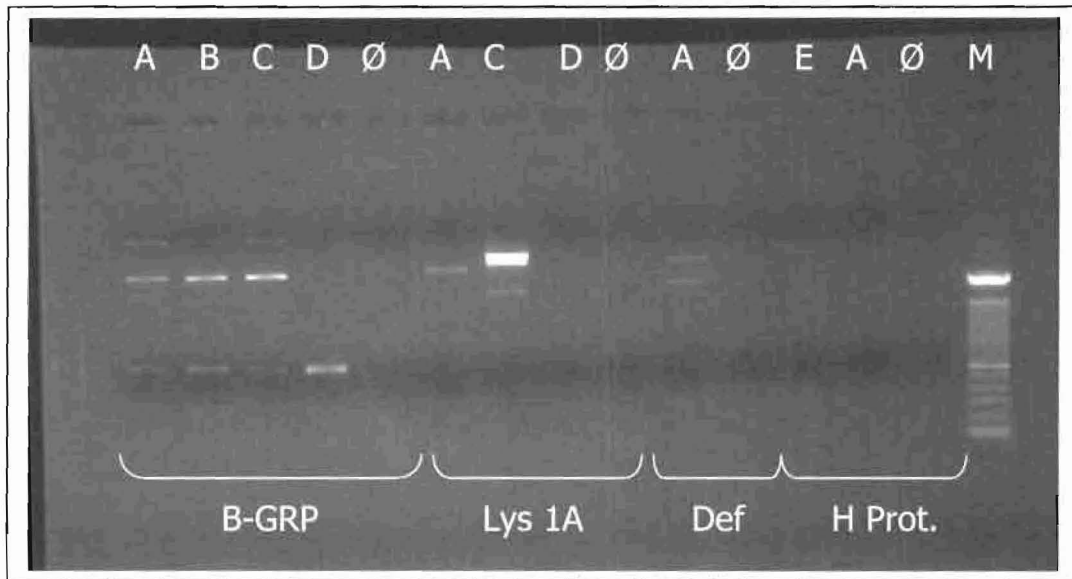
Figure 3-3 iPCR results.

Genomic DNA was digested with the respective restriction enzymes, self-ligated and then used in a PCR reaction with inverse oriented primers and a high performance DNA polymerase enzyme (iProof). Genes investigated were: β1-3 glucan recognition protein (β-GRP), Lysozyme, Defensin and hemolymph proteinase (HP). A: DpnI; B: EcoRI; C: EcoRV; D: RsaI; E: BamHI; Ø:No template control; M: 100bp DNA molecular weight marker.
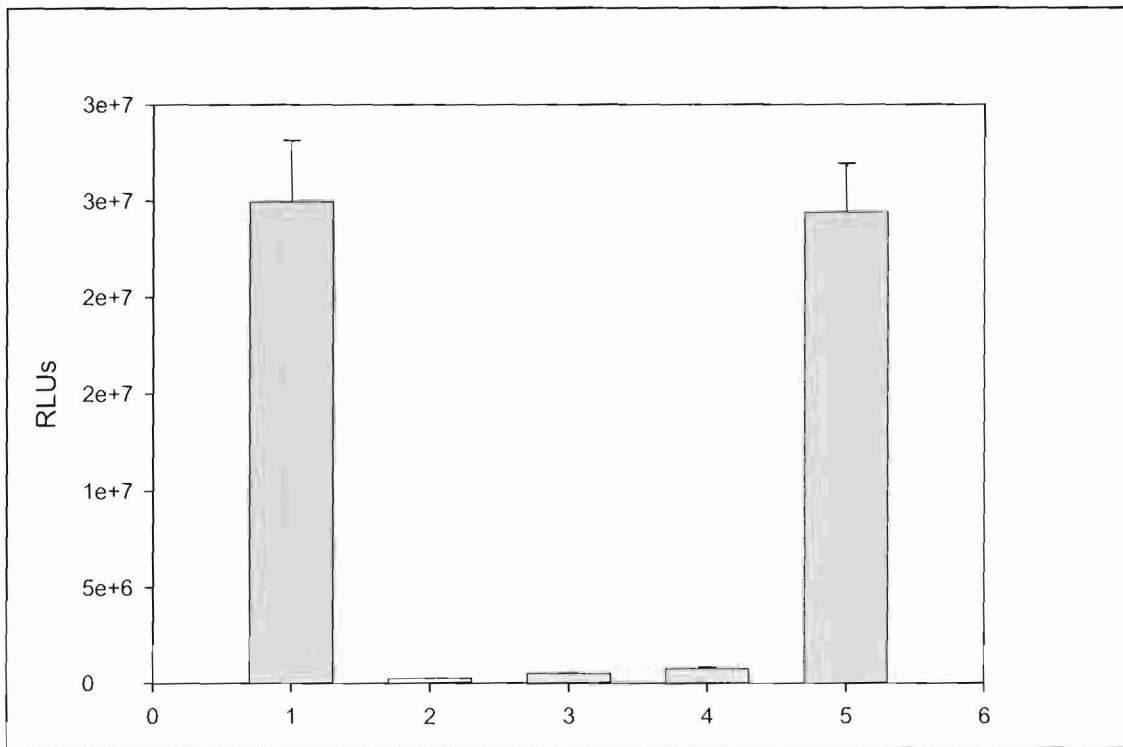
Figure 3-4 Recombinant protein binding activity.

Five micrograms of bacterial cell extracts were assayed for binding to the NF-κB sequence 5'GGGACTTTCC3'. 1: positive reagent control; 2: negative reagent control; 3: Un-induced cell extracts; 4: RHD Induced; 5: Dorsal 1C Induced. Y axis: relative light units (RLUs). Error bars represent s.d. of 2 independent trials in triplicate wells.
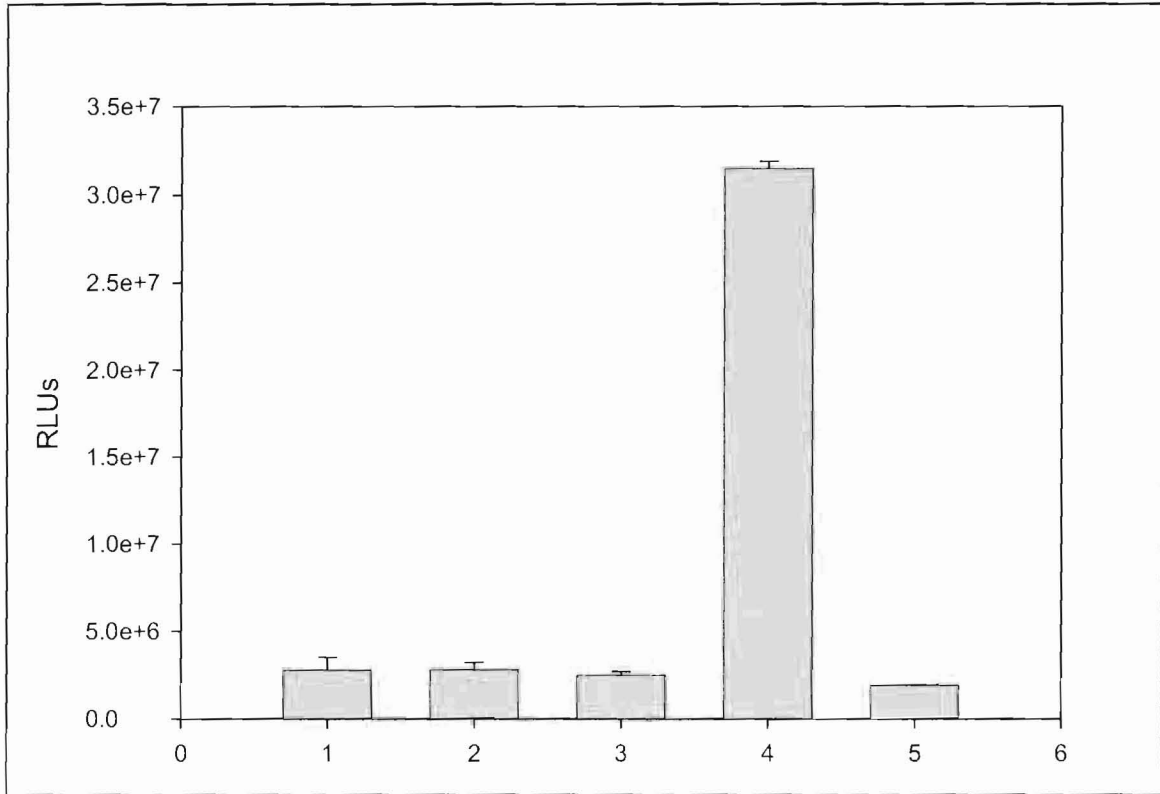
Figure 3-5 *R. prolixus* nuclear extract activity.

Nuclear extracts were isolated from fat body and assayed two hours post inoculation to quantify NF-κB activity. 1: *E. coli*; 2: M. luteus; 3: Sterile injection; 4: positive control reagent; 5: negative control reagent. Y axis: relative light units (RLUs). Bars represent s.d. of 3 independent trials with triplicate wells.

Tables

Table 3-1 Differential gene expression of *R. prolixus* Dorsal transcription factors.

cDNAs were screened with PCR primers designed to the 5' UTR and the stop codon as described in material and methods. +: presence; -: absence of amplicon.

	Egg	Midgut Naïve	Midgut 24h Bact.	FB Naïve	FB 24h Bact.	FB 24h Tc
Do 1C	−	−	−	−	+	+
Do 1A	+	−	+	+	+	+
Do 1B	+	−	+	+	−	−

Table 3-2 iPCR primer pairs.

Oligo Name	Sequence 5' to 3'	Tm °C	%GC
iPCR Lys1A F	AACTACGACGGAAGCTATGATAATGG	65.7	42.31
iPCR Lys1A R	TAGTGAACACCCTAGCTTGTGTGG	66.1	50
iPCR βGRP F	AGAATTAGAATATCTAGAAGCTGGCG	62.8	38.46
iPCR βGRP R	CAGAACATGTTGCTATGAAGAGG	62.5	43.48
iPCR DefA F	AGGTAACCGAAGAACATGTCGC	66.1	50
iPCR DefA R	GGCCACCAAGAACAGAGTAACC	65.7	54.55
iPCR HP F	ATTCTAGGCATAAACCAGGAGTG	62.1	43.48
iPCR HP R	TCCAAAGCAAACTAATCCGAC	62.8	42.86

Table 3-3 Putative transcription factor binding sites.

Putative transcription binding sites were identified using Alibaba 2.1 software using lazy restriction parameters. Alibaba predicts transcription factor binding sites by context dependent matrices generated from TRANSFAC 4.0 public sites. Location of the putative binding site is indicated relative to the methionine start codon. The clone indicators refer to the trace data files available at the NCBI trace data archives.

Gene	Clone	NF-κB	GATA
β-GRP	NAAX-acc45d02		GATATAAAAA -160 CATTAAGATTT -50
Hemolymph Proteinase	NAAD-aab82b10		TGATAATGTTT -35 TGTTTCAGATA -18
Defensin A	NAAX-adj13d02	TTCTTCCTCT -479 AAGAAAATCC -372 GGGATTCCCCC -162	CTATAAACAA -276
Defensin B	NAAX-abh21d09	TGGAATCCCC -166 GGATATTCCAC -30	
Lys 1A	iPCR Dpn1 NADD-aee07e10	GGAACTTTCAA -64 ATTAGGAAATAC -49	TGTTTCAGATC -115 CTTATATTTCT -42
Lys 1B	NAAX-ady62g11	TAGGAAATGAC -181	TTTGAGCAGAA -356 TTATTATTTTT -302

References

Barillas-Mury, C., Charlesworth, A., Gross, I., Richman, A., Hoffmann, J.A. & Kafatos, F.C., 1996. Immune factor Gambif1, a new rel family member from the human malaria vector, *Anopheles gambiae. The EMBO journal,* 15, 4691-701.

Chapman, A.D., 2005. *Numbers of Living Species in Australia and the World.* Toowoomba.

Dushay, M.S., Asling, B. & Hultmark, D., 1996. Origins of immunity: Relish, a compound Rel-like gene in the antibacterial defense of *Drosophila. Proceedings of the National Academy of Sciences of the United States of America,* 93, 10343-7.

Engstrom, Y., Kadalayil, L., Sun, S.C., Samakovlis, C., Hultmark, D. & Faye, I., 1993. kappa B-like motifs regulate the induction of immune genes in *Drosophila. Journal of molecular biology,* 232, 327-33.

Evans, J.D., Aronstein, K., Chen, Y.P., Hetru, C., Imler, J.L., Jiang, H., Kanost, M., Thompson, G.J., Zou, Z. & Hultmark, D., 2006. Immune pathways and defence mechanisms in honey bees *Apis mellifera. Insect molecular biology,* 15, 645-56.

Gross, I., Georgel, P., Oertel-Buchheit, P., Schnarr, M. & Reichhart, J.M., 1999. Dorsal-B, a splice variant of the *Drosophila* factor Dorsal, is a novel Rel/NF-kappaB transcriptional activator. *Gene,* 228, 233-42.

Harshman, L.G. & James, A.A., 1998. Differential gene expression in insects: transcriptional control. *Annual review of entomology,* 43, 671-700.

Hedengren, M., Asling, B., Dushay, M.S., Ando, I., Ekengren, S., Wihlborg, M. & Hultmark, D., 1999. Relish, a central factor in the control of humoral but not cellular immunity in *Drosophila. Molecular cell,* 4, 827-37.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P.S., Pagni, M. & Sigrist, C.J., 2006. The PROSITE database. *Nucleic acids research,* 34, D227-30.

Ip, Y.T., Reach, M., Engstrom, Y., Kadalayil, L., Cai, H., Gonzalez-Crespo, S., Tatei, K. & Levine, M., 1993. Dif, a dorsal-related gene that mediates an immune response in *Drosophila. Cell,* 75, 753-63.

Kadalayil, L., Petersen, U.M. & Engstrom, Y., 1997. Adjacent GATA and kappa B-like motifs regulate the expression of a *Drosophila* immune gene. *Nucleic acids research,* 25, 1233-9.

Lemaitre, B. & Hoffmann, J., 2007. The host defense of *Drosophila melanogaster*. *Annual review of immunology,* 25, 697-743.

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., Deweese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Yamashita, R.A., Yin, J.J., Zhang, D. & Bryant, S.H., 2005. CDD: a Conserved Domain Database for protein classification. *Nucleic acids research,* 33, D192-6.

Meister, S., Kanzok, S.M., Zheng, X.L., Luna, C., Li, T.R., Hoa, N.T., Clayton, J.R., White, K.P., Kafatos, F.C., Christophides, G.K. & Zheng, L., 2005. Immune signaling pathways regulating bacterial and malaria parasite infection of the mosquito *Anopheles gambiae*. *Proceedings of the National Academy of Sciences of the United States of America,* 102, 11420-5.

Meredith, J.M., Munks, R.J., Grail, W., Hurd, H., Eggleston, P. & Lehane, M.J., 2006. A novel association between clustered NF-kappaB and C/EBP binding sites is required for immune regulation of mosquito Defensin genes. *Insect molecular biology,* 15, 393-401.

Pal, S., Wu, J. & Wu, L.P., 2008. Microarray analyses reveal distinct roles for Rel proteins in the *Drosophila* immune response. *Developmental and comparative immunology,* 32, 50-60.

Petersen, U.M., Kadalayil, L., Rehorn, K.P., Hoshizaki, D.K., Reuter, R. & Engstrom, Y., 1999. Serpent regulates *Drosophila* immunity genes in the larval fat body through an essential GATA motif. *The EMBO journal,* 18, 4013-22.

Sagisaka, A., Tanaka, H., Furukawa, S. & Yamakawa, M., 2004. Characterization of a homologue of the Rel/NF-kappaB transcription factor from a beetle, *Allomyrina dichotoma*. *Biochimica et biophysica acta,* 1678, 85-93.

Sen, R. & Baltimore, D., 1986. Inducibility of kappa immunoglobulin enhancer-binding protein NF-kappa B by a posttranslational mechanism. *Cell,* 47, 921-8.

Senger, K., Armstrong, G.W., Rowell, W.J., Kwan, J.M., Markstein, M. & Levine, M., 2004. Immunity regulatory DNAs share common organizational features in *Drosophila*. *Molecular cell,* 13, 19-32.

Senger, K., Harris, K. & Levine, M., 2006. GATA factors participate in tissue-specific immune responses in *Drosophila* larvae. *Proceedings of the National Academy of Sciences of the United States of America,* 103, 15957-62.

Shin, S.W., Kokoza, V., Ahmed, A. & Raikhel, A.S., 2002. Characterization of three alternatively spliced isoforms of the Rel/NF-kappa B transcription factor Relish from the mosquito *Aedes aegypti*. *Proceedings of the National Academy of Sciences of the United States of America,* 99, 9978-83.

Shin, S.W., Kokoza, V., Bian, G., Cheon, H.M., Kim, Y.J. & Raikhel, A.S., 2005. REL1, a homologue of *Drosophila* dorsal, regulates toll antifungal immune pathway in the female mosquito *Aedes aegypti*. *The Journal of biological chemistry,* 280, 16499-507.

Steward, R., 1987. Dorsal, an embryonic polarity gene in *Drosophila*, is homologous to the vertebrate proto-oncogene, c-rel. *Science,* 238, 692-4.

Stoven, S., Ando, I., Kadalayil, L., Engstrom, Y. & Hultmark, D., 2000. Activation of the *Drosophila* NF-kappaB factor Relish by rapid endoproteolytic cleavage. *EMBO reports*, 1, 347-52.

Strand, M.R., 2008. The insect cellular immune response. *Insect Science,* 15, 1-14.

Tanaka, H., Yamamoto, M., Moriyama, Y., Yamao, M., Furukawa, S., Sagisaka, A., Nakazawa, H., Mori, H. & Yamakawa, M., 2005. A novel Rel protein and shortened isoform that differentially regulate antibacterial peptide genes in the silkworm *Bombyx mori*. *Biochimica et biophysica acta,* 1730, 10-21.

Triglia, T., 2000. Inverse PCR (iPCR) for obtaining promoter sequence. *In* M.J. Tymms (ed.) *Transcription factor protocols*. Totowa, NJ, 79 - 83.

Ursic-Bedoya, R.J. & Lowenberger, C.A., 2007. *Rhodnius prolixus*: identification of immune-related genes up-regulated in response to pathogens and parasites using suppressive subtractive hybridization. *Developmental and comparative immunology,* 31, 109-20.

Ursic-Bedoya, R.J., Nazzari, H., Cooper, D., Triana, O., Wolff, M. & Lowenberger, C., 2008. Identification and characterization of two novel lysozymes from *Rhodnius prolixus*, a vector of Chagas disease. *Journal of insect physiology,* 54, 593-603.

Ursic Bedoya, R.J., Mitzey, A.M., Obraztsova, M. & Lowenberger, C., 2005. Molecular cloning and transcriptional activation of lysozyme-encoding cDNAs in the mosquito *Aedes aegypti*. *Insect molecular biology,* 14, 89-94.

Connecting statement 3

In the previous chapter, we reported the cloning and characterization of three different transcripts homologues of the *Drosophila melanogaster* transcription factor Dorsal. *Rhodnius prolixus* Dorsal 1A, and 1B were found to have a dual role in early development and immunity, whereas Dorsal 1C appears to be involved solely in immunity. Dorsal proteins are known transcriptional activators of antimicrobial peptides (AMPs), the effector arm of humoral immunity in insects. Among the AMPs we found that contained the necessary DNA motif to confer inducibility by Rel/NF-κB transcription factors was defensin. In chapter 4, we describe a combination of biochemical and molecular methods to characterize three isoforms of *Rhodnius prolixus* defensins, perhaps the most evolutionary conserved AMP as it has been identified in mammals, plants and insects.

CHAPTER 4: ISOLATION AND CHARACTERIZATION OF A NOVEL INSECT DEFENSIN FROM *RHODNIUS PROLIXUS*, A VECTOR OF CHAGAS DISEASE

LOPEZ L.[a], MORALES G.[a], URSIC BEDOYA R.[b], WOLFF M.[a], LOWENBERGER C. [b]

[a]Universidad de Antioquia, Instituto de Biologia Calle 67 No 53-108, Medellin, Colombia
[b]Simon Fraser University
Department of Biological Sciences
Burnaby, British Columbia, V5A 1S6, Canada

As third author, I was responsible for: isolating RNA from tissues, generating cDNAs, using 5'-3' RACE to obtain the full length of the gene, cloning, DNA sequencing and phylogenetic analysis.

Abstract

An antimicrobial peptide belonging to the defensin family of small cationic peptides associated with innate immunity in insects was isolated from the hemolymph of *Rhodnius prolixus*, a vector of Chagas disease. This peptide, designated *R. prolixus* Defensin A, was purified and sequenced. The active peptide contains 43 residues and aligns well with other insect defensins. However the pre-pro region of the sequence has little shared identity with other insect defensins. We identified 3 isoforms of *R. prolixus* defensin from cDNA clones obtained from RNA isolated from the whole bodies of immune activated insects. Northern analysis and Real-Time Quantitative PCR indicated that there was a very low baseline transcription of this peptide in naive insects, and that transcription increased significantly in the fat body of immune-activated insects. In addition, transcription of this peptide occurred in the intestine 24h post activation suggesting that the midgut/intestine of this species is active in the immune response against pathogens.

Introduction

Insect immunity is the rapid, germ-line encoded anti-infection response insects use to protect themselves from pathogens and parasites (Boman, 1998). A major component of this immune response is the production of a potent arsenal of immune peptides, most of which are produced in the fat body or hemocytes of the insect and released into the hemolymph. This immune response is based on the recognition of the pathogen as nonself, the induction of suitable genes and biochemical pathways that result in the production of potent antimicrobial peptides (Barillas-Mury *et al.*, 2000, Hoffmann and Reichhart, 2002). Immune peptides are expressed de novo, usually in the fat body and hemocytes (Dimarcq *et al.*, 1994), and delivered to the appropriate site to defend the host in a manner that is neither learned nor acquired, unlike the hallmark characteristics of classical immunology (Boman, 1998).

Innate immunity is not limited to insects; immune peptides are highly conserved members of the innate response of highly diverse taxa, including single celled organisms (Leippe, 1999), various classes of invertebrates (for a review see (Hetru *et al.*, 1998), plants (Broekaert *et al.*, 1995) and vertebrates (Lehrer and Ganz, 1996). However, the majority of studies over the last 10 years have been carried out on insects, especially *Drosophila melanogaster* in which strong similarities have been demonstrated between the signal transduction pathways of the innate immune response of insects and the acute phase response of vertebrates (Hoffmann *et al.*, 1999). There are distinct advantages to the innate response; anti-microbial peptides can act with low specificity against a wide range of microorganisms, there are no memory requirements, they can be synthesized at relatively low metabolic cost to their hosts, and can be stored in high concentrations (Shai, 1998). A keystone element of innate immunity is the speed with which these responses occur: transcripts can be found within minutes of stimulation, and proteins found within hours.

Several studies have suggested that inducible immune peptides can limit parasite development in vectors (Jaynes *et al.*, 1989, Rodriguez *et al.*, 1995, Lowenberger *et al.*, 1996, Possani *et al.*, 1998, Shahabuddin *et al.*, 1998, Lowenberger *et al.*, 1999b, Lowenberger, 2001). Most of these studies have looked at vectors, such as mosquitoes, in which the parasites have direct contact with hemolymph factors as they move from their site of development to the mouthparts or salivary glands for transmission to vertebrates.

In Chagas Disease vectors, however, the parasites never leave the intestinal tract, and there is no such direct contact between parasite and hemocyte or hemolymph factors. Instead, the ingested parasites form epimastigotes in the vector midgut and metacyclic trypomastigotes that inhabit the hindgut and rectum (Brener, 1973, Garcia and Azambuja, 1991, Azambuja *et al.*, 1999, Garcia *et al.*, 1999). As the insect feeds and engorges, a fecal droplet containing infective parasites is deposited on the skin of the host and the parasites then are rubbed into the puncture or existing skin lesions. We hypothesize that this apparently inefficient, but successful mode of transmission has evolved to avoid contact with lethal components of the vectors' immune response. To test this hypothesis, we first must identify the repertoire and activity spectra of the inducible immune response in vectors of this parasite. We report here the isolation and characterization of a member of the insect defensin family from *Rhodnius prolixus*, a major vector of Chagas disease.

Material and Methods

Insect maintenance, immune activation and hemolymph collection.

A colony of *R. prolixus* has been maintained at the Institute of Biology, Universidad de Antioquia, Medellin, Colombia for over 5 years. Bacteria (*Escherichia coli* and *Micrococcus luteus*) were grown at 37°C as described previously (Lowenberger *et al.*, 1995). Overnight cultures of each bacterial culture were combined, pelleted, and the supernatant discarded. A minuten pin (0.15mm) was dipped into the moist bacterial pellet and inserted into the

hemocoel of 4th and 5th instar nymphs of *R. prolixus*. At 6, 12, and 24 h post inoculation, a minuten pin was inserted through the thoracic pleura of the insects, and the exuding drop of hemolymph was collected. Approximately 10µl of hemolymph was collected from each insect and the hemolymph from 30 animals was combined. Because there was rapid melanization of hemolymph, the material was stored in an ammonium acetate buffer pH 3.5 (final concentration 25mM), supplemented with PMSF (final concentration 1.5µM) as a protease inhibitor, and phenylthiourea (final concentration 20µM) as a melanization inhibitor.

Peptide isolation and antimicrobial assay

After gentle stirring for 30 min in an ice-cold water bath the samples were centrifuged at 10,000g for 20 min at 4°C. The supernatant was loaded onto a C18 cartridge (Burdick and Jackson) previously equilibrated with acidified water (0.05% trifluoroacetic acid) and stepwise elution performed with 2, 40, and 80% Acetonitrile (ACN) in acidified water (Bulet *et al.*, 1993). The 40% fraction was analyzed by reverse-phase HPLC on an Aquapore 300 C8 column (250 × 7.0mm) using a linear gradient of 2–100% ACN in acidified water for 160 min at a flow rate of 1.3 mL/min. The fractions were collected every minute during the run, lyophilized, and re-suspended in 100µL of distilled water. Antimicrobial assays were carried out in two ways. Initially, sterile discs of Whatman filter paper were impregnated with each fraction and placed on an agar plate previously overlayed with either *M. luteus* or *E. coli*, and incubated at 37°C. The plates were assessed for halos of growth inhibition for each fraction 24h later. Subsequently isolated fractions were assayed for activity in a liquid assay as described previously (Hetru and Bulet, 1997). Active fractions were further purified using a biphasic gradient of ACN in acidified water from 2–25% over 10 min and 25–32% over 40 min at a flow rate of 0.2mL/min. For all HPLC analyses, the effluent was measured at 225nm and fractions hand collected and concentrated under vacuum.

Capillary zone electrophoresis and microsequence analysis.

The purity of peptides was ascertained by capillary zone electrophoresis as described previously (Lowenberger *et al.*, 1995) using a 3D Hewlett Packard Capillary electrophoresis system equipped with silica capillary. Separation was done from anode to cathode in 20mM citrate buffer at pH 2.5. Detection was carried out at 30°C, at 200nm. Two very active fractions peaks were transferred to a PVDF membrane, and submitted to the Medical College of Wisconsin for peptide sequence analysis via Edman degradation.

RNA isolation and cDNA determination.

Total RNA was collected from whole bodies, fat bodies, or intestinal tracts of immune activated or non-inoculated (control) *R. prolixus* at 0, 6, 12, 24h post inoculation using TRIreagent (Molecular Research Center) following the manufacturers instructions. Total RNA was quantified using a Biophotometer (Eppendorf, Germany) and 1μg of total RNA was reverse transcribed as described previously (Lowenberger *et al.*, 1999c) using the primer 5'-CGGGCAGTGAGCGCAACGT14- 3'. Degenerate PCR was carried out using the RT primer and two forward degenerate primers designed to amplify the sequence encoding the amino acid sequence we obtained via Edman degradation. The primer sequences were: Primer1: 5'GTNACNCCNAAYCAYGCNGG3';

Primer2: 5'GCNCAYCAYYTNTTYMGNYTNGG3'. PCR conditions were 95°C-3 min, and 30 cycles of 95°C (10 s), 50°C (10 s) 72°C (30 s) followed by a 5 min extension period at 72°C on an Idaho Technologies Indy Cycler (Idaho technologies, USA). PCR products were size-fractionated on a 1.2% low melting point agarose gel and visualized on a BioDoc gel documentation system (UVP, USA). Bands of the predicted size were excised from the gel, placed at 65°C to liquefy the fragment, vortexed briefly, and cloned directly into P-GEM-T vector (Promega, USA) following the manufacturer's instructions. Blue-white screening of XL1-Blue cells (Stratagene, USA) was used to identify potential transformants. These colonies were grown overnight in 5mL LB medium with 5μL ampicillin (100μg/uL) and purified using the Wizard Plus Miniprep DNA Purification System

(Promega, USA). Sequencing of these clones was carried out on an ABI 310 sequencer (Applied Biosystems, USA) using Big Dye chemistry. Sequences were compared with known sequences in the NCBI database. We designed specific primers for the defensin clone based on the sequences obtained from the degenerate primer PCR. Full length cDNA sequences were obtained using the Marathon cDNA synthesis kit (Clontech, USA) using our specific primers and the flanking primers obtained with the kit. PCR amplification, cloning, transformation and sequencing were carried out as described above. Phylogenetic analysis and multiple alignment of *R. prolixus* defensins with defensins from other sources was carried out using DNA Star software (DNASTAR, Inc USA) using Clustal W method with PAM 250 matrix.

Northern analysis.

Northern analysis was performed as described previously (Lowenberger *et al.*, 1999a) using 5μg of total RNA from whole bodies and specific tissues of control or immune-activated insects. RNA was separated on a formaldehyde-agarose gel (Sambrook *et al.*, 1989), transferred to a nylon membrane, and UV crosslinked. ^{32}P probes were made using 50ng of the entire coding region (283 bp) of *R. prolixus* isoform A in a PCR reaction described previously (Severson and Kassner, 1995). Membranes were hybridized with the probe, and subsequently with a ^{32}P-labelled probe made from a 200bp fragment of an actin gene isolated from *R. prolixus*. Preparation of the probe, removal of free dNTPs, hybridization conditions and washes were carried out as described previously (Lowenberger *et al.*, 1999c).

Real time quantitative PCR

For a more precise estimation of comparative transcription rates, Real-Time Quantitative PCR (qPCR) was used. Reverse transcription was done as described above with 1μg of RNA from whole bodies or specific tissues. Initially, 1μL of the RT reaction was used in a QPCR reaction using primers to amplify a partial fragment of a *R. prolixus* actin (Brackney, Lowenberger, and Wolff

unpublished) as a control for similar amounts of cDNA in each sample. Samples were run on a BioRad iCycler machine under the conditions: 95°C (2 min), and 40 cycles of 95°C (0.5 min), 62°C (0.5 min) 72°C (1 min). The PCR reagents were similar to the regular PCR with the addition of 1μl of a 1/1000 dilution of Sybr-Green I (Sigma, USA) to measure the amounts of double stranded DNA produced in the reaction and 2.5μl of a 1/1000 dilution of fluorescein to control for background fluorescence. Sample volumes were adjusted in subsequent PCR reactions to ensure similar amplification profiles for actin. Subsequently these volumes of the cDNAs were used to amplify a partial fragment of our defensin sequence. qPCR was carried out on samples collected from different batches of immune stimulated or naïve insects.

Results

In this study, we assessed the hemolymph of immune activated *R. prolixus* for the presence of antimicrobial peptides. Based on our previous studies with mosquito defensins (Lowenberger *et al.*, 1995) we concentrated our analysis on the fractions eluted with 40% ACN. This fraction was subjected to RP-HPLC analysis (Figure 4-1A).

Fractions (1mL) were collected each minute during the run, dried, resuspended in ddH2O, and assessed for antibacterial activity using the impregnated Whatman technique (Figure 4-1B). Active fractions were subjected to a second reverse-phase chromatography, and fractions tested against *E. coli* and *M. luteus*. Capillary electrophoresis of one active peak indicated two bands that were separated on a gel, transferred to PVDF, and sequenced. We obtained a sequence: VTPNHAGCAHHLFRLGNRG. This was submitted to the NCBI BLAST program for comparisons with other reported peptides. The highest match (57% identity, 72% positives) was a segment of the insect defensin from *Pyrrhocoris apterus* (Figure 4-2). We performed PCR on the cDNA using the degenerate primers described in the Materials section. The initial amplification of a fragment of *R. prolixus* defensin cDNA using primer 1 produced clones which

indicated three distinct cDNA sequences (Figure 4-3). We then designed 3 reverse primers that would distinguish between these three isoforms for use during the 5' RACE protocols, using the Marathon kit (Clontech, USA) to amplify full length cDNAs. The cDNAs contain a 5'UTR of differing lengths: 46bp (isoform A), 56bp (isoform B) and 51bp (isoform C). All isoforms have a pre-pro defensin of 153bp beginning with the ATG start codon and terminating with AAG AGA, which translates to the K-R cleavage site present in many insect defensins.

The coding region for the mature defensin is 129bp (that translates to a peptide of 43 residues) in all isoforms, a stop codon TGA, followed by a 3' UTR of differing lengths: 117, 113, and 129 bp, respectively, for isoforms A, B, & C. The cDNAs for isoform A and C have a poly-adenylation consensus sequence (AATAAA) 16bp before the poly-A tail, whereas isoform B has no such signal (Figure 4-3).

We aligned the sequences of the mature peptide obtained with sequences available in GenBank (Figure 4-2A). There is a strong conservation of the size of defensin within the insects and the position of the conserved 6 cysteines typical for members of this family of immune peptides. A phylogenetic alignment of the mature peptides (Figure 4-2B) demonstrated that the phylogenetic relationship of the Defensins paralleled the relative evolutionary placement of the individual organisms in the three of life.

Transcriptional Profile: Northern analysis (Figure 4-4) indicates a low level of transcription in naive insects, and 6 h after inoculation a weak response in the intestine of *R. prolixus*. However by 24 h after inoculation there is a strong transcriptional activity for defensin in the intestine. In contrast, the fat body shows a high level of transcription 6 h post inoculation and remains high for 24h. In Real Time quantitative PCR analysis we measured transcript presence as a percentage of the controls (Figure 4-5). The levels found in naive midguts and fat bodies were arbitrarily given a value of 1. Whereas transcription in immune activated midguts did not increase over controls at 6h post stimulation, there was a 7-fold increase 24h post stimulation. In the fat bodies there was a 25- and 29-fold increase in defensin transcripts at 6 and 24 h, respectively, after immune

stimulation. These data indicate that the fat body is the major immune organ in the insect as has been reported in other insect systems. However the midgut, where pathogens and obligate symbionts reside, is also an immune competent tissue.

Discussion

The data presented here establish that the cell free hemolymph of immune-challenged *R. prolixus* contains several compounds with activity against bacteria. One of these is a member of the Defensin family, which are ubiquitous immune peptides described from several groups of invertebrates, plants and vertebrates. In common with all preprodefensin sequences there is a definite signal peptide region (residues 1–24), a prodefensin region (residues 25–51) that terminates with a KR cleavage site. This conserved cleavage site is common among the insects (Hetru *et al.*, 1998, Lowenberger *et al.*, 1999c). A comparison with established peptides in the databanks established that our sequence is a member of the insect defensin family. At the mature protein level the *R. prolixus* peptide A shares 88 and 77% identity with defensins isolated from *P. apterus* and *Palomena prasina* respectively, both of which are members of the same insect order, the Hemiptera, and 66% shared identity with the Coleopteran, *Oryctes rhinoceros*. However we failed to obtain any positive comparisons when we subjected only the pre-pro region of our peptide to the databases, either as nucleotide or as the translated peptide sequences. Many of the sequences in the databases for closely related insects are for the mature peptide sequence only of the defensins, and this may limit the number of comparisons available. These data suggest that whereas the mature peptide region has been highly conserved through evolutionary time, the prepro region of this peptide has been modified from a precursor molecule, and it is possible that *R. prolixus*, and possibly other members of the true bugs (Homoptera/Hemiptera), have a different pre-pro peptide than other insect orders. The exception is the KR cleavage site at the end of the pre-pro region (Figure 4-3) that is highly conserved in *R. prolixus* as

112

well as in several other very diverse species for which full-length cDNAs are available. These data and the phylogenetic analysis (Figure 4-2B) suggest a very strong conservation of the mature peptide sequence among all organisms, especially the location of the 6 cysteine residues that form 3 disulphide bridges. Such conservation in all likelihood exists due to the importance of defensins as key components of the very effective immune response of invertebrates to pathogen invasion that has allowed these organisms to thrive in environments full of potential pathogens. We tested an antibody raised against *A. aegypti* (data not shown) Defensin A peptide on the hemolymph collected from naive and immune activated *R. prolixus*, and found a positive band at the expected location, albeit at a lower signal intensity that hemolymph collected from immune activated A. *aegypti*. The *A. aegypti* and *R. prolixus* peptides share approximately 55% identity, and the positive Western result indicates a high level of conservation of the peptide structure of defensins from different orders of insects. In addition, we have isolated, by PCR, three genomic sequences containing introns of 90–97bp. The introns are located in the same position in all three sequences; in all three sequences introns are found within the codon coding for amino acid D31. This is similar to the situation with *Ae. aegypti* in which three isoforms of insect defensins were isolated (Lowenberger *et al.*, 1995, Lowenberger *et al.*, 1999a). Similarly, there is a gene cluster encoding three cecropin genes in *Hyalophora cecropia* (Gudmundsson *et al.*, 1991), a compact gene cluster for three cecropin genes in *D. melanogaster* (Kylsten *et al.*, 1990) and it is conceivable that the same situation occurs in *R. prolixus* defensins as Southern analysis suggests there are three defensin genes in *R. prolixus*. We will be able to resolve these questions by obtaining full length genomic clones through screening a genomic library. Multiple sequence alignment (Figure 4-2) of mature defensin peptide amino acid sequences demonstrates a high level of conservation within closely related phylogenetic groups. Because defensin is such an ancient peptide found in a wide array of organisms, we would expect that peptides from more closely-related organisms would be most similar. However, as we encounter novel peptides, the similarities exhibited may reflect diet, environment, or previous

113

pathogen exposure as opposed to taxonomic associations. Transcription of defensin in *R. prolixus* was determined 0–24h after inoculation with bacteria. Northern analyses were carried out using a probe of the entire 282bp coding region of our clone. RpDEF-A is contained within a 463bp message that is detectable 8h after inoculation, and increases in strength by 24 h. Using primers designed against *R. prolixus* defensin sequences we evaluated, using Q-PCR, transcript levels in whole bodies and specific tissues of naive or immune activated nymphs at various times after immune challenge (Figure 4.5). We can detect defensin transcripts at very low levels in naive insects as was demonstrated in mosquitoes (Lowenberger *et al.*, 1999a). However, there is a significant increase due to immune activation. Immune activation not only increases transcription significantly in the fat body as we would expect, but also increases defensin transcription 4 fold in the midgut/intestine 24h after activation, whereas there is little increase in midguts/intestine 6h after stimulation suggesting that indeed the midgut/intestine is an immune responsive tissue. There is not the speed of transcription, nor the same quantity of message in the midgut/intestine that we see in the fat body. However, in these insects bacteria were injected into the hemocoel, and we would expect an extremely rapid and strong response in this important immune responsive tissue. These data suggest that there may be a systemic factor in the *R. prolixus* immune response in which tissues not stimulated directly by the inoculation (intestine) have a delayed transcription of immune peptides. This might be due to a cascade of signaling molecules released in response to the immune activation process. Alternatively, injected bacteria or other products of bacterial degradation may interact directly with the hemocoel side of the midgut/intestine inducing the expression of defensins. Because the transcriptional response in these tissues is less than that of the fat body, and assuming all transcripts are translated, it is possible that parasites in the intestine may be protected from lethal concentrations of these peptides. In addition, high levels of immune peptides in the midgut may be not be advantageous if the obligate bacterial symbionts that reside there are susceptible to the immune peptides. We may hypothesize that the relegation of *T. cruzi* to its

lifecycle in the intestine of *R. prolixus*, and its relatively inefficient mode of transmission, may be an evolutionary consequence of its susceptibility to insect immune peptides. Durvasula et al. (1997) demonstrated that the presence of bacterial symbionts engineered to express cecropin A in the midgut lumen of *R. prolixus* had no traces of *T. cruzi*, whereas individuals containing non-engineered symbionts were positive for the parasite. However to test this hypothesis requires that we explore more fully the repertoire of immune molecules in this species, their ability to be produced in the intestine, or to be transported to this region, and their ability to affect the development of the parasite in vivo. The characterization of the defensin described here and the identification of novel immune peptides in vectors such as *R. prolixus* will increase our knowledge of general aspects of insect immunity. In addition, expanding these studies to vectors with different modes of parasite transmission will allow us a greater understanding of the mechanisms used by these insects to protect themselves, and the stresses placed on the parasites within the insect vectors.

Acknowledgements
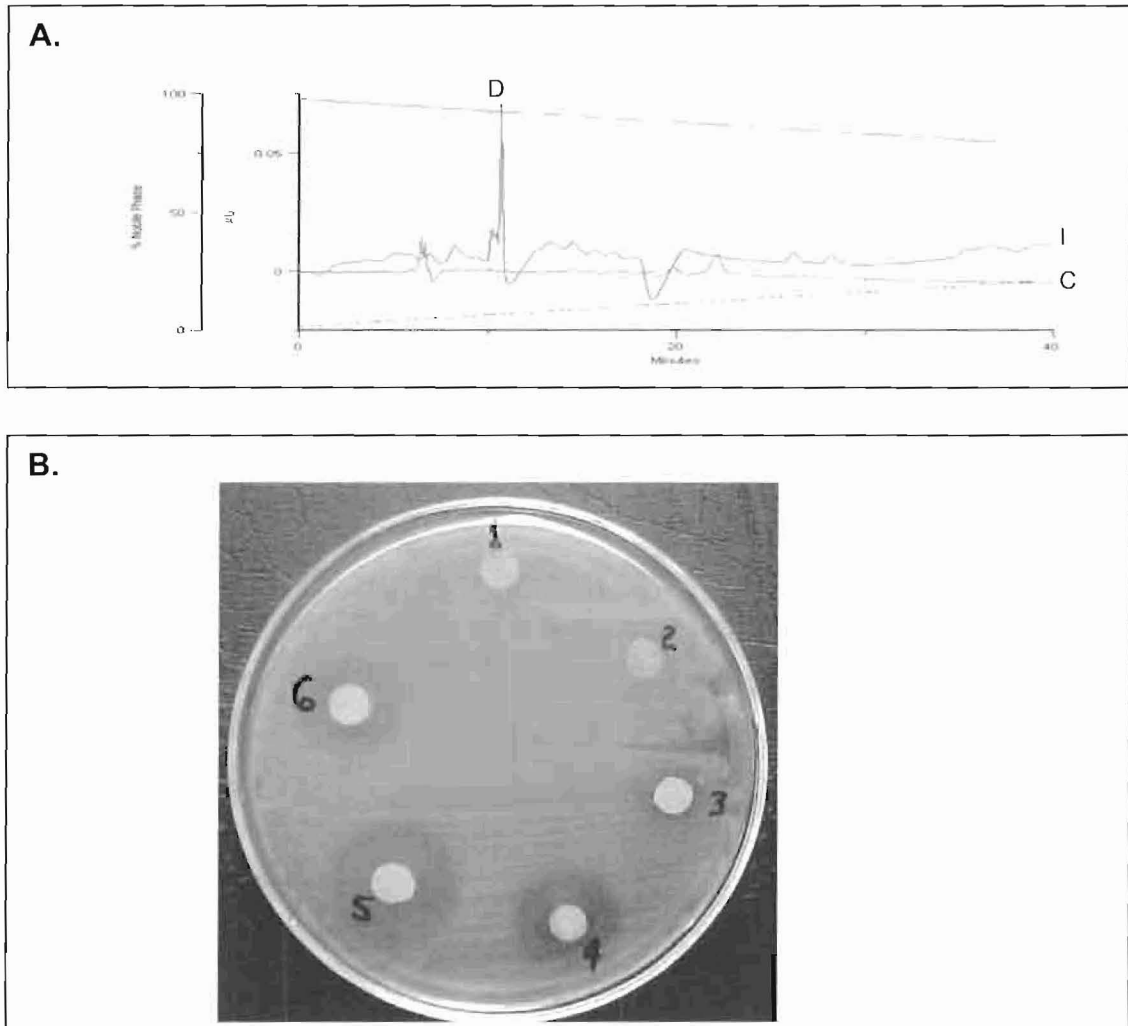
Figures

Figure 4-1 Biochemical isolation of *R. prolixus* Defensin.

A: Reverse phase chromatograph of hemolymph collected from immunized (I) or control (C) *Rhodnius prolixus*. The peak labeled "D" contains the *R. prolixus* defensin. Pre-purified samples were applied to a reverse-phase C18 column and eluted with an acetonitrile gradient. Fractions (1mL) were collected each minute, dried under vacuum and resuspended in 100µl of water. Sterile discs were impregnated with 25µl of this liquid and placed on an agar plate previously seeded with either *E. coli* or *M. luteus*.

B: HPLC fractions were collected, dried and resuspended in ddH2O and used in disk diffusion assays against *E. coli* and *M. luteus* as described in Material and Methods. These qualitative assays were used solely to identify fractions with activity. Fractions demonstrating activity were reanalyzed and individual peptides obtained.

A:

```
                        Loop              α-helix       β-sheet
                    |-------------------|----------|-----------------
R. prolixus A       ATCDLFSFRSKWVT---P---NHAACAAHCLLRGNRGGRC-KGTICHCRK
R. prolixus B       ATCDLLSFSSKWVT---P---NHAGCAAHCLLRGNRGGHC-KGTICHCRK
R. prolixus C       ATCDLLSLTSKWFT---P---NHAGCAAHCIFLGNRGGRC-VGTVCHCRK
N. glauca           ATCDLLSFSTPWFT---A---NDAACAGHCLVKGYKGGKC-RNGICHCRN
P. apterus          ATCDILSFQSQWVT---P---NHAGCALHCVIKGYKGGQC-KITVCHCRR
P. prasina          ATCDALSFSSKWLT---V---NHSACAIHCLTKGYKGGKC-VNTICNCRN
P. maculoventris    ATCDALSFSSKWLT---I---NHSACAIHCLTKGYKGGRC-MNTICNCRN
Ae. aegypti A       ATCDLLSG-FG------V---GDSACAAHCIARGNRGGYCNSKKVCVCRN
An. Gambiae         ATCDLASG-FG------V---GSSLCAAHCIARRYRGGYCNSKAVCVCRN
G. morsitans        VTCNI---------------GEWVCVAHCNSKSKKSGYC-SRGVCYCTN
S. calcitrans       ITCDLLSL-WK------V---GHAACAAHCLVLGDVGGYCTKEGLCVCKE
T. molitor          VTCDILSVEAK------GVKLNDAACAAHCLFRGRSGGYCNGKRVCVCR
Z. atrobas          FTCDVLGFEIA------GTKLNSAACGAHCLALGRRGGYCNSKSVCVCR
A. luxuriosa        FTCDVLSVEAK------GVKLNHAACGIHCLFRRRTGGYCNKKRVCICR
A. mellifera        VTCDLLSFKGQ------V---NDSACAANCLSLGKAGGHC-EKGVCICRKTSFKDLWDKRF
A. dichotoma        VTCDLLSFEAK---GFAA---NHSLCAAHCLAIGRRGGSC-ERGVCICRR
```

B:

Figure 4-2 *Rhodnius prolixus* defensins amino acid analysis.

A: Amino acid sequence alignments of three *R. prolixus* defensins. Alignments were compared with defensins from taxonomically related organisms. Each sequence has the classical loop, alpha helix and B sheet structure, and the highly conserved 6 cysteines that maintain the correct conformation.

B: Phylogenetic analysis of defensins from different sources. The defensins isolated from *R. prolixus* are most closely aligned with those from other Hemipterans, and then other insect orders indicating that modifications in the defensin sequence of *R. prolixus* may have occurred after the separation of insects in evolutionary time. Sequences were compared using Clustal W method with PAM 250 residue weight table. Units indicate the number of substitution events.

Defensin A: CTAGAATACTCCACTCAACCGCAACCAGACATC
Defensin B: AGCACAGTACCTAGGATATTCCACTCAACCGCAACCAGACATC
Defensin C: GTGCAGTACAGTCCTAATACCTAGCCACTCCTCACATC

Figure 4-3 Alignment of the cDNA sequences encoding three isoforms of *Rhodnius prolixus* defensin.

Dashes represent identical nucleotides in each sequence as compared with isoform A. Deduced amino acids are presented as single letter codes above each codon only when the same amino acid occurs in all three isoforms. Each sequence terminates with a stop codon (TGA) indicated by an asterisk. The putative polyadenylation consensus sequence is underlined. The position of the intron found in genomic clones is indicated by the arrow, and inserts G-intron-AT in amino acid D31 in the signal peptide of all three sequences.

118

Figure 4-4 Northern blot of *R. prolixus* Defensin.

Northern blot autoradiography demonstrating transcriptional activity for insect defensins in specific tissues collected from *Rhodnius prolixus* at various times after immune stimulation. Each lane contains 5μg of total RNA run on a formaldehyde-agarose gel and probed as described in the Methods. Lower panel: A probe generated from a 300 bp fragment of a *R. prolixus* actin clone was used as a control. Lanes: 1) Control RNA from fat body, 2) intestine—6h PI, 3) intestine—24h PI, 4) fat body—6h PI, 5) fat body—24h PI.
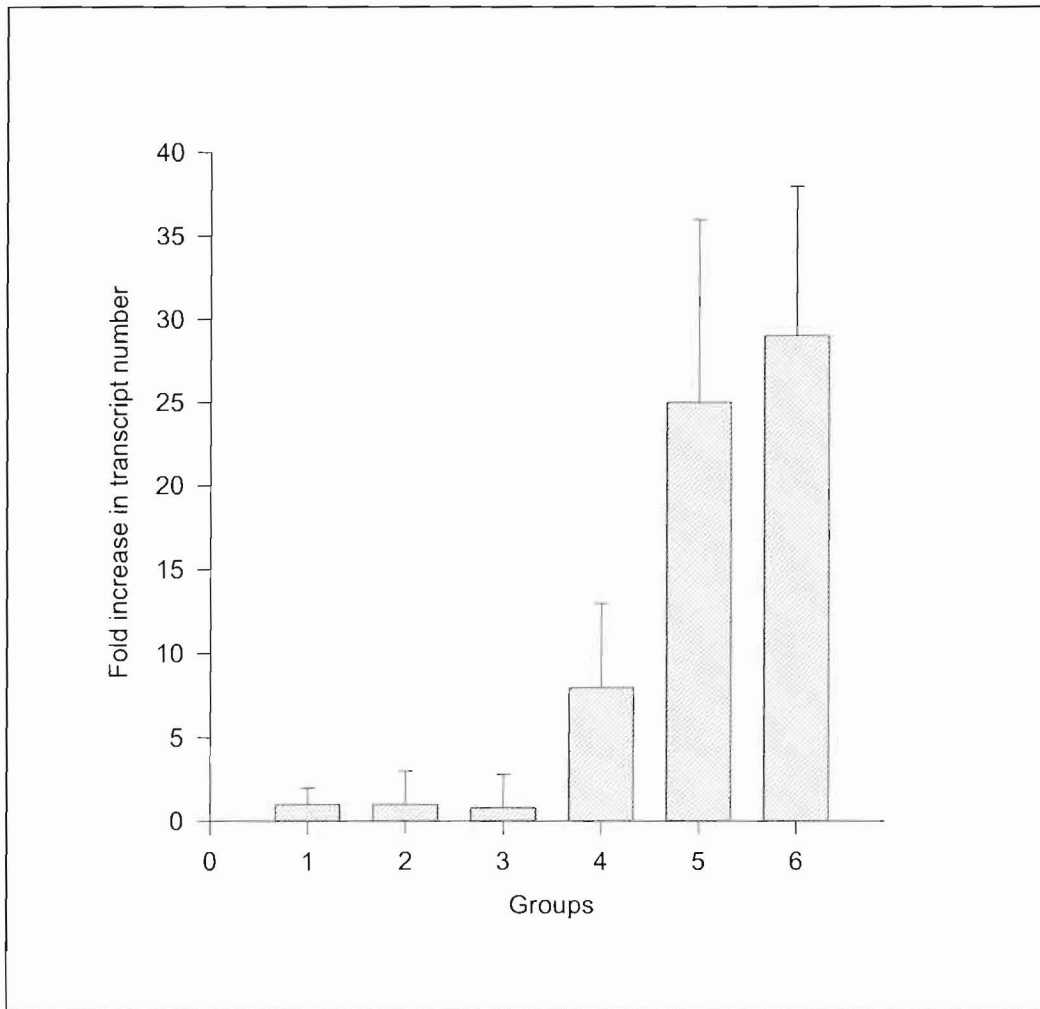
Figure 4-5 Real-Time Quantitative PCR of *Rhodnius prolixus* defensin A.

RNA was isolated from fat bodies and intestines at various times post immune stimulation. The raw data were compared with known amounts of pure template and the values calculated for graphical purposes. Values for control lanes (fat body or intestine) were arbitrarily given values of 1 and the vertical axis represents the fold increase in the tissues as compared with the same tissue controls. Lanes: 1) Control intestine, 2) Control Fat body, 3) Intestine 6 h PI, 4) Intestine 24 h PI, 5) Fat body 6 h PI, 6) Fat body 24 h PI.

References

Azambuja, P., Feder, D., Mello, C., Gomes, S. & Garcia, E., 1999. Immunity in *Rhodnius prolixus*: trypanosomatid-vector interactions. *Memorias do Instituto Oswaldo Cruz,* 94 Suppl 1, 219-22.

Barillas-Mury, C., Wizel, B. & Han, Y.S., 2000. Mosquito immune responses and malaria transmission: lessons from insect model systems and implications for vertebrate innate immunity and vaccine development. *Insect biochemistry and molecular biology,* 30, 429-42.

Boman, H.G., 1998. Gene-encoded peptide antibiotics and the concept of innate immunity: an update review. *Scandinavian journal of immunology,* 48, 15-25.

Brener, Z., 1973. Biology of *Trypanosoma cruzi*. *Annual review of microbiology,* 27, 347-82.

Broekaert, W.F., Terras, F.R., Cammue, B.P. & Osborn, R.W., 1995. Plant defensins: novel antimicrobial peptides as components of the host defense system. *Plant physiology,* 108, 1353-8.

Bulet, P., Dimarcq, J.L., Hetru, C., Lagueux, M., Charlet, M., Hegy, G., Van Dorsselaer, A. & Hoffmann, J.A., 1993. A novel inducible antibacterial peptide of *Drosophila* carries an O-glycosylated substitution. *The Journal of biological chemistry,* 268, 14893-7.

Dimarcq, J.L., Hoffmann, D., Meister, M., Bulet, P., Lanot, R., Reichhart, J.M. & Hoffmann, J.A., 1994. Characterization and transcriptional profiles of a *Drosophila* gene encoding an insect defensin. A study in insect immunity. *European journal of biochemistry / FEBS,* 221, 201-9.

Durvasula, R.V., Gumbs, A., Panackal, A., Kruglov, O., Aksoy, S., Merrifield, R.B., Richards, F.F. & Beard, C.B., 1997. Prevention of insect-borne disease: an approach using transgenic symbiotic bacteria. *Proceedings of the National Academy of Sciences of the United States of America,* 94, 3274-8.

Garcia, E., Gonzalez, M. & Azambuja, P., 1999. Biological factors involving *Trypanosoma cruzi* life cycle in the invertebrate vector, *Rhodnius prolixus*. *Memorias do Instituto Oswaldo Cruz,* 94 Suppl 1, 213-6.

Garcia, E.S. & Azambuja, P., 1991. Development and interactions of *Trypanosoma cruzi* within the insect vector. *Parasitology today,* 7, 240-4.

Gudmundsson, G.H., Lidholm, D.A., Asling, B., Gan, R. & Boman, H.G., 1991. The cecropin locus. Cloning and expression of a gene cluster encoding three antibacterial peptides in *Hyalophora cecropia*. *The Journal of biological chemistry,* 266, 11510-7.

Hetru, C. & Bulet, P., 1997. Strategies for the isolation and characterization of antimicrobial peptides of invertebrates. *In* W.M. Shafer (ed.) *Methods in molecular biology.* Totowa: Humana press, 35-49.

Hetru, C., Hoffmann, D. & Bulet, P., 1998. Antimicrobial peptides from insects. *In* P.T. Brey & D. Hultmark (eds.) *Molecular mechanisms of immune responses in insects.* London: Chapman and Hall, 40-66.

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

Hoffmann, J.A. & Reichhart, J.M., 2002. *Drosophila* innate immunity: an evolutionary perspective. *Nature immunology,* 3, 121-6.

Jaynes, J.M., Julian, G.R., Jeffers, G.W., White, K.L. & Enright, F.M., 1989. In vitro cytocidal effect of lytic peptides on several transformed mammalian cell lines. *Peptide research,* 2, 157-60.

Kylsten, P., Samakovlis, C. & Hultmark, D., 1990. The cecropin locus in *Drosophila*; a compact gene cluster involved in the response to infection. *The EMBO journal,* 9, 217-24.

Lehrer, R.I. & Ganz, T., 1996. Endogenous vertebrate antibiotics. Defensins, protegrins, and other cysteine-rich antimicrobial peptides. *Annals of the New York Academy of Sciences,* 797, 228-39.

Leippe, M., 1999. Antimicrobial and cytolytic polypeptides of amoeboid protozoa- -effector molecules of primitive phagocytes. *Developmental and comparative immunology,* 23, 267-79.

Lowenberger, C., 2001. Innate immune response of *Aedes aegypti. Insect biochemistry and molecular biology,* 31, 219-29.

Lowenberger, C., Bulet, P., Charlet, M., Hetru, C., Hodgeman, B., Christensen, B.M. & Hoffmann, J.A., 1995. Insect immunity: isolation of three novel inducible antibacterial defensins from the vector mosquito, *Aedes aegypti. Insect biochemistry and molecular biology,* 25, 867-73.

Lowenberger, C., Charlet, M., Vizioli, J., Kamal, S., Richman, A., Christensen, B.M. & Bulet, P., 1999a. Antimicrobial activity spectrum, cDNA cloning, and mRNA expression of a newly isolated member of the cecropin family from the mosquito vector *Aedes aegypti. The Journal of biological chemistry,* 274, 20092-7.

Lowenberger, C.A., Ferdig, M.T., Bulet, P., Khalili, S., Hoffmann, J.A. & Christensen, B.M., 1996. *Aedes aegypti*: induced antibacterial proteins reduce the establishment and development of *Brugia malayi. Experimental parasitology,* 83, 191-201.

Lowenberger, C.A., Kamal, S., Chiles, J., Paskewitz, S., Bulet, P., Hoffmann, J.A. & Christensen, B.M., 1999b. Mosquito-Plasmodium interactions in response to immune activation of the vector. *Experimental parasitology,* 91, 59-69.

Lowenberger, C.A., Smartt, C.T., Bulet, P., Ferdig, M.T., Severson, D.W., Hoffmann, J.A. & Christensen, B.M., 1999c. Insect immunity: molecular cloning, expression, and characterization of cDNAs and genomic DNA encoding three isoforms of insect defensin in *Aedes aegypti. Insect molecular biology,* 8, 107-18.

Possani, L.D., Zurita, M., Delepierre, M., Hernandez, F.H. & Rodriguez, M.H., 1998. From noxiustoxin to Shiva-3, a peptide toxic to the sporogonic development of *Plasmodium berghei. Toxicon,* 36, 1683-92.

Rodriguez, M.C., Zamudio, F., Torres, J.A., Gonzalez-Ceron, L., Possani, L.D. & Rodriguez, M.H., 1995. Effect of a cecropin-like synthetic peptide (Shiva-3) on the sporogonic development of Plasmodium berghei. *Experimental parasitology,* 80, 596-604.

Sambrook, J., Fritsch, E.F. & Maniatis, T., 1989. *Molecular cloning: a laboratory manual,* 2nd ed. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

Severson, D.W. & Kassner, V.A., 1995. Analysis of mosquito genome structure using graphical genotyping. *Insect Molecular Biology,* 4, 279-286.

Shahabuddin, M., Fields, I., Bulet, P., Hoffmann, J.A. & Miller, L.H., 1998. Plasmodium gallinaceum: differential killing of some mosquito stages of the parasite by insect defensin. *Experimental parasitology,* 89, 103-12.

Shai, Y., 1998. Mode of action of antibacterial peptides. *In* P.T. Brey & D. Hultmark (eds.) *Molecular mechanisms of immune responses in insects.* London: Chapman and Hall, 111-134.

Connecting statement 4

In chapter 4, we described the characterization of an inducible Defensin antimicrobial peptide. Defensins are mainly active against Gram positive bacteria, yet insects encounter a much wider variety of potentially pathogenic microorganisms. For example *Drosophila melanogaster* expresses 7 different AMPs, and it is logical that *R. prolixus* also should express other AMPs. In contrast to *D. melanogaster*, *R. prolixus* is a strictly hematophagous hemipteran and requires the presence of bacterial symbionts in its intestinal tract to reach adulthood. Indiscriminate growth of these bacterial populations must be regulated and AMPs expressed in the gut likely play a role in this process. In the next chapter I describe the identification and characterization of two differentially expressed lysozyme genes. Real time PCR studies showed the compartmentalized expression of these two genes suggesting that they may have different roles in immunity and digestion. The presence of lysozymes in the intestinal tract has potential implications for the obligate bacterial symbionts and for the survival of protozoan human pathogen *T. cruzi*.

CHAPTER 5: IDENTIFICATION AND CHARACTERIZATION OF TWO NOVEL LYSOZYMES FROM *RHODNIUS PROLIXUS*, A VECTOR OF CHAGAS DISEASE

URSIC BEDOYA R.[a], NAZZARI H.[b], COOPER D.[a], TRIANA O.[c], WOLFF M.[c], LOWENBERGER C.[a]

[a]Department of Biological Sciences, Simon Fraser University
Burnaby, BC, V5A 1S6 Canada.
[b]Department of Cellular and Physiological Sciences, University of British Columbia, Vancouver V6T 1Z3, Canada.
[c]Instituto de Biología, Universidad de Antioquia, Calle 67 No 53-108, Medellín, Colombia

Abstract

Lysozymes have been described in invertebrates as both digestive and immune molecules. We report here the characterization of two novel c-type lysozymes, RpLys-A (EU250274) and RpLys-B (EU250275), isolated from the fat body and digestive tract of immune-stimulated *Rhodnius prolixus,* a major vector of Chagas disease. Transcriptional profiles indicate that the temporal and spatial expression patterns of these two peptides are very different. RpLys-A was expressed predominantly in the midgut after ingestion of *Trypanosoma cruzi* in a blood meal, or after injection of bacteria into the hemocoel. RpLys-B was expressed primarily in the fat body after bacterial injection. Phylogenetic alignments indicated that RpLys-A aligned best with molecules from other hemipterans whose major expression is found in the intestinal tract whereas RpLys-B aligned best with mosquito and tick molecules whose expression is found principally in hemocytes and fat body and whose role has been described as immune-related. These data suggest a differential tissue distribution and role of these two closely-related molecules; one for immunity in the hemocoel and the other for digestion in the midgut.

Introduction

Chagas disease, or American trypanosomiasis, is caused by the parasitic protozoan *Trypanosoma cruzi* that is transmitted to humans principally by hematophagous triatomine insects such as *Rhodnius prolixus*. Chagas disease remains prevalent in many areas of the Americas, ranging from southern Argentina to the southern United States, and afflicts over 17 million people in these locations (Dutra *et al.*, 2005). A number of studies have suggested that inducible immune peptides can limit parasite development in vectors (Jaynes *et al.*, 1988, Rodriguez *et al.*, 1995, Lowenberger *et al.*, 1996, Boisbouvier *et al.*, 1998, Possani *et al.*, 1998, Shahabuddin *et al.*, 1998, Lowenberger *et al.*, 1999, Lowenberger, 2001, Vizioli *et al.*, 2001). The majority of these studies examined associations in which the parasites make direct contact with hemolymph factors as they move from their site of development to the salivary glands for the subsequent transmission to vertebrates (Lowenberger *et al.*, 1999). *Trypanosoma cruzi*, however, never leaves the intestinal tract and is voided in the faces during blood ingestion, and therefore has no direct contact with hemolymph factors (Azambuja and Garcia, 1987, Lopez *et al.*, 2003). This inefficient transmission of *T. cruzi* may be an evolutionary adaptation by the parasite to avoid contact with lethal components of the innate immune response of the vector (Lopez *et al.*, 2003). This concept is supported by studies in which *T. cruzi* was injected and subsequently killed and cleared in the hemocoel of *R. prolixus* (Azambuja and Garcia, 1987) and by paratransgenic studies, which demonstrated the susceptibility of *T. cruzi*, in vivo, to Cecropin A, another AMP (Durvasula *et al.*, 1997).

Because previous studies had indicated the presence of a ~15kDa protein in the hemolymph of *R. prolixus* following immune activation that had antibacterial activity similar to other insect lysozymes (Lopez, Wolff, Triana, and Lowenberger unpublished), and because lysozyme-like activity had been reported in this

species (Ribeiro and Pereira, 1984, Azambuja and Garcia, 1987) we sought to identify cDNAs encoding lysozymes from this vector species. Lysozymes hydrolyze the 1,4-Beta-linkage between N-acetylmuramic acid and N-acetylglucosamine of the cell wall peptidoglycans of bacteria (Grunclova *et al.*, 2003). As such, lysozymes may function in a digestive role for insects that ingest large numbers of bacteria (Regel *et al.*, 1998), as immune related molecules to prevent colonization of the hemocoel by pathogens, and in some insects different isoforms of lysozymes may serve both functions (Ursic Bedoya *et al.*, 2005).

In *Drosophila melanogaster*, lysozymes are found in the gastrointestinal tract and are involved in digestion (Kylsten *et al.*, 1992). In many Lepidoptera and nematoceran Diptera, lysozymes are found in the hemolymph but not in the gut (Lemos and Terra, 1991). The recruitment of lysozymes as digestive enzymes, and their adaptation to an acidic midgut, may have occurred after the divergence of Cyclorrhapha from the Nematocera (Hultmark, 1996).

Lysozymes may be expressed constitutively to regulate gut flora and help initiate the rapid immune response of the insect. Previous studies have shown that peptidoglycan fragments, produced by the enzymatic action of lysozymes on bacterial cell walls, are very potent inducers of the fat body response (Dunn *et al.*, 1985). However, the kissing bugs that transmit *T. cruzi* (such as *R. prolixus*) are a much more ancient insect family and the role lysozymes may play in digestion and/or immunity is unknown. We report here the isolation and characterization of two chicken type (c-type) lysozymes isolated from *R. prolixus* and their temporal and spatial expression in response to blood feeding, immune activation with bacteria, and the ingestion of a blood meal containing the human pathogen, *T. cruzi*.

Material and Methods

Insect maintenance, immune activation, and exposure to *Trypanosoma cruzi.*

A colony of *R. prolixus* has been maintained at the Institute of Biology, Universidad de Antioquia, Medellín Colombia for over 10 years and at Simon Fraser University, British Columbia, Canada for 5 years. The bacteria used for immune activation were grown and maintained as described previously (Lowenberger *et al.*, 1996). Briefly, *Escherichia coli* and *Micrococcus luteus* were grown in Luria-Bertani's rich nutrient medium (LB medium) overnight at 37°C while shaking at 350 rpm. Cultures were combined, pelleted by centrifugation, and insect minuten pins (0.10mm diameter) were dipped into the pellet and inserted directly into the hemocoel of adult insects as described (Lopez *et al.*, 2003). Sterilized pins were inserted into the hemocoel of control insects. For detecting peptide expression in response to ingested parasites, insects were allowed to feed on mice infected, or not, with the HA strain of *T. cruzi* at the Institute of Biology at the University of Antioquia in Medellin, Colombia.

Tissue Collection.

We isolated the intestinal tracts or fat body tissue for RNA extraction from bacteria-inoculated, blood fed, or *T. cruzi*-exposed *R. prolixus* adults at various times after treatment. Tissues from bacteria-inoculated insects were collected 8 and 24 hours post inoculation. Tissues were collected from blood-fed or *T. cruzi*-exposed insects at 0, 2, 7 and 14 days post feeding. These times correspond to different developmental stages and location of the parasite in the insect intestinal tract.

RNA isolation and cDNA synthesis.

Total RNA was extracted from the selected tissues of immune-activated and naive insects at various time points after inoculation or blood feeding using TRI REAGENT (Molecular Research Centre, USA) following manufacturer's

instructions. RNA was quantified using a Biophotometer (Eppendorf, Germany) and 2.5 µg of total RNA was used for reverse transcription as described previously (Lowenberger *et al.*, 1999) using an oligo dT primer (MG) with a 5' extension (5'CGGGCAGTGAGCAACGT$_{12}$3'). Degenerate forward primers (5'GAYAAYGGNYTNTTYCARAT3' and 5'GGNGGNCCNAAYAARAAYGGN3') were designed against a partial protein sequence obtained previously (Lopez, Wolff, Triana, and Lowenberger unpublished) and conserved regions of other insect lysozymes. These primers were used with the MG primer in a PCR reaction with the conditions: 95°C (3 min), and 30 cycles of 95°C (10 s), 53°C (10 s), 72°C (2 min) on an Idaho Technologies Rapid Cycler (Salt Lake City, USA). The products of these reactions were size-fractioned on a 1.2% low melting point agarose gel. Bands of predicted size were excised from the gel, heated to 65°C, and cloned directly into pGEM-T vector (Promega, USA) using the manufacturer's protocols. Transformations using XL1-Blue cells, and blue-white screening of presumed transformants were done following manufacturer's protocols. Selected colonies were grown overnight in 5mL LB medium containing ampicillin (100 µg/mL) and purified using the Wizard Plus Minipreps DNA Purification system (Promega, USA). Sequencing of these clones was carried out on an ABI-310 automated sequencer using Big Dye v3.1 chemistry. Sequences were compared with available sequences in the NCBI database. Two unique sequences that aligned well with insect lysozymes were obtained and specific primers for each cDNA sequence were designed. The 5'end of each sequence was obtained using specific reverse primers in a RACE reaction (Marathon cDNA synthesis kit; Clontech, USA). PCR was done under the conditions: 95°C (3min), and 30 cycles of 95°C (1 min), 60 °C (30 s), 72°C (1 min) and the resulting products separated, ligated, screened, and sequenced as described above. SeqMan (DNA STAR, USA) was used to align overlapping sequences of our two clones. Specific primers then were designed to amplify each of the two-cDNA *R. prolixus* sequences: A and B, with no cross amplification.

Genomic DNA extraction.

Genomic DNA was isolated from five starved adult insects. Insects were ground in a glass tissue grinder with 1.5mL of fresh DNA extraction buffer (EB) (0.5% SDS, 0.2 M NaCl, 25 mM EDTA, 10 mM Tris pH 8) and 1.5mL of phenol. After incubation at room temperature for 15 min, the homogenate was transferred to a 15mL Corex tube and centrifuged at 8800g for 20 min at 4°C on an Allegra 64R (Beckman Coulter, USA) centrifuge. The aqueous phase was transferred to a new tube to which an equal volume of phenol:chloroform (1:1) was added. The mixture was homogenized by vortexing and centrifuged at 6650g for 20 min at 4°C. The supernatant was transferred to a new tube, mixed with an equal volume of chloroform, and centrifuged at 6650g for 15 min at 4°C. The resulting supernatant was transferred to a new tube and 1:10 volume of 4 M ammonium acetate plus 2.5 volumes of 95% ethanol were added, and the mixture was stored at -20°C for 1h. The DNA was pelleted by centrifugation at 6650g for 30 min at 4°C. The resulting pellet was dissolved in 100µL of EB buffer (10 mM Tris-HCl pH 8.5), treated with RNAse A (50 mg at 37°C for 30 min), and the DNA was further extracted with 100uL of phenol:chloroform:iso-amylalcohol (25:25:1), washed with 95% ethanol, dried, resuspended in 100uL of EB buffer and quantified using a Biophotometer (Eppendorf, Germany).

Target gene identification.

Inverse PCR (iPCR) (Triglia, 2000) was used to amplify regions of genomic DNA upstream of the coding region of our genes to identify potential transcription factor binding sites. We used NEBcutter v2.0 (New England Biolabs, USA) restriction digest analysis to identify restriction enzymes that would digest the cDNA Lys 1A gene within the first 500bp of the initial methionine of the coding region. The enzymes used were: Dpn1, Rsa1 and EcoRV (New England Biolabs, USA). One microgram of genomic DNA was digested separately with 10U of each restriction enzyme in an air incubator at 37°C for 3h. Restriction enzymes were heat inactivated according to manufacturer's instructions or the digested DNA was isolated by a phenol:chloroform extraction. Approximately

200ng of each digested genomic DNA were self-ligated with 12U of T4 Ligase (Promega, USA) at 16°C for 16h in 100µL reactions. Two microliters of the ligation reaction were used in a PCR reaction using iProof DNA polymerase (Bio-Rad, USA). The inverse oriented primers used were: F: 5'CCAACTACGACGGAAGCTATGATAATGGA3' and R: 5'CTAGTGAACACCCTAGCTTGTGTGGC3'. Amplicons obtained were cloned into pGem- T-easy and transformed into *E. coli* JM109 and sequenced as described above. In addition to the molecular approach, we also used a bioinformatic approach to confirm our findings. We searched for contigs from the recently released trace data from the *R. prolixus* genome sequencing project using Mega Blast searches (http://www.ncbi.nlm.nih.gov/blast/mmtrace.shtml) using the first 200 nucleotides of the open reading frame from our cDNA sequences. Contigs containing the identified genomic clones and the remaining regions of the open reading frames were constructed using the SeqManII module of DNAstar software with loose assembling parameters to accommodate large gaps corresponding to introns. Putative transcription binding sites were identified using Alibaba 2.1 software (Grabe, 2002) using lazy restriction parameters. Alibaba predicts transcription factor binding sites by context dependent matrices generated from TRANSFAC 4.0 public sites.

Sequence identity analysis.

Multiple sequence alignments of *R. prolixus* lysozymes and other invertebrate lysozymes were carried out using MegAlign (DNA Star, USA) using the Clustal W method with PAM 250 matrix. Prediction of the signal peptide was performed using SignalP v.3.0. Theoretical isoelectric points (pI) and molecular weights were determined using Expasy ProtParam program.

Transcriptional profile using Quantitative RT-PCR.

We have generated a battery of cDNAs from various tissues of *R. prolixus* adults after different immune stimulations. We used these cDNAs in a Real-Time Quantitative-PCR (Q-PCR) analysis using a Rotor-Gene 3000 (Corbett

Research, Australia) to compare expression patterns of transcripts for both identified cDNAs. We constructed standard curves with known concentrations of purified lysozyme cDNAs. A 150bp fragment of *R. prolixus* β-actin gene was used to normalize cDNA samples. Standard 25μl PCR reactions containing 1.0μl of SYBRGreen (Molecular Probes, Sigma, St. Louis, MO) and primers that distinguished between the two lysozyme sequences (RpLys-A-forward: 5'ATGAAAGCTGTTTTCTTACTGGC3' and reverse: 5'AAAGCAAACGTTGATATCTGGTA3' and RpLys-B-forward: 5'ATGATTGCAAATCTAGTTTTAACACTATTGC3' and reverse: 5'TTAACAAACCAATGGAGGCAAC3') were used in a PCR program of 95 °C (3 min) and 35 cycles of 95° C (10 s), 58°C for (10 s), and 72°C (30 s). Quantification, melt curve analysis and sample comparison were done with the Rotor-Gene version 5 software (Corbett Research). Three independently synthesized cDNAs for each time point were evaluated in these studies and each cDNA was analyzed at least 5 times to detect levels of the lysozyme and β-actin sequences.

Results

Sequence Analysis.

The *R. prolixus* lysozyme cDNAs contain deduced open reading frames of 417 and 414 nucleotides for RpLys-A and RpLys-B respectively, that encode proteins of 139 and 138 residues with predicted sizes of 15.8 and 15.1 kDa respectively The coding region contains a stop codon and a 3' untranslated region (UTR) of 96 and 100 nucleotides for RpLys-A and RpLys-B respectively, a putative polyadenylation consensus signal (AATAAA) and 15 and 19 additional nucleotides before the poly-A tail for RpLys-A and -B respectively (Figure 5-1). Each deduced protein sequence contains a putative signal peptide comprising the first 18 residues that terminates with an alanine residue (Ala18) based on SignalP v3.0 Expasy tools (Bendtsen *et al.*, 2004). The calculated theoretical pI

values are 8.5 and 6.84 for the active regions of RpLys-A and RpLys-B respectively (Expasy ProtParam).

Identification of upstream promoter sites.

iPCR successfully amplified only one amplicon that contained a segment of the RpLys-A gene (data not shown). This fragment contained only 60bp of upstream sequence, but nonetheless contained a putative NF-κB site. Initially we could not find the sequence of RpLys-A in the trace data files but as the *R. prolixus* genome is completed and annotated we were able to identify a genomic clone containing 170bp of upstream sequence. Data mining in this manner produced a strong match for RpLys-B and we obtained a 415bp sequence upstream of the start ATG codon. Analysis with Alibaba 2.1 detected several regions identified as potential transcription factor binding sites. Of particular interest were GATA-1 and NF-κB sites (Table 5-1) present in the upstream region of both genes. We cannot eliminate the possibility of other binding factor sites further upstream.

Multiple alignments and phylogenetic analysis.

Multiple protein sequence alignment with lysozymes from selected organisms indicates RpLys-A and RpLys-B share significant identity with other insect c-type lysozymes. All lysozymes documented here have the conserved Glu50 Asp68 in the active site and 8 structural Cysteine residues. There is a conserved active functional domain region (FQIND) found in the vast majority of insect c-type lysozymes (Figure 5-2). It is apparent from the alignment that while there is significant conservation of regions and motifs within the active proteins, there is apparently no such conservation in the signal peptide region. Similarly there does not appear to be a conserved pattern within the signal peptides of proteins identified in closely related species or in a specific tissue (e.g. fat body). Therefore, subsequent analysis only included the sequences of the active proteins from which the signal peptides had been removed. RpLys-A shares only 49% identity with RpLys-B. RpLys-A shared the greatest identity (79, 61, and

78%) with lysozymes from closely related organisms; *T. infestans* lysozyme -1 (Kollien *et al.*, 2003), T. infestans lysozyme-2 (Balczun *et al.*, 2008), and *T. brasiliensis* (Araujo *et al.*, 2006) respectively. RpLys-B shares 53, 46, and 52% identity respectively with these same molecules. Therefore the two lysozymes identified from *R. prolixus* are very different from each other. Comparison of the RpLys-A and –B sequences with other insect lysozymes using CLUSTAL W (v.3.2.2) indicated a shared identity with *Ae. aegypti*-A (41%), *Ae. aegypti*-B (41%), *Ae. aegypti*-S (34%), *Ae. albopictus* (38%), *An. gambiae* (40%), *D. andersoni* (41%), *D. variabilis* (41%) *D. melanogaster* B (45%), *D. melanogaster* D (45%), *D. melanogaster* P (42%), *H. cecropia* (48%), *H. virescens* (45%), *M. domestica* (44%), *S. cynthia* (51%), *T. ni* (48%) (Figure 5-2). A comparison of selected invertebrate lysozymes was performed at the amino acid level (Figure 5-3) using only the active regions of the proteins. The cladogram (Figure 5-3) shows a general separation of lysozymes based on function: lysozymes described as having a principal role in immune function are separate from those whose function has been described mainly as digestive. The sequence of the termite, *R. speratus*, is significantly different from all of the other sequences used here, and is appropriately on its own branch. Lysozymes from the Lepidoptera, mosquitoes, and ticks (*Dermacentor* sp.), whose molecules have been described as having more of an immune function, group together. The lysozymes from the flies, *Triatoma* sp. and a tick (*O. moubata*), whose function has been described as digestive, group in another major branch. Our molecule, RpLys-A, whose expression is greatest in the intestine groups with the branch of triatome lysozymes found in the clade that contains digestive lysozymes whereas RpLys-B, found in the fat body, is found in the clade of molecules whose function has been described as immune related, including molecules from distantly related Lepidoptera, Diptera, and ticks.

Induction of lysozyme genes.

Real time Quantitative PCR (Q-PCR) was used to compare expression patterns of RpLys-A and RpLys-B in different tissues, and at various time points,

after inoculation with bacteria or ingestion of a blood meal containing *T. cruzi*. Constitutively expressed transcripts were found in fat body tissues and the intestinal tract for both lysozyme sequences. After introduction of bacteria into the hemocoel we found a differential induction of transcription of both molecules. RpLys-B increased 12 and 18 fold in fat body tissues at 8h and 24h post inoculation respectively, but showed no significant difference in midgut expression at these time points (Figure 5-4). In contrast, RpLys-A was up-regulated 24 fold transiently in the midgut of *R. prolixus* 8h after the inoculation of bacteria into the hemocoel. Transcript levels decreased to baseline amounts in midgut tissues 24 h post inoculation (Figure 5-4A). There was no induction of RpLys-A in the fat body 8 and 24 h post inoculation (Figure 5-4B).

In blood-fed insects we measured minimal differences in expression of RpLys A between 0 and 48h post ingestion of a sterile blood meal or a meal containing the parasite *T. cruzi*. Subsequently, RpLys A transcripts increased >20 fold in midgut and intestinal tissues extracted 7 and 14 days post ingestion of the parasite laden blood meal, but no differences were determined after the ingestion of a parasite free blood meal. The presence or absence of *T. cruzi* in the blood meal did not produce significant changes in RpLys-B expression (Figure 5-5).

Discussion

In a previous study (Lopez *et al.*, 2003) we identified, using HPLC, an inducible protein approximately 15kDa with lysozyme-like activity in the hemolymph of immune activated *R. prolixus*. Lysozyme-like activity had been reported previously in the hemolymph of this vector, (Azambuja and Garcia, 1987) and in intestinal homogenates obtained from adult insects, with 2 peaks 3 days and 3 weeks after feeding (Ribeiro and Pereira, 1984). A similar activity was observed in response to *M. lysodeikticus* (Azambuja and Garcia, 1987) or *T. cruzi* (Mello *et al.*, 1995) injection into the hemolymph but no characterization or sequencing of these proteins was done (Azambuja and Garcia, 1987). Our data corroborate the conclusions of these authors that the effects they observed were

likely due to lysozyme expression. We report here the expression pattern of two novel lysozymes in this vector of Chagas disease.

Lysozymes are ubiquitous proteins described in many groups of invertebrates and vertebrates and which have been characterized as immune related molecules (Roxstrom-Lindquist *et al.*, 2004), digestive enzymes (Grunclova *et al.*, 2003), or multifunctional molecules (Li *et al.*, 2005, Ursic Bedoya *et al.*, 2005). At the protein level, RpLys-A and RpLys-B share the greatest identity with lysozymes isolated from the closely related insects, *T. infestans* (Kollien *et al.*, 2003) and *T. brasiliensis* (Araujo *et al.*, 2006) (Figure 5-3). A sequence analysis of the active regions of these lysozymes indicates a general grouping determined more by function (immune or digestion) or possibly location (hemocytes/fat body or digestive tract). RpLys-B, found in the fat body, aligns closest to a group containing immune related lysozymes isolated from the distantly related Lepidoptera or the ticks, *D. andersoni* and *D. variabilis* (Simser *et al.*, 2004), rather than organisms that are more closely related taxonomically. RpLys-A, found mainly in the digestive tract of *R. prolixus*, aligns best with the lysozyme found in the digestive tract of the other hemipterans, *T. infestans* and *T. brasiliensis*, and falls within the general grouping of digestive lysozymes. By including sequences from termites and the distantly related ticks in the analysis it appears that the conservation of specific aspects of these very similar molecules is based on their functional role or tissue origin (fat body/hemocytes) rather than solely by taxonomic relatedness. All invertebrate lysozymes compared in this study share a common theme; 8 cysteine residues that form 4 disulphide bridges, and, with the exception of T. infestans-Lys2 (Balczun *et al.*, 2008), all share the conserved catalytic sites of glutamic and aspartic acid residues. This conservation of common structural components in all insect and tick lysozymes suggests a major role for these molecules in invertebrates.

In *Drosophila melanogaster*, it has been widely documented that promoter sequences in antimicrobial peptide (AMP) genes contain combinations of

transcription factor binding sites responsible for their tissue and signal-dependent specificities (Uvell and Engstrom, 2007). Two different classes of transcription factors are implicated in the transcriptional activation of AMPs in the *Drosophila* fat body; the NF-κB factors Dorsal, Dif, and Relish (RHD containing proteins), and the GATA factor Serpent (Senger *et al.*, 2004, Senger *et al.*, 2006). Generally, GATA sites are located within 20bp of the NF-κB sites in functionally important promoter regions (Kadalayil *et al.*, 1997). Furthermore, regulatory regions for Diptericin and Metchnikowin require GATA sites for their activation in the midgut and a second GATA factor, dGATAe, mediates a Toll-independent immune response in the midgut (Senger *et al.*, 2006).

Although three other lysozyme genes have been characterized at the molecular level from kissing bugs (Kollien *et al.*, 2003, Araujo *et al.*, 2006, Balczun *et al.*, 2008), no information concerning their regulatory sequences has been published. Our data concurs with information available from higher order dipteran insects. Both of the genes we describe in this paper contained both GATA and NF-κB sites in the promoter's proximal region. No GATA transcription factor homologue has been identified in triatomes to date and our early attempts using homology searches to *Drosophila*'s Serpent did not yield any results. However, we have identified and described in chapter 3 a Dorsal homologue (Ursic-Bedoya and Lowenberger, 2007) and current work is underway to characterize this gene. Further work also will focus on mapping regulatory sequences from other immunity genes in *R. prolixus*, which we have identified through a series of SSH libraries (Ursic-Bedoya and Lowenberger, 2007), to conduct a comparative analysis among different insect orders.

We detected very low baseline transcripts for both sequences in naive insects. In response to bacteria inoculation in the hemocoel we detected significant increases in transcription of RpLys-B in the fat body 8h and 24h after stimulation. This inoculation also produced a spike of expression of RpLys-A in the midgut/intestine, which indicates that these tissues are immune responsive, and

that they may be activated by stimuli received elsewhere in the body. These data suggest a systemic and coordinated immune response, possibly mediated by cytokines, in which the stimulation of one region results in an increased expression of peptides in another region as has been reported previously (Lowenberger *et al.*, 1996, Dimopoulos *et al.*, 1997, Ursic Bedoya *et al.*, 2005).

In blood-fed insects, with or without parasites in the blood meal, we found no significant changes in the transcriptional profile for RpLys-B, neither in the midgut nor the fat body, suggesting that the interactions between parasite and digestive tract do not result in a systemic response in the fat body. In addition, because this parasite normally does not cross the digestive tract, there is no direct contact between parasite and hemolymph factors such as hemocytes and fat body cells.

However, we do see a significant increase in RpLys-A expression in the digestive tract in response to the presence of the *T. cruzi*, 7 and 14 days post ingestion. This up-regulation did not occur in insects that ingested a parasite-free blood meal, suggesting that a specific interaction occurred between *T. cruzi* and midgut and intestinal tissues over the first 14 days of parasite development.

Although we have demonstrated an inducible response in *R. prolixus* to *T. cruzi* in the digestive tract, these enzymes expressed at normal physiological levels are not lethal to the parasite. *Rhodnius prolixus* relies on obligate mutualistic bacterial symbionts in the intestine to provide essential nutrients (Azambuja and Garcia, 1987, Garcia *et al.*, 2007). While hemocoel AMPs kill pathogens, AMPs in the intestine may regulate the proliferation of microbes in the digestive tract but must not eliminate these essential symbionts by expressing high concentrations of AMPs to which the symbionts are susceptible. *Trypanosoma cruzi* is susceptible to insect AMPs, as demonstrated in vitro and in vivo (Durvasula *et al.*, 1997, Beard *et al.*, 2001, Lopez *et al.*, 2003) suggested that its exclusive relegation to the intestine may have evolved to permit development and multiplication in an area of low AMP expression. Molecules such as lysozymes may have evolved to play dual roles in digestion and defense.

Recently, a number of studies that have examined the expression of lysozymes in *T. brasiliensis* (Araujo *et al.*, 2006) and *T. infestans* (Kollien *et al.*, 2003, Balczun *et al.*, 2008) showed significant up-regulation in the digestive tract. The molecules from *T. infestans* also showed an up-regulation in response to molting and feeding. These data are supported by previous studies that demonstrated lysozyme-like activity in the intestine (Ribeiro and Pereira, 1984) or after pathogens were injected into the hemocoel (Azambuja and Garcia, 1987, Mello *et al.*, 1995).

Characterization of the seven c-type lysozymes found in *Anopheles gambiae* suggests that lysozymes may have been adapted more recently to play a role in immunity (Li *et al.*, 2005). We have identified two lysozymes that we propose may act in a compartmentalized manner to protect the hemocoel or to aid in digestion and maintain intestinal flora at acceptable levels. Our data suggest that lysozymes evolved for important roles in both digestion and immunity in the more 'ancient' hemimetabolous triatominae but also play a role in the 'higher' Diptera and Lepidoptera. To speculate on the evolution of lysozymes and to develop a prediction of which physiological function, digestion or immunity, preceded the other would require a detailed examination of these molecules throughout the invertebrates.

The parasite *T. cruzi* lives in the milieu of the digestive tract and appears not to be affected by normal physiological levels of lysozymes, or other molecules, expressed in these tissues. Further studies will reveal if hemocoel components, such as RpLys-B, can reduce *T. cruzi* viability. Whether the molecules described here originated for a digestive or defense role cannot be determined. Our results lend support for the dual role of lysozymes in invertebrates. The identification of novel immune peptides in vectors such as *R. prolixus* will increase our knowledge of general insect immunity, and the evolutionary origins of these and other immune peptides, and may provide insight into vector-parasite interactions that affect and regulate parasite transmission.

Acknowledgements

Figures

```
A:
Aggtcattcgcaaa
M    K    A    V    F    L    L    A    I    F    A    L    L    G    A    T    Q    A    R    V
ATG  AAA  GCT  GTT  TTC  TTA  CTG  GCT  ATT  TTC  GCC  CTN  CTT  GGC  GCC  ACA  CAA  GCT  AGG  GTG
F    T    R    C    G    L    A    R    E    L    A    R    Q    G    L    P    R    H    D    L
TTC  ACT  AGA  TGC  GGT  CTA  GCG  CGG  GAA  TTG  GCT  AGG  CAA  GGA  CTT  CCA  CGC  CAC  GAT  TTG
A    N    W    V    C    L    I    E    A    E    S    G    R    N    T    R    A    R    G    G
GCT  AAT  TGG  GTA  TGC  CTG  ATT  GAA  GCA  GAA  AGT  GGC  AGA  AAT  ACC  AGA  GCC  AGA  GGT  GGC
P    N    Y    D    G    S    Y    D    N    G    L    F    Q    I    N    D    R    I    W    C
CCC  AAC  TAC  GAC  GGA  AGC  TAT  GAT  AAT  GGA  CTA  TTC  CAG  ATC  AAT  GAT  AGA  ATT  TGG  TGT
M    N    G    R    P    G    H    A    C    H    V    R    C    E    D    L    R    T    D    D
ATG  AAC  GGT  AGA  CCT  GGA  CAT  GCT  TGC  CAC  GTC  AGA  TGT  GAA  GAT  TTA  AGA  ACA  GAC  GAC
I    R    A    S    V    R    C    A    V    Q    I    K    Q    Q    Q    G    W    S    A    W
ATC  AGA  GCC  AGT  GTG  AGG  TGT  GCT  GTC  CAA  ATC  AAA  CAG  CAA  CAG  GGC  TGG  TCA  GCT  TGG
Y    G    W    Q    Y    H    C    R    G    R    P    L    P    D    I    N    V    C    F    *
TAC  GGT  TGG  CAG  TAC  CAT  TGC  AGA  GGT  CGT  CCC  CTA  CCA  GAT  ATC  AAC  GTT  TGC  TTT  TAA

AGCAATCTAATTATTTTAAACTATTAACAAATTTTGTTAATTGTTTATTAGTATGCTTTGATTGTAATTGAGACTATTTGtaaa
cattacccatcgaataaaatattattgaatttcAn


B:
ctgatcatatacggac
M    I    A    N    L    V    L    T    L    L    L    L    F    T    V    S    S    A    K    V
ATG  ATT  GCA  AAT  CTA  GTT  TTA  ACA  CTA  TTG  CTG  CTG  TTT  ACT  GTC  AGT  TCA  GCC  AAA  GTG
F    T    D    C    E    L    A    N    V    L    E    N    A    G    F    P    K    D    Q    L
TTC  ACC  GAT  TGT  GAA  CTG  GCA  AAT  GTA  TTG  GAA  AAT  GCT  GGA  TTC  CCA  AAA  GAT  CAA  CTA
K    D    W    I    C    L    A    K    A    E    S    S    L    N    T    T    A    V    G    G
AAA  GAC  TGG  ATT  TGT  CTA  GCT  AAA  GCA  GAA  AGC  TCA  CTG  AAC  ACC  ACG  GCC  GTC  GGA  GGA
P    N    K    N    G    S    Y    D    Y    G    L    F    Q    I    N    D    H    I    W    C
CCA  AAT  AAA  AAT  GGA  AGC  TAT  GAT  TAT  GGT  TTA  TTT  CAG  ATA  AAC  GAT  CAT  ATA  TGG  TGT
D    P    E    K    R    G    G    D    C    N    V    K    C    S    D    L    V    L    E    D
GAT  CCA  GAA  AAA  AGA  GGA  GGT  GAT  TGT  AAT  GTG  AAA  TGT  TCA  GAT  CTT  GTT  CTT  GAA  GAT
D    I    G    P    S    M    N    C    A    K    I    V    Y    K    V    Q    G    F    K    A
GAC  ATT  GGA  CCC  AGT  ATG  AAT  TGT  GCA  AAA  ATA  GTT  TAT  AAA  GTT  CAA  GGA  TTC  AAG  GCA
W    N    G    L    D    K    K    C    K    G    K    K    L    P    P    L    V    C    *
TGG  AAT  GGT  TGG  ATC  AAG  AAA  TGT  AAG  GGC  AAA  AAG  TTG  CCT  CCA  TTG  GTT  TGT  TAA

AAAGAAGAAGAATAAGAAGAAGGAAAGCTAGTTCCTAAAATCATTGTTTATaaatggtcaaattgtaaaagaatttgttatttg
caataattttagaataaattaaactgaatttattaata
```

Figure 5-1 cDNA and translated amino acid sequences of *Rhodnius prolixus* lysozymes.

The termination codons are marked with an asterisk. The bold amino acid sequence and arrowhead indicate the predicted signal sequence and the N-terminus of the mature protein, respectively. The putative polyadenylation signal is double underlined. The nucleotide sequences used as primer sites for Real Time Quantitative PCR are underlined. The putative active site residues are in bold and italic (Glu50 and Asp68).

142
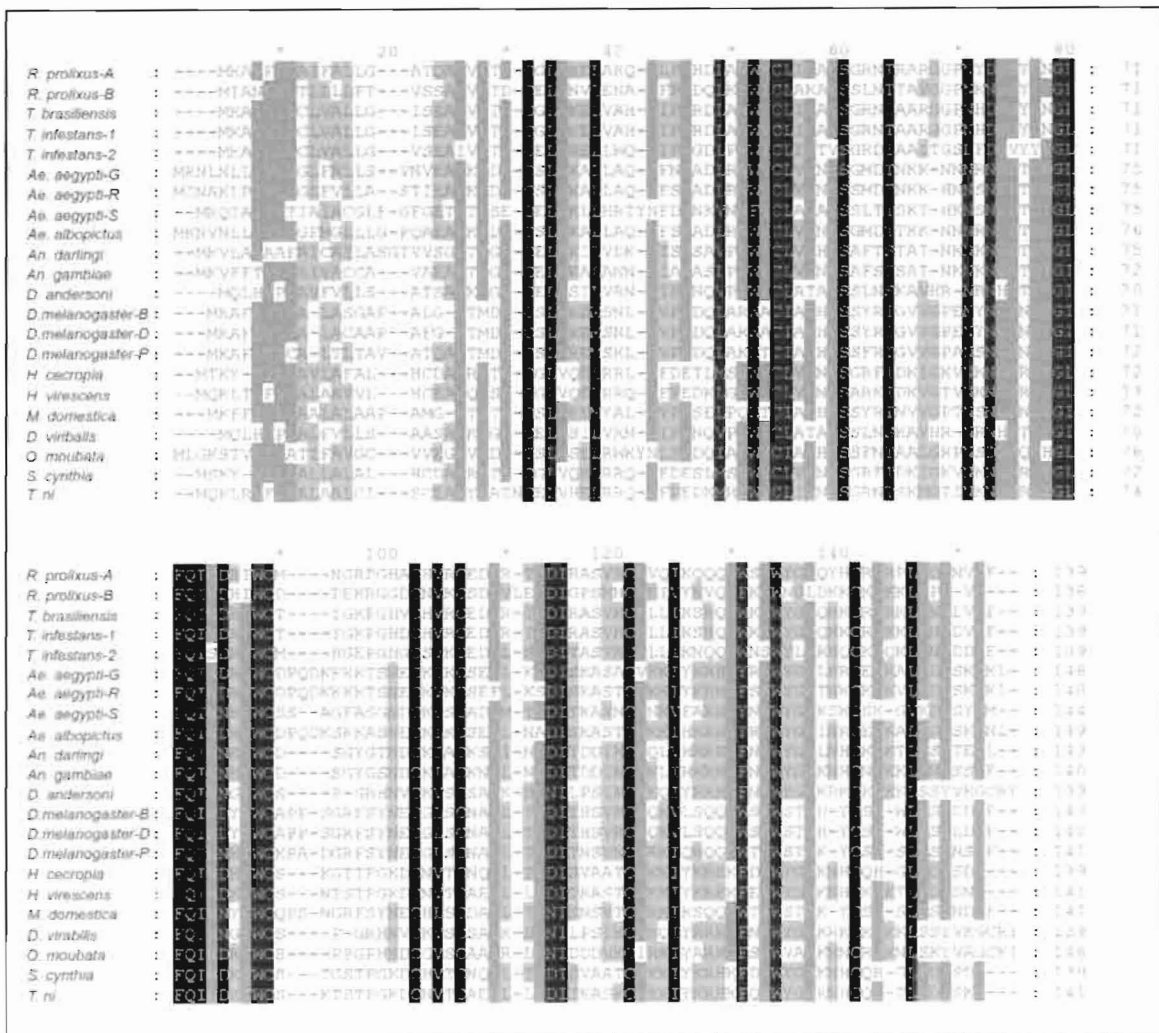
Figure 5-2 Multiple sequence alignments of the amino acid sequences of *R. prolixus* lysozymes and other invertebrate lysozyme precursors.

The shading reflects the level of conservation throughout these molecules. The sequences used in this analysis were: *R. prolixus*-A (EU250274) and -B (EU250275); *A. albopictus* (AY089957); A. *aegypti*- A (AJ290428) –B (AY693973) and - S (AF466591); *A. darlingi* (AF003945); *A. gambiae* (Q17005); *D. andersoni* (AY207371); *D. melanogaster*- B (Z22225), - D: (X58382) and -P: (X58383); *D. variabilis* (AY183671); *H. cecropia* (P05105); *H. virscens* (U50551); *M. domestica* (AY344588); *O. moubata* (AF425264); *S. Cynthia* (AB048258); *T. infestans* Lys 1 (AY253830) and Lys 2 (Balczun, et al., 2008) ; *T. brasiliensis* (AAU04569) and *T. ni* (P50718)

143
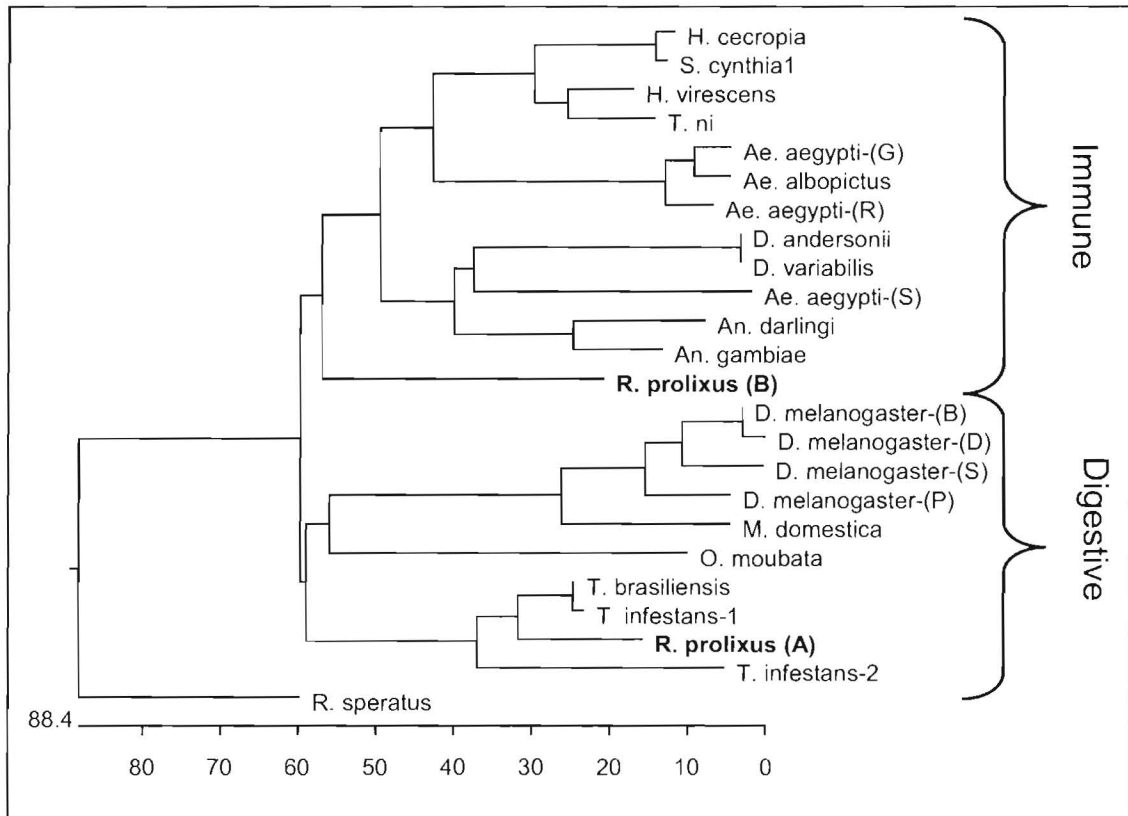
Figure 5-3 Cladogram of selected insect lysozyme sequences.

Lysozymes represented include molecules from diverse insects and for which a known function has been described. Alignments were constructed with the Clustal W method with the PAM250 residue weight table using the active regions of the lysozymes from which the signal peptide regions had been removed.
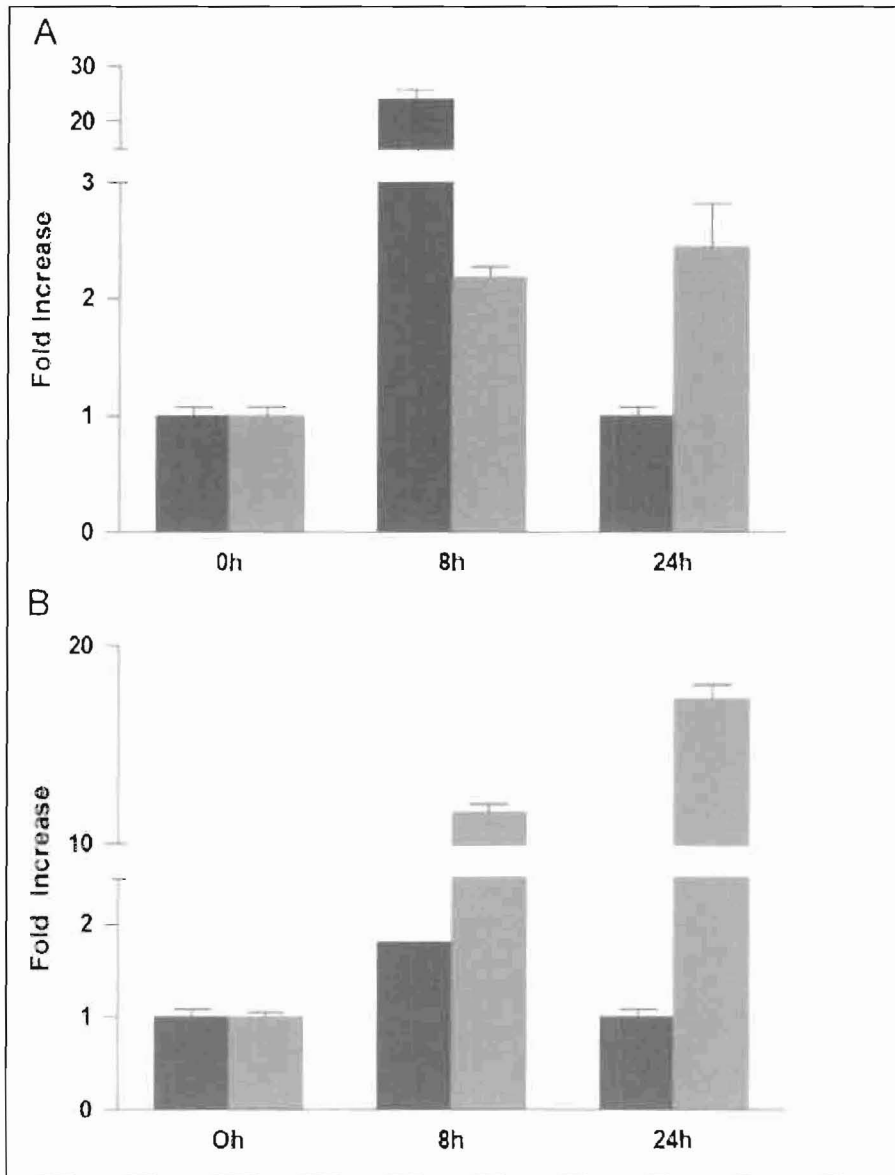
Figure 5-4 Real time quantitative PCR (Q-PCR) profile of R. *prolixus* lysozymes.

Lysozymes-A and –B in midgut (A) and fat body (B) tissues after bacterial inoculation. Standard curves were established for RpLys-A and RpLys-B and a 150bp fragment of *R. prolixus* β-actin. Products from each sample were normalized to β-actin levels in each cDNA. The levels in control insects that were injected with a sterile needle were arbitrarily designated as 1, and all other levels were expressed as a fold increase over controls. Each bar represents the mean (+ SD) fold increase of 5 replicates from three independently derived cDNAs. Black bars= RpLys-A, Grey bars= RpLys-B.
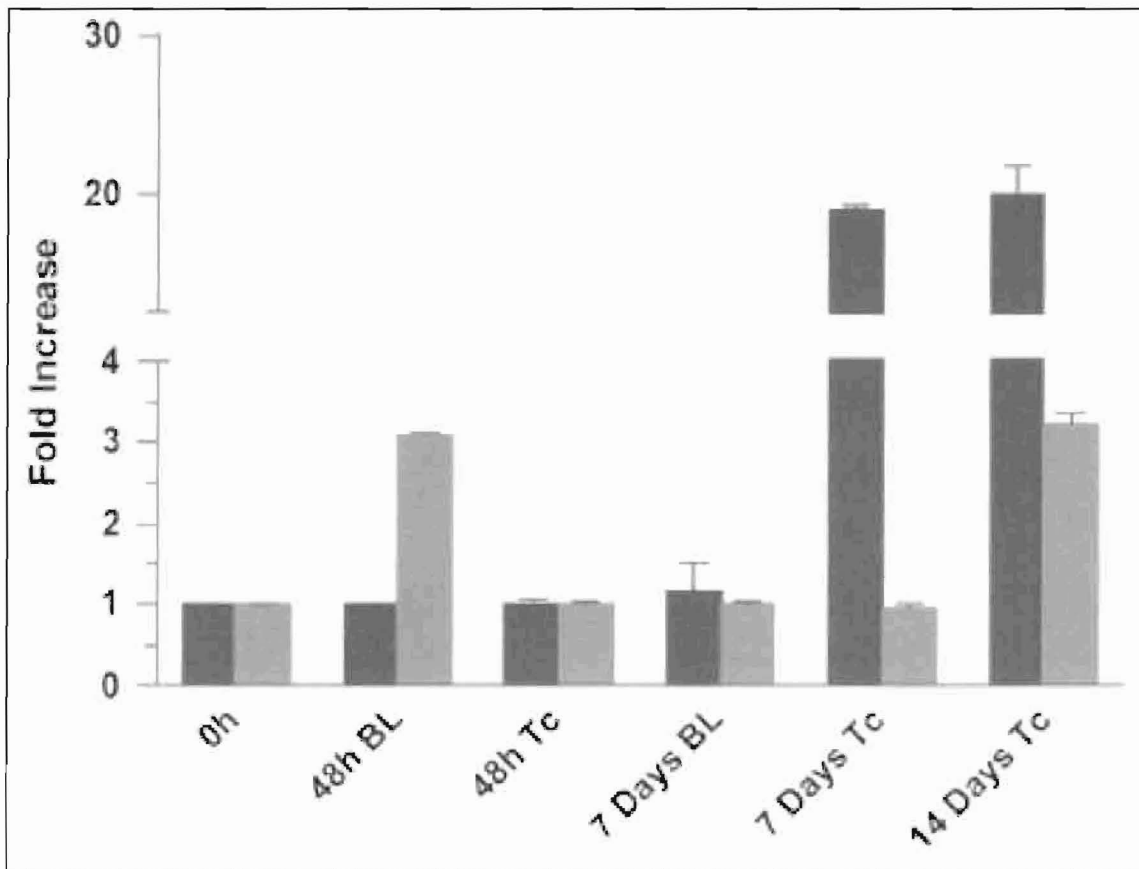
Figure 5-5 Real time quantitative PCR profile of *R. prolixus* lysozymes-A and –B in the midgut.

Standard curves were established for RpLys-A and RpLys-B and a 150bp fragment of *R. prolixus* β-actin. Products from each sample were normalized to β-actin levels in each cDNA. The levels in naive insects were arbitrarily designated as 1, and all other levels were expressed as a fold increase over controls. Each bar represents the mean (+ SD) fold increase of 5 replicates from three independently derived cDNAs. Black bars= RpLys-A, Grey bars= RpLys-B, BL = sterile blood meal, Tc = blood meal containing *T. cruzi*.

Tables

Table 5-1 *R. prolixus* lysozymes putative transcription factor binding sites.

Putative transcription binding sites were identified using Alibaba 2.1 software using lazy restriction parameters. Alibaba predicts transcription factor binding sites by context dependent matrices generated from TRANSFAC 4.0 public sites. Location of the putative binding site is indicated relative to the methionine start codon. The clone indicators refer to the trace data files available at the NCBI trace data archives.

Gene	Clone	NF-κB	GATA
Lys 1A	iPCR Dpn1 NADD-aee07e10	GGAACTTTCAA -64 ATTAGGAAATAC -49	TGTTTCAGATC -115 CTTATATTTCT -42
Lys 1B	NAAX-ady62g11	TAGGAAATGAC -181	TTTGAGCAGAA -356 TTATTATTTTT -302

References

Araujo, C.A., Waniek, P.J., Stock, P., Mayer, C., Jansen, A.M. & Schaub, G.A., 2006. Sequence characterization and expression patterns of defensin and lysozyme encoding genes from the gut of the reduviid bug *Triatoma brasiliensis. Insect biochemistry and molecular biology,* 36, 547-60.

Azambuja, P. & Garcia, E.S., 1987. Characterization of inducible lysozyme activity in the hemolymph of *Rhodnius prolixus. Brazilian journal of medical and biological research,* 20, 539-48.

Balczun, C., Knorr, E., Topal, H., Meiser, C.K., Kollien, A.H. & Schaub, G.A., 2008. Sequence characterization of an unusual lysozyme gene expressed in the intestinal tract of the reduviid bug *Triatoma infestans* (Insecta). *Parasitology research,* 102, 229-32.

Beard, C.B., Dotson, E.M., Pennington, P.M., Eichler, S., Cordon-Rosales, C. & Durvasula, R.V., 2001. Bacterial symbiosis and paratransgenic control of vector-borne Chagas disease. *International journal for parasitology,* 31, 621-7.

Bendtsen, J.D., Nielsen, H., Von Heijne, G. & Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. *Journal of molecular biology,* 340, 783-95.

Boisbouvier, J., Prochnicka-Chalufour, A., Nieto, A.R., Torres, J.A., Nanard, N., Rodriguez, M.H., Possani, L.D. & Delepierre, M., 1998. Structural information on a cecropin-like synthetic peptide, Shiva-3 toxic to the sporogonic development of *Plasmodium berghei. European journal of biochemistry / FEBS,* 257, 263-73.

Dimopoulos, G., Richman, A., Muller, H.M. & Kafatos, F.C., 1997. Molecular immune responses of the mosquito *Anopheles gambiae* to bacteria and malaria parasites. *Proceedings of the National Academy of Sciences of the United States of America,* 94, 11508-13.

Dunn, P.E., Dai, W., Kanost, M.R. & Geng, C.X., 1985. Soluble peptidoglycan fragments stimulate antibacterial protein synthesis by fat body from larvae of *Manduca sexta. Developmental and comparative immunology,* 9, 559-68.

Durvasula, R.V., Gumbs, A., Panackal, A., Kruglov, O., Aksoy, S., Merrifield, R.B., Richards, F.F. & Beard, C.B., 1997. Prevention of insect-borne disease: an approach using transgenic symbiotic bacteria. *Proceedings of the National Academy of Sciences of the United States of America,* 94, 3274-8.

Dutra, W.O., Rocha, M.O. & Teixeira, M.M., 2005. The clinical immunology of human Chagas disease. *Trends in parasitology,* 21, 581-7.

Garcia, E.S., Ratcliffe, N.A., Whitten, M.M., Gonzalez, M.S. & Azambuja, P., 2007. Exploring the role of insect host factors in the dynamics of *Trypanosoma cruzi-Rhodnius prolixus* interactions. *Journal of insect physiology,* 53, 11-21.

Grabe, N., 2002. AliBaba2: context specific identification of transcription factor binding sites. *In silico biology,* 2, S1-15.

Grunclova, L., Fouquier, H., Hypsa, V. & Kopacek, P., 2003. Lysozyme from the gut of the soft tick *Ornithodoros moubata*: the sequence, phylogeny and post-feeding regulation. *Developmental and comparative immunology,* 27, 651-60.

Hultmark, D., 1996. Insect lysozymes. *In* P. Jollès (ed.) *Lysozymes--model enzymes in biochemistry and biology.* Basel: Birkhäuser Verlag, 87-102.

Jaynes, J.M., Burton, C.A., Barr, S.B., Jeffers, G.W., Julian, G.R., White, K.L., Enright, F.M., Klei, T.R. & Laine, R.A., 1988. In vitro cytocidal effect of novel lytic peptides on *Plasmodium falciparum* and *Trypanosoma cruzi. The FASEB journal,* 2, 2878-83.

Kadalayil, L., Petersen, U.M. & Engstrom, Y., 1997. Adjacent GATA and kappa B-like motifs regulate the expression of a *Drosophila* immune gene. *Nucleic acids research,* 25, 1233-9.

Kollien, A.H., Fechner, S., Waniek, P.J. & Schaub, G.A., 2003. Isolation and characterization of a cDNA encoding for a lysozyme from the gut of the reduviid bug *Triatoma infestans. Archives of insect biochemistry and physiology,* 53, 134-45.

Kylsten, P., Kimbrell, D.A., Daffre, S., Samakovlis, C. & Hultmark, D., 1992. The lysozyme locus in *Drosophila melanogaster*: different genes are expressed in midgut and salivary glands. *Molecular & general genetics,* 232, 335-43.

Lemos, F.J. & Terra, W.R., 1991. Digestion of bacteria and the role of midgut lysozyme in some insect larvae. *Comparative biochemistry and physiology,* 100, 265-8.

Li, B., Calvo, E., Marinotti, O., James, A.A. & Paskewitz, S.M., 2005. Characterization of the c-type lysozyme gene family in *Anopheles gambiae. Gene,* 360, 131-9.

Lopez, L., Morales, G., Ursic, R., Wolff, M. & Lowenberger, C., 2003. Isolation and characterization of a novel insect defensin from *Rhodnius prolixus,* a vector of Chagas disease. *Insect biochemistry and molecular biology,* 33, 439-47.

Lowenberger, C.A., 2001. Form, function and phylogenetic relationships of mosquito immune peptides. *Advances in experimental medicine and biology,* 484, 113-29.

Lowenberger, C.A., Ferdig, M.T., Bulet, P., Khalili, S., Hoffmann, J.A. & Christensen, B.M., 1996. *Aedes aegypti:* induced antibacterial proteins reduce the establishment and development of *Brugia malayi. Experimental parasitology,* 83, 191-201.

Lowenberger, C.A., Kamal, S., Chiles, J., Paskewitz, S., Bulet, P., Hoffmann, J.A. & Christensen, B.M., 1999. Mosquito-Plasmodium interactions in response to immune activation of the vector. *Experimental parasitology,* 91, 59-69.

Mello, C.B., Garcia, E.S., Ratcliffe, N.A. & Azambuja, P., 1995. *Trypanosoma cruzi* and *Trypanosoma rangeli:* interplay with hemolymph components of Rhodnius prolixus. *Journal of invertebrate pathology,* 65, 261-8.

Possani, L.D., Zurita, M., Delepierre, M., Hernandez, F.H. & Rodriguez, M.H., 1998. From noxiustoxin to Shiva-3, a peptide toxic to the sporogonic development of *Plasmodium berghei. Toxicon,* 36, 1683-92.

Regel, R., Matioli, S.R. & Terra, W.R., 1998. Molecular adaptation of Drosophila melanogaster lysozymes to a digestive function. *Insect biochemistry and molecular biology,* 28, 309-19.

Ribeiro, J.M. & Pereira, M.E., 1984. Midgut glycosidases of *Rhodnius prolixus. Insect Biochemistry,* 14, 103-108.

Rodriguez, M.C., Zamudio, F., Torres, J.A., Gonzalez-Ceron, L., Possani, L.D. & Rodriguez, M.H., 1995. Effect of a cecropin-like synthetic peptide (Shiva-3) on the sporogonic development of *Plasmodium berghei. Experimental parasitology,* 80, 596-604.

Roxstrom-Lindquist, K., Terenius, O. & Faye, I., 2004. Parasite-specific immune response in adult *Drosophila melanogaster:* a genomic study. *EMBO reports,* 5, 207-12.

Senger, K., Armstrong, G.W., Rowell, W.J., Kwan, J.M., Markstein, M. & Levine, M., 2004. Immunity regulatory DNAs share common organizational features in *Drosophila. Molecular cell,* 13, 19-32.

Senger, K., Harris, K. & Levine, M., 2006. GATA factors participate in tissue-specific immune responses in *Drosophila* larvae. *Proceedings of the National Academy of Sciences of the United States of America,* 103, 15957-62.

Shahabuddin, M., Fields, I., Bulet, P., Hoffmann, J.A. & Miller, L.H., 1998. *Plasmodium gallinaceum:* differential killing of some mosquito stages of the parasite by insect defensin. *Experimental parasitology,* 89, 103-12.

Simser, J.A., Macaluso, K.R., Mulenga, A. & Azad, A.F., 2004. Immune-responsive lysozymes from hemocytes of the American dog tick, *Dermacentor variabilis* and an embryonic cell line of the Rocky Mountain wood tick, *D. andersoni. Insect biochemistry and molecular biology,* 34, 1235-46.

Terra, W.R., 1990. Evolution of Digestive Systems of Insects. *Annual Review of Entomology,* 35, 181-200.

Triglia, T., 2000. Inverse PCR (iPCR) for obtaining promoter sequence. *In* M.J. Tymms (ed.) *Transcription factor protocols.* Totowa, NJ, 79 - 83.

Ursic-Bedoya, R.J. & Lowenberger, C.A., 2007. *Rhodnius prolixus*: identification of immune-related genes up-regulated in response to pathogens and parasites using suppressive subtractive hybridization. *Developmental and comparative immunology,* 31, 109-20.

Ursic Bedoya, R.J., Mitzey, A.M., Obraztsova, M. & Lowenberger, C., 2005. Molecular cloning and transcriptional activation of lysozyme-encoding cDNAs in the mosquito *Aedes aegypti. Insect molecular biology,* 14, 89-94.

Uvell, H. & Engstrom, Y., 2007. A multilayered defense against infection: combinatorial control of insect immune genes. *Trends in genetics,* 23, 342-9.

Vizioli, J., Bulet, P., Hoffmann, J.A., Kafatos, F.C., Muller, H.M. & Dimopoulos, G., 2001. Gambicin: a novel immune responsive antimicrobial peptide from the malaria vector *Anopheles gambiae. Proceedings of the National Academy of Sciences of the United States of America,* 98, 12630-5.

Connecting statement 5

In the previous chapter we described two lysozyme genes and their expression profiles with an emphasis on the presence of the human pathogen *T. cruzi*. Lysozymes, like defensins, represent a subset of AMPs that have complex secondary structures and are highly conserved through evolutionary time. Not all AMPs exhibit these features. In Chapter six, we describe the identification and characterization of a third class of AMPs, specific to *Rhodnius prolixus* which we named Prolixin. Prolixin is a short, linear and glycine rich peptide expressed by both fat body and midgut tissues. The antibacterial activity of this peptide seems to target selected bacterial species. Our phylogenetic analysis suggests that this peptide is an ancestor, or a precursor molecule, of glycine-rich AMPs such as the Diptericins and the Attacins that to date have been described in the Diptera.

CHAPTER 6: PROLIXIN: A NOVEL MEMBER OF THE DIPTERICIN ANTIMICROBIAL PEPTIDE FAMILY ISOLATED FROM THE HEMIPTERAN *RHODNIUS PROLIXUS*.

URSIC BEDOYA Raul, LOWENBERGER Carl
Department of Biological Sciences
Simon Fraser University
Burnaby BC, V5A1S6 Canada.

Abstract

We have characterized Prolixin, a novel antimicrobial peptide isolated from midgut tissues of the hemipteran, *Rhodnius prolixus.* Prolixin belongs to the Diptericin family of antimicrobial peptides and it is the first member to be isolated outside the Diptera. Prolixin is an 11kDa peptide containing a 21 amino acid signal peptide, 2 putative phosphorylation sites, lacks glycosylation sites and is produced by both fat body and midgut tissues of nymphs or adults in response to bacterial infection of the hemolymph or the midgut. The *Prolixin* gene does not seem to be under the control of NF-κB binding sites unlike most antibacterial peptides, but its promoter region contains several GATA sites. The recombinant protein has strong bactericidal activity against the Gram positive bacterium *Bacillus subtilis* and also bacteriostatic activity against the Gram negative bacterium *Escherichia coli* in liquid culture.

Introduction

Arthropod-borne infectious and parasitic diseases cause severe human mortality and morbidity throughout the world particularly in developing countries where health conditions are poor. Malaria, yellow fever, dengue, African and American trypanosomiasis, encephalitis, leishmaniasis, filariasis, and onchocerciasis are all transmitted to humans by insects. According to the World Health Organization, all combined, these diseases cause the death of over 2 million individuals per year, with malaria being the deadliest of them all (WHO, 2002b).

Parasites have exploited the blood-feeding behavior of insects for transmission to suitable hosts. Insects are one of the most successful groups of organisms and have colonized almost every ecological niche on this planet. This success is in part due to their powerful immune system that eliminates most pathogenic organisms.

Insects defend themselves from pathogens and foreign organisms using a variety of cellular and humoral mechanisms, but lack the antibody-mediated responses found in higher organisms. *Drosophila melanogaster* is currently the model organism for studies in insect immunity and much of the work done in other systems is based on, or compared to *Drosophila* data. Recently a significant amount of work has also been done with insects of economic and medical relevance, particularly *Plasmodium*-transmitting mosquitoes (Chen *et al.*, 2008).

We have been studying the immune response of *Rhodnius prolixus*, a hematophagous hemipteran insect. *R. prolixus* is a major vector of *Trypanosoma cruzi*, the etiologic agent of American trypanosomiasis (Chagas disease) in humans. An estimated 18 million people are infected with Chagas disease, 100 million are at risk of acquiring it and approximately 14,000 per year die directly

because of this disease (WHO, 2002a). This disease is highly correlated with poverty and poor housing conditions and therefore affects mostly impoverished rural populations in the Americas. *Trypanosoma cruzi* infects a broad range of mammals and is transmitted to humans through blood transfusion, congenital transmission, but mostly (80%) by blood-sucking Reduvidae insects. Whereas most insect transmitted pathogens exit the vector in the saliva as the vector bites the vertebrate host, this is not the case for *T. cruzi*. Instead, the parasite remains in the intestine and rectum of the insect (Brener, 1973). As the insect engorges, it defecates and fecal droplets containing the parasites are deposited on the host's skin and may enter via the bite site or a mucosal membrane. Although inefficient, this mode of transmission allows the parasite to avoid components of the vectors' immune response present in the hemolymph.

An insect's innate immune response consists of a stepwise progression of three functional processes: (i) non-self recognition; (ii) signal transduction and (iii) the production of effector molecules responsible for eliminating the invading organism. The activation of the immune system relies on the very basic ability of recognizing and discriminating self from non-self.

Once activated the immune response regulates responses through signaling pathways that trigger two types of parallel responses, cellular and humoral. The cellular response is assumed by hemocytes, which engage in different defense mechanisms depending on their lineage. In *Drosophila*, plasmatocytes are specialized for phagocytosis; crystal cells are responsible for initiating the phenoloxidase (PO) cascade and lamellocytes are adhesive hemocytes (found only in larvae) specialized in the encapsulation of large foreign microorganisms (Strand, 2008). In *R. prolixus*, five different types of hemocytes have been described (prohemocytes, plasmatocytes, granulocytes, oenocytoids and adipohemocytes), yet plasmatocytes are also the only cell type involved in phagocytosis (Borges *et al.*, 2007). In addition to their role in cellular immunity, hemocytes are also known to participate in the humoral response as they can

produce and secrete antimicrobial peptides (AMPs) into the hemolymph (Lavine and Strand, 2002).

Humoral defenses are characterized by a battery of potent AMPs, reactive intermediates of nitrogen or oxygen and a complex enzymatic cascade involved in clotting or hemolymph melanization (Lemaitre and Hoffmann, 2007). In general, AMPs are small proteins and peptides, usually less than 100 amino acid residues long, with sizes ranging from 2 to 25kDa, which play a pivotal role in clearing microbial infections. Over 500 different AMPs have been described and in the hemolymph their concentrations range from 1 to 100µM. AMPs are produced primarily by the fat body but also by the midgut and hemocytes.

In *Drosophila,* eight classes of AMPs have been classified into three families based on the organisms they selectively eliminate: Gram negative bacteria, Gram positive bacteria, or fungi (Bulet and Stocklin, 2005). One family of AMPs are the Diptericins, originally identified in *Phormia terranovae* (Dimarcq *et al.*, 1988), and which to date, have only been found in dipteran insects. Diptericins are less than 100 amino acids long, can be glycosylated, are expressed primarily in the digestive tract and are active against Gram negative bacteria. Insects secrete AMPs into the gut lumen to protect themselves from microorganisms acquired through feeding but also to digest and regulate their resident bacterial flora. This is true for *R. prolixus* which harbors *Rhodococcus rhodnii*, an actinomycete acquired by coprophagy and required by the insect to reach sexual maturity (Wigglesworth, 1936).

Recently, through a functional genomics approach using Suppressive Subtractive Hybridization (SSH) we identified several up-regulated genes in response to bacterial or *T. cruzi* invasion of the hemocoel of *R. prolixus* (Ursic-Bedoya and Lowenberger, 2007). We describe here the molecular characterization of a novel, glycine-rich AMP whose initial EST sequence was obtained in this previous study

from a midgut library, belonging to the Attacin/Diptericin superfamily of antimicrobial peptides.

Material and Methods

Gene sequencing and characterization.

Initial DNA sequence data for *Rhodnius prolixus* prolixin was obtained from our previous study deposited at the NCBI corresponding to dbEST accession numbers EB084328 and EB084378 (Ursic-Bedoya and Lowenberger, 2007). To obtain the full length sequence of the gene, we extracted RNA from fat body tissues 24h after bacteria inoculation as described (Lopez *et al.*, 2003). Full length cDNAs were obtained using 5'-3' Rapid amplification of cDNA ends (RACE) with the Marathon cDNA synthesis kit (Clontech, USA) using primers AP2: 5'TGGGAATTCCACCAGACCTCCCAC3' and gene specific Rstop: 5'TCCCCAGGTGTGTTCTATCC3'. After cloning the resulting PCR amplicons into pGEM-Teasy (Promega, USA) and transforming into *E. coli* JM109, DNA sequencing was performed with BigDye v3.1 (Applied Biosystems, USA) chemistry as previously described (Ursic-Bedoya and Lowenberger, 2007). The resulting DNA sequences were assembled into a single contig using Lasergene's module SeqMan (DNAstar, USA). We subjected the full length cDNA sequence to bioinformatics analyses to identify known structural features and putative post translational modification sites. Peptide conserved domains were identified using NCBI's conserved domain database (Marchler-Bauer *et al.*, 2005). Signal peptide prediction (Emanuelsson *et al.*, 2007) and putative post translational modifications were performed using tools from the Center for Biological Sequence Analysis (http://www.cbs.dtu.dk/services/) (Hansen *et al.*, 1998, Blom *et al.*, 2004).

Protein alignment and phylogenetic analysis.

Multiple sequence analysis of all diptericins was performed using Lasergene's MegAlign module (DNAstar, USA) as described (Lopez *et al.*, 2003) using a protein data set containing: *Protophormia terraenovae* Diptericin

(CAB57822.1); *Mayetiola destructor* Diptericin (ABG21230.1); *Drosophila melanogaster* Diptericin A (NP_476808.1), B (NP_523787.2) *Glossina morsitans morsitans* Diptericin A (AAL34111.1) *Stomoxys calcitrans* Diptericin A (AAY98016.1), *Rhodnius prolixus* Prolixin (bankit1060780 EU448993). Phylogenetic analysis was performed using the active protein sequence of the selected proteins and the following additional sequences: *Bombyx mori* Attacin (AAB34519.1) *Drosophila melanogaster* Attacin (CAA86995.1); *Glossina morsitans morsitans* Attacin A (AAL34113.1); *Manduca sexta* Attacin 2 (AAO74640.1); *Holotrichia diomphalia* Holotricin 2 (BAA02890.1); *Allomyrina dichotoma* Coleoptericin A (BAB40436.1), B(BAB40437.1), C (BAB40438.1) and *Oryctes rhinoceros* Rhinocerosin (O76145.1) and *Acalolepta luxuriosa* Acaloleptin (P81592.1). The phylogenetic tree was built using Megalign's Jotun Hein method (Hein, 1990).

Genomic sequence and promoter analysis.

Rhodnius prolixus genomic DNA was obtained as described (Ursic-Bedoya *et al.* 2008), using a phenol-chloroform extraction protocol. Prolixin's genomic DNA sequence was obtained by PCR amplification using primers F1: 5'ACAATTTTGGTGGTGGTTGTC3' and Rstop: 5'TCCCCAGGTGTGTTCTATCC3' and 1ng of genomic DNA as template under the following conditions: 95°C, 2min; 30 cycles of 95°C, 15s; 62°C 10s; 72°C, 45s and a final extension of 2min at 72°C with Platinum Taq DNA polymerase (Invitrogen, city USA). The resulting DNA amplicon was cloned, sequenced and analyzed as described above. To obtain DNA sequence containing upstream regulatory sequences we searched *R. prolixus* WGS database deposited on NCBI's trace archives (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi). The resulting hits were assembled into a genomic contig to manually annotate the intron/exon boundary and to identify the 5' upstream sequence of the gene. Upstream DNA sequence was analyzed using Alibaba2 v2.1 software (Grabe, 2002) accessed through Biobase biological databases (http://www.gene-regulation.com) using default parameters to identify putative transcription factor binding sites.

Gene expression profiles.

Immune challenge of the insects was done as described (Ursic-Bedoya and Lowenberger, 2007). Entire eggs or insect tissues were carefully dissected and RNA was extracted immediately with Trizol (Invitrogen, USA) as described (Ursic-Bedoya and Lowenberger, 2007). cDNAs used for tissue expression profiling were generated starting with 1µg of total RNA with SuperScript II enzyme (Invitrogen, USA) and MGdT primer (5'CGGGCAGTGAGCGCAACGT$_{14}$-3') following manufacturer's recommendations whereas cDNAs used for real time PCR were synthesized with 5µg of DNAse treated total RNA using MMLV-RT (Promega, USA).

For a qualitative analysis of prolixin's transcription we used RT-PCR to identify tissues which express this AMP. We extracted RNA from eggs, salivary glands, midgut and fat body tissues and generated cDNAs as described above. We screened these cDNAs with prolixin-specific primers F1 (5'ACAATTTTGGTGGTGGTTGTC3') and qR (5'GCTTGAGCTCTGGTCCTTCC3') to determine tissue specific expression.

For a more precise estimation of comparative transcription rates, Real-Time Quantitative PCR (qPCR) was used in tissues that tested positive for prolixin expression with primers F1-qR for prolixin and

Act-nqF: (5'AGTTATGCCTTGCCACACG3'),

Act-Rstop: (5'TTAGAAGCATTTGCGGTGG3') for beta actin. PCR conditions were optimized using Platinum SYBR green qPCR SuperMix-UDG (Invitrogen, USA) so that amplification efficiencies of both amplicons were equivalent. PCR conditions used were: 50°C 2min; 95°C 2 min; 40 cycles of 95°C 10s; 61°C 15s; 72°C 20s; 83°C 20s. Fluorescence acquisition was done at 83°C to eliminate nonspecific signal generated by primer dimers. QPCR data were analyzed using the -2ΔΔCt method (Livak and Schmittgen, 2001).

Recombinant protein expression and isolation.

We expressed Prolixin's open reading frame using Novagen's pET system. Initially, we PCR amplified prolixin using primers pET-F (5'GACGACGACAAGATGTTCAAAACAATTTTGG3') and pET-R (5'GAGGAGAAGCCCGGTTTAATTTCCCCAGGTG3'). Amplicons were cloned into pET_{32} or pET_{46} expression vectors by ligation independent cloning (LIC) as described in Novagen's pET System manual (Novagen, 2005). Recombinant plasmid DNA was first transformed into non expression host *E. coli* NovaBlue cells by heat shock, grown over night at 37°C in liquid LB containing Carbenicillin at 50mg/mL , 80ug/mL X-gal, 0.5mM IPTG and then isolated using the WizardPlus Miniprep DNA Purification System (Promega, USA). DNA sequencing confirmed cloning into the correct reading frame prior to transformation into the bacterial expression host. Five nanograms of plasmid DNA were transformed into *E. coli* Origami 2(DE3) or BL21 by heat shock at 42°C for 30s following manufacturer's recommendations (Novagen, USA). The recombinant bacteria were plated on LB agar supplemented with 50µg/ml carbenicillin and incubated overnight at 37°C. The next morning a single colony forming unit was used to inoculate 100mL of fresh LB-carbenicillin (50µg/mL) liquid media and grown at 37°C with vigorous shaking until $OD_{600} \approx$ 0.8. Recombinant protein expression was induced by adding IPTG to a final concentration of 1mM and bacterial cultures were incubated at room temperature for 7h with shaking. Bacterial growth was monitored every hour. Following induction, the recombinant bacteria were isolated by centrifugation at 10,000g for 10 min and washed once with ice cold 20mM Tris-Cl. The remaining pellet was stored at -70°C until further protein purification. The bacterial cell pellet was resuspended in 20mM Tris-Cl (5mL/gr bacterial pellet), then lysed on ice using a Vibra cell VC50T sonicator (Sonics & Materials, Newtown CT, USA) at 50% intensity for 20 second intervals over 5-7 minutes. The resulting solution was centrifuged at 4°C for 15 min at 10,000g. Induction of the recombinant protein was confirmed by SDS-PAGE (5% stacking, 10% resolving) and stained with Coomassie brilliant blue R-250 (Bio-Rad, Canada).

Antibacterial activity assays:

Bacterial strains (Gram negative: *E. coli*, *S. marcescens* and Gram positive: *B. subtilis, S. epidermis, M. luteus*) were obtained from Fan Sozzi-Guo at Simon Fraser University's biology teaching laboratories. Bacterial cultures were grown on 5mL of liquid LB media at 37°C (or 25°C for *M. luteus*) with vigorous shaking. The next day, fresh LB agar plates were streaked with a sterile loop dipped in the overnight culture and incubated overnight at 37°C. A single colony forming unit from the LB agar plates was used to inoculate 3mL of LB liquid media and grown at 37°C until OD_{600} reached 1.0. A sterile cotton swab was dipped in these cultures and was used to inoculate fresh LB agar plates and create a homogeneous bacterial lawn.

Cell lysates from induced or un-induced recombinant bacteria ($\approx$1mg of total protein) were spotted on sterile 8mm diameter chromatography paper (Whatman, USA) circles and then placed on the bacterial lawns. Antibacterial activity was assessed the next morning after incubation at 25°C.

For a more precise analysis of Prolixin's activity spectrum, we used a liquid broth assay in 96 well plates. A 5mL over night liquid culture of each bacterial strain was obtained as above, then diluted to OD_{600} of 0.1 with poor Broth medium (Hetru and Bulet, 1997). Fifty microliters of cell lysate obtained above from induced or un-induced recombinant bacteria was mixed with 50µL of the diluted individual bacterial cultures in a sterile 96 well polystyrene microplate. The mixture was incubated at 37°C with shaking for six hours. Cell growth was monitored hourly using an Expert plus microplate reader (ASYS Hitech, Eugendorf, Austria) at 620nm.

Prolixin's activity in liquid culture against *E. coli* was assayed by comparing growth profiles of recombinant bacteria expressing Prolixin or a NF-κB transcription factor (Dorsal 1C) or a 300aa portion encoding the Rel Homology domain (RHD). These molecules were amplified using primers:

Rel-adF (5'gacgacgacaagatgaaccaatctgttcggaga3') and

Rel-adR (5'gaggagaagcccggtttagtttactttttgttgttcg3') or

RHD-adR (5'gaggagaagcccggttaagttaataattttgtattcg3') respectively, cloned and expressed as described above.

Results

Molecular characterization.

In our previous study to identify *R. prolixus* fat body and midgut genes expressed in response to bacterial or *T. cruzi* injection into the hemocoel, we found approximately 30% of the ESTs for which we were not able to predict any putative function (Ursic-Bedoya and Lowenberger, 2007). We compared these again to the constantly updated databases at the NCBI using Blast X against non redundant databases. Our query resulted in an identical significant hit for two of these ESTs. Clones 2.57 and 1.15 originated from the midgut SSH library (dbEST accession numbers EB084328 and EB084378) having sequence homology with a diptericin gene from the Hessian fly *Mayetolia destructor* (E value: 0.027). After aligning clones 2.57 (337bp) and 1.15 (180bp) we discovered that they were the same EST, albeit clone 2.57 was an incomplete digestion product of Rsa1 used to build the SSH library (Chapter 2). 5'-3' RACE was used to obtain both 5' and 3' UTR sequences which were missing to complete the full length gene. The complete cDNA consisted of 528bp with an open reading frame (ORF) of 297bp, a 5'UTR of 85bp and a 3'UTR of 146bp. The genomic sequence revealed a single 93bp intron (Figure 6-1).

Analysis of the ORF revealed a putative protein 98 amino acids long with an approximate molecular weight of 11kDa. The peptide contains 15 strongly basic (K,R), 12 strongly acidic (D,E), 27 hydrophobic (A,I,L,F,W,V) and 24 polar amino acids (N,C,Q,S,T,Y). Similarity search using BLAST-P reconfirmed the homology with *M. destructor*'s Diptericin (E value: 0.007). As other members of the Diptericin family, Prolixin is glycine rich (13%) however it only contains two proline residues which is in sharp contrast with other members of the family which contain a Proline rich region preceding the G region. The proline rich

domain is characteristic of other AMPs, which include Apidaecins, Drosocin, Metchinkowin, Lebocin and others (Imler and Bulet, 2005).

Further bioinformatic analysis identified conserved protein and functional domains of the putative peptide. The amino terminal end contains a 21 residue hydrophobic signal peptide, predicted to form an alpha helix. Additionally, in the carboxyl terminal fraction, homology to an Attacin C conserved motif (Pfam accession number: PF03769) expanding from residue K^{29} to the N^{98} (covering 70% of the pro-peptide) was found thus strongly suggesting the peptide encoded by this gene is an antimicrobial peptide (AMP).

We used tools from the Center for Biological Sequence Analysis (http://www.cbs.dtu.dk/services/) to identify common sites associated with post-translational modifications. Although *D. melanogaster* and *P. terranovae* diptericins each have two predicted O-linked glycosylation sites, none were identified in the prolixin molecule. Other post translational modification sites were found; two phosphorylation sites on residues S^{23} and T^{46}.

Protein sequence alignment with the few other members of the Diptericin family revealed that prolixin only has 33% identity with *M. destructor*, and only 20.8% with *D. melanogaster* and *P. terranovae* diptericins. However, there is a strong conservation of Glycine residues as well as conservation of residues S^{59}, D^{60} and R^{63} (Figure 6-2).

We used a phylogenetic analysis of Prolixin and other known glycine rich AMPs from different insect orders, using a method optimized for highly evolved gene families that have clear evolutionary relationships (Hein, 1990). The coleopteran glycine-rich AMPs exhibit an early divergence from the glycine rich AMPs found in the Attacin/Diptericin superfamily of molecules identified in the Diptera, Lepidoptera and Hemiptera.

Within the Attacin family, clustering of molecules follows insect taxonomical classification in lepidopterans and dipterans. Within the Diptericins, the phylogenetic tree suggests a common ancestor for Prolixin and the Diptericins, however Prolixin clearly belongs to a different taxon (Figure 6-3).

Promoter analysis.

We used a bioinformatics approach to identify prolixin's genomic clones containing promoter sequences. We used BLASTN to search the Trace Archive databases from the *R. prolixus* genome sequencing project. This search identified two contigs in the 5' upstream region of our genomic sequence and provided 516bp of proximal upstream sequence ahead of the start ATG codon. Fifty two potential transcription factor binding sites were identified in this sequence. Of particular interest were 4 putative GATA-1 sites (Table 1) given their relevance in insect immunity (Kadalayil *et al.*, 1997, Tingvall *et al.*, 2001, Senger *et al.*, 2006). Of note is that no putative NF-κB sites were identified.

Prolixin expression profile.

Our *in silico* analysis of the cDNA and deduced amino acid sequence strongly pointed out to the identification of a diptericin like AMP in *R. prolixus* midgut. To further strengthen this hypothesis, we analyzed the transcriptional profile of this gene first in different tissues and subsequently in response to different immune challenges.

A qualitative RT-PCR using RNA collected from five eggs or tissue from a single adult insect revealed that diptericin is constitutively produced and is up-regulated in response to bacteria in the hemocoel. Midgut and fat body tissues had the strongest detectable transcript levels which concurs with the initial discovery of the *Prolixin* EST and the fact that the fat body is the major organ for AMP biosynthesis in insects. Eggs (red colored, 10 days after hatching) did not produce a detectable signal, whereas all other tissues examined did (Figure 6-4), including the salivary glands.

The apparent up regulation in the fat body and midgut in response to bacterial challenge led us to use qPCR to quantify Prolixin's transcripts levels in response to different stimuli in the Fat body. We artificially infected adult insects by injecting them with bacteria, *T. cruzi* promastigotes, LIT media or by poking them with a sterile needle to control for the immune response generated by cuticular damage. Transcripts levels were compared to a naïve (non-challenged) sample and the results were reported as fold increase over this sample (Figure 6-5). Cuticular damage caused a small increase in transcript levels, whereas bacterial presence in the hemolymph caused a significant ($\approx$200 fold) increase over the non challenged samples. Injection of 5µL of sterile LIT (a rich media) caused a similar increase to that of bacterial challenge; however the presence of *T. cruzi* caused a very significant increase (over 500 fold) over the naïve control and about three times that of LIT injected.

Recombinant protein activity.

The activity spectrum of the recombinant Prolixin was assayed in different formats. First we assayed the activity of the recombinant protein using a diffusion assay against four different bacterial strains using two different expression vectors. Two Gram positive and two gram negative strains were tested using either pET_{32} or pET_{46}. Novagen's pET_{46} adds an N-terminal His tag to the recombinant protein whereas pET_{32} adds a 17kDa N-terminal tag which enhances protein solubility and facilitated protein detection during SDS PAGE. In this assay, the least sensitive, antibacterial activity was detected against *B. subtilis* and marginally against *E. coli* (Figure 6-6).

To further investigate Prolixin's *in vitro* antibacterial properties we set up a more sensitive assay in 96 well format using cell lysates from induced or uninduced recombinant bacteria mixed with diluted bacterial cultures. After six hours of incubation at 37°C with shaking, OD_{600} was measured, and negative effects on bacterial growth were apparent for *B. subtilis* and to a lesser extent for *E. coli*, which supports our results on the plate assay. The activity against *M. luteus* is

not evident due to poor bacterial growth under assay conditions and no antibacterial activity was seen against *S. epidermis* or *S. marcescens* (Figure 6-7). Induction of prolixin over expression in liquid culture showed Prolixin's toxicity to *E. coli*. Although protein over expression reduced bacterial growth in comparison to non induced bacteria, recombinant bacteria expressing an exogenous transcription factor did not show the bacteriostatic effect seen for Prolixin (Figure 6-8).

Discussion

We have identified and characterized Prolixin, a glycine-rich peptide with strongest homology to members of the Attacin/Diptericin superfamily of AMPs, and which represents the first diptericin-like AMP from a non-dipteran insect.

Initial *in silico* analysis of the amino acid sequence indicates that Prolixin is not glycosylated, in contrast to similar glycine-rich molecules. All fly Diptericins and Drosocin are naturally glycosylated. However, glycosylation is not a *sine qua non* condition for antibacterial activity. Chemically synthesized glycosylated and non-glycosylated forms were shown to retain activity against gram negative bacteria *E. coli D22, 1106* and *S. typhimurium* (Cudic *et al.*, 1999, Winans *et al.*, 1999). The exact role of glycosylation of AMPs is not known and further research should address the affinity between sugar moieties of AMPs and those on the bacterial surface, as a means to target activity and increase killing specificity.

While proline- and glycine-rich peptides like Diptericin are predominantly active against Gram-negative bacteria (Keppi *et al.*, 1989), Prolixin, like cecropins, is active against both Gram positive and Gram-negative bacteria (Moore *et al.*, 1996, Lowenberger *et al.*, 1999). In addition to their antibacterial properties, proline-rich peptides have been evaluated for anti-parasite activity. Purified recombinant AttacinA1 from *Glossina* sp. has inhibitory effects against both the mammalian bloodstream forms and the insect stages of *T. brucei* and strong

antimicrobial activity against *E. coli-K12*, but not against the enteric gram-negative symbiont of tsetse, *Sodalis glossinidius* (Hu and Aksoy, 2005). Similarly, Hao *et al.* (2001) showed that *Trypanosoma brucei rhodesiense* was susceptible to the Tse-Tse fly Diptericin *in vitro* at relatively high concentrations (IC_{50} of 10μM), whereas susceptible bacteria (i.e. *E. coli* DH5α) had an IC_{50} of 0.2μM (Hao *et al.*, 2001). *Trypanosoma cruzi* is susceptible to AMPs as demonstrated in laboratory conditions using transgenic bacteria to over express Cecropin A, a pore-forming AMP of *Hyalophora cecropia* (Durvasula *et al.*, 1997). Cecropins, however have not been identified in hemipterans, and we do not know if endogenous *R. prolixus* AMPs can regulate *T. cruzi* development. There may not be sufficient concentrations of lethal AMPs in the midgut of the vector to kill *T. cruzi*, or these AMPs may be affected by the digestive proteases. Alternatively the vector may not express high concentrations of AMPs because these might kill the obligate bacterial symbionts that reside in the midgut and on which the vector is dependent for survival. Antimicrobial peptides in the gut might be used as a mechanism to regulate bacterial populations within acceptable levels without eliminating all symbionts. The fact that we find transcriptional activation of several AMPs (Defensin, Lysozyme and Prolixin) in gut tissues indirectly supports this hypothesis.

Prolixin's deduced amino acid sequence suggests a shorter and simpler structure than that of other glycine-rich AMPs. Diptericins comprise a hydrophobic N terminal signal peptide, a proline-rich (P domain) sequence and a glycine-rich moiety (G domain) in the C terminus. Likewise, Attacins contain a signal peptide, a P domain and two G domains (Hedengren *et al.*, 2000). This is not the case for Prolixin, where the P domain is missing and where only a signal peptide and a single G domain are noticeable. Further evidence that the P domain is missing is that the conserved furin cleavage site, found in Diptericins and Attacins is missing. Furin functions to cleave other proteins downstream a canonical sequence, R-X-(R/K)-R, into their mature/active forms; and it has been

168

suggested that *D. melanogaster* Diptericin B is activated after removal of both the signal peptide and the P domain in a two step process.

Despite the relatively low sequence similarity in the N terminal end among Prolixin, Diptericins and Attacins, we believe these peptides share a potential common origin because of their glycine-rich C-terminal region. As several insect genomes have been sequenced, comparative genomic studies on immune genes have revealed interesting evolutionary characteristics of immune gene families implicated in the three main phases of the immune response. Signal transduction genes are maximally divergent in sequence. This is not the case with effector molecules such as AMPs, which have diversified by gene duplication producing new families of compounds and even species-specific genes (Waterhouse *et al.*, 2007). Certain types of AMPs, however, are omnipresent in phylogenetically very diverse organisms. Defensins are small cationic, cysteine-rich peptides which possess activity against Gram positive bacteria, and have been described in invertebrates, vertebrates and plants. Cecropins are cationic linear peptides active against gram negative bacteria and are found in selected insect orders, Coleoptera, Diptera and Lepidoptera but not yet in ancient hemimetabolous insects such as hemipterans. Prolixin appears to be a simpler molecule, possibly a precursor of the Diptericin family of antimicrobial peptides which, to date, have only been characterized in the higher insect order Diptera.

A glycine-rich N-terminal end motif is common to Prolixin, Diptericins and Attacins. It has been suggested that Diptericins and Attacins evolved from a common ancestor (Hedengren *et al.*, 2000). However, Attacins and Diptericins have proline-rich domains, found in other antimicrobial peptides such as Drosocin and Metchnikowin, but these are missing in Prolixin. In view of these observations and if we assume a modular structure of the different motifs of AMPs we could speculate that the addition of a proline-rich section (P domain) gave rise to Diptericins in flies and that a subsequent gene duplication added the second, C terminal glycine-rich motif found in Attacins. A similar mechanism has been suggested in the molecular evolution of Sarcotoxin II of *Sarcophaga*

peregrina, where this molecule evolved from an attacin-like ancestor by the addition of two novel domains. A proline-rich extension in the N terminal, and a sequence insertion near the amino terminal of the G1 domain differentiate these two AMPs (Asling *et al.*, 1995). As more genomes from diverse insect groups become available for comparison we will be able to validate or disprove the theories of the origins and interrelationships among these similar, but distinct AMPs.

The temporal and spatial expression of antimicrobial peptides is tightly regulated via complex intracellular pathways. In *Drosophila*, AMP expression is regulated via two main signaling pathways, Toll and Imd. Both of these pathways culminate in the activation of NFκB transcription factors. The Toll pathway leads to the activation of two *Drosophila* NFκB transcription factor homologs, Dorsal and Dif, resulting in the transcriptional activation of antifungal Metchnikowin and Drosomycin. The Imd pathway, through the activation of another NFκB homolog, Relish, activates the expression of AMPs Cecropin, Drosocin, Defensin and Diptericin (Lemaitre and Hoffmann, 2007). Promoter mapping of *D. melanogaster's* diptericin gene revealed the presence of proximal DNA motifs required for transcriptional activation. These include a combination of NF-κB and GATA binding sites, where GATA sites are recognized by the transcription factor Serpent (Kadalayil *et al.*, 1997, Tingvall *et al.*, 2001). We were not able to amplify a Serpent homolog in *R. prolixus* using *Drosophila* cactus degenerate primers, but the release of the *R. prolixus* genome should allow us to identify members of this family of transcription factors in this hemipteran insect. The absence of NF-κB sites within 500bp of the prolixin start methionine codon is unusual as these sites have been described in the promoter regions of many insect AMP genes including Diptericins (Georgel *et al.*, 1993, Meister *et al.*, 1994). Recently, we identified both NF-κB and GATA-1 sites in the promoter region of two *R. prolixus* lysozyme genes (Ursic-Bedoya *et al.*, 2008) as well as in several other immune-related genes (Ursic-Bedoya unpublished). At this point we can only speculate about the existence of a novel signaling pathway in hemipterans which would not

be regulated by NFκB transcription factors. Toll and Imd receptors homologues in hemipteran insect have not yet been identified; however, insect comparative genomics studies point out to the strong conservation of intracellular immune signaling pathways (Toll, Imd and JAK/SAT) and for the evolutionary diversification of other gene families like PGRPs, clip-domain proteins, serpins, Toll-related receptors, and AMPs (Waterhouse *et al.*, 2007, Zou *et al.*, 2007).

Most AMPs are synthesized by the fat body and released into the hemolymph in response to pathogen invasion. Prolixin's transcriptional pattern concurs with that of other AMPs, which are predominantly expressed in the fat body but also in other tissues which come into contact with foreign organisms, such as the midgut epithelium. Upon oral infection with *Ecc15* and *P. entomophila*, *Drosophila* responds through a systemic immune response initiated in the gut which signals to, and initiates, the fat body production of AMPs. In triatomes, similar responses have been shown for defensins (Lopez *et al.*, 2003, Araujo *et al.*, 2006) and lysozymes (Kollien *et al.*, 2003, Balczun *et al.*, 2008, Ursic-Bedoya *et al.*, 2008). Prolixin also is expressed in the midgut, and the concept that the insect gut is an immune competent tissue is not new. In *Drosophila*, Diptericins or Attacins are produced by epithelial cells under the control of the Imd pathway upon recognition of PGN released by Gram-negative bacteria and a local epithelial response can trigger a systemic response by the fat body. In hemipterans, such as *R. prolixus*, bacterial symbionts are obligatory residents of the intestinal tract. They are responsible for vitamin B procurement, and their elimination via aposymbiosis has severe developmental effects, that can be reversed by a diet rich in vitamin B (Schaub and Eichler, 1998). Several bacterial species reside in the intestinal tract of *R. prolixus*; *Rhodococcus rhodnii,* the main symbiont, is found primarily in the cardia and anterior midgut (stomach). Upon a blood meal, hemolysis and water adsorption occurs in the anterior midgut which initially decreases the bacterial population. Within days as the blood meal is circulated to the posterior midgut (intestine) for proteolysis by cathepsins, the bacterial population increases significantly (Eichler and Schaub, 2002, Azambuja *et al.*,

2004). As the intestine is relatively free of bacteria, AMPs such as Prolixin may be released into this compartment for two functions: to aid in digestion of blood products and as a local immune response to prevent massive bacterial proliferation.

Bacteriolysis would occur in the intestine to prevent bacteria from sequestering nutrients obtained from blood meal digestion and to release vitamins essential to the development of the insect. Antimicrobial peptides participating in the digestive process include lysozymes in Drosophila (Kylsten *et al.*, 1992, Daffre *et al.*, 1994) and we recently characterized a lysozyme from *R. prolixus* which we believe has the same function (Ursic-Bedoya *et al.*, 2008).

The forthcoming release of the annotated genome sequence of *R. prolixus* should allow for the identification of more components of the still largely unknown puzzle that comprises the immune system in this and other hemimetabolous insects. New post-genomic technologies including proteomics and RNA interference (RNAi), which has been used recently in this hemipteran insect to study gene function (Araujo *et al.*, 2006b), have expanded our knowledge of immune system regulation in medically relevant insect vectors. Furthermore, comparative genomic studies with an evolutionary ecology perspective should help us delineate the evolution of the innate immune system in general, on specific AMPs such as Prolixin, and will allow us to speculate on ways we could use this knowledge to better address and reduce vector borne diseases which cause so much morbidity and mortality worldwide.

Figures

```
tatttttttt ttttttttttg gggtntcctc ggcctntcng ctgtgtacaa actataacga
gtgaactata agacaaaaag aaaaa
M    F    K    T    I    L    V    L    V    V    A    L    N    L    L    G    A
ATG  TTC  AAA  ACA  ATT  TTG  GTG  CTG  GTT  GTC  GCC  CTG  AAC  TTG  CTT  GGC  GCA
F    T    Y    A    S    P    R    P    D    D    K    K    N    Q    G    S    A
TTC  ACT  TAT  GCA  AGT  CCG  AGG* CCT  GAT  GAC  AAG  AAA  AAT  CAA  GGA  AGT  GCC
S    V    D    V    Q    N    E    R    G    E    G    T    K    V    D    A    R
TCA  GTA  GAC  GTG  CAG  AAC  GAA  CGA  GGA  GAA  GGT  ACC  AAA  GTT  GAT  GCA  AGA
V    R    Q    E    L    W    R    S    D    D    G    R    T    R    A    Q    A
GTT  AGA  CAA  GAA  TTG  TGG  CGA  AGT  GAC  GAT  GGA  AGG  ACC  AGA  GCT  CAA  GCT
Y    G    H    W    D    R    T    Y    G    G    R    N    H    G    E    R    S
TAT  GGA  CAT  TGG  GAT  AGA  ACA  TAC  GGT  GGA  CGA  AAC  CAC  GGT  GAA  AGG  AGT
Y    G    G    G    M    R    I    E    H    T    W    G    N    stop
TAC  GGT  GGC  GGA  ATG  AGG  ATA  GAA  CAC  ACC  TGG  GGA  AAT  TAA
ataaaattat tgataagtac caaaaattga attgcttaag tgataatttg taatttgtta
ccgccattaa acacgcttaa atttaaataa atatgttcat aataaagttc atagctaaaa
tgttttatgt tgtcaaaaaa aaaaaa

*gtaagtggtc atttcttaag taaccatatc aaatggtata aaaaaaaaaa tctacaatta
aataaatgac cgtctcataa ttttcttaaa tag
```
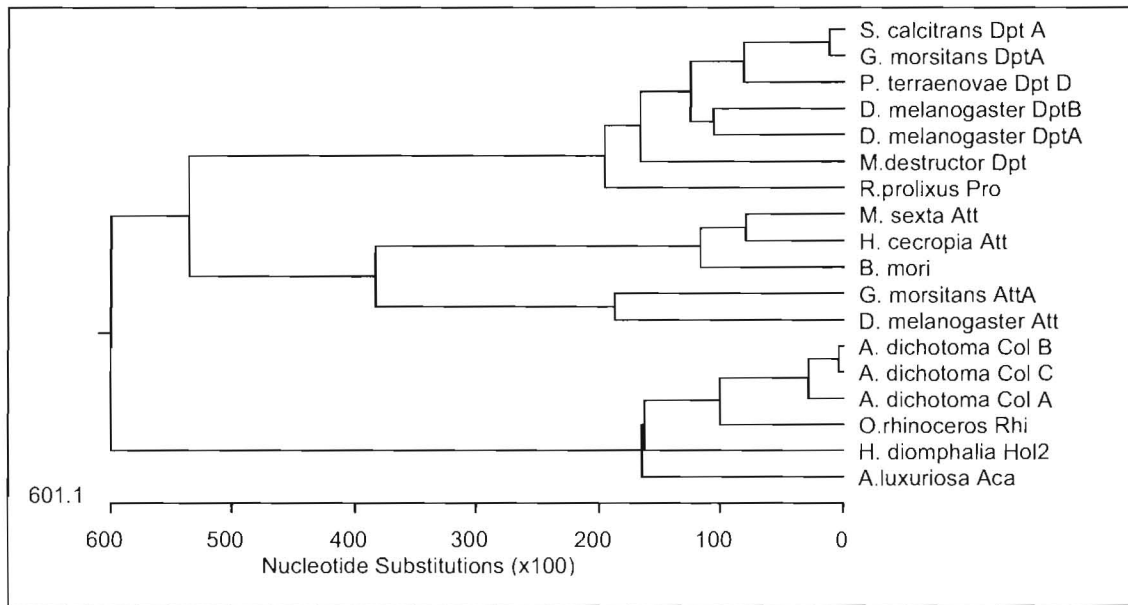
Figure 6-1 *R. prolixus* prolixin gene, genomic and deduced amino acid sequence.

Analysis of the ORF revealed a putative protein 98 amino acids long with an approximate molecular weight of 11kDa. A 21 hydrophobic signal peptide (underlined) is predicted to from an alpha helix; and highlighted is an Attacin C motif (Pfam accession number: PF03769) expanding from residue K^{29} to the N^{98}. Putative phosphorylation sites are in bold. *indicates intron insertion site. UTRs and intron sequences are in lowercase.

Figure 6-2 Prolixin alignment with other members of the diptericin family.

Multiple sequence analysis was performed using Clustal W with Lasergene's MegAlign module (DNAstar, USA). Sequences used in the alignment were: *Protophormia terraenovae* Diptericin (CAB57822.1); *Mayetiola destructor* Diptericin (ABG21230.1); *Drosophila melanogaster* Diptericin A (NP_476808.1), B (NP_523787.2) *Glossina morsitans morsitans* Diptericin A (AAL34111.1) *Stomoxys calcitrans* Diptericin A (AAY98016.1), and *Rhodnius prolixus* Prolixin (bankit1060780 EU448993). Boxed amino acids represent identities to the consensus and residue numbering is based on Prolixin.
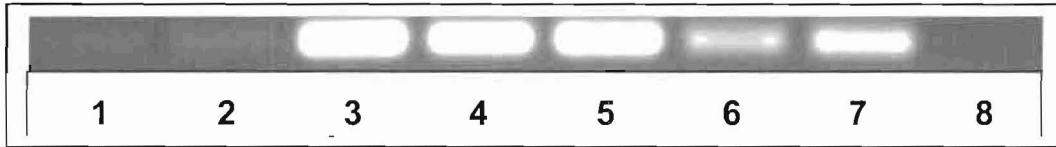
**Figure 6-3: Phylogenetic reconstruction of insect glycine-rich AMPs.**

Diptericin, Attacin and coleopteran families of glycine-rich peptides were analyzed. Signal peptides were removed from each precursor molecule, and then aligned with DNAstar Megalign's Jotun Hein Method (PAM250).

Figure 6-4 Prolixin's transcript tissue distribution.

Qualitative RT-PCR using RNA collected from five eggs or tissue from a single adult insect revealed that diptericin is constitutively produced and is up-regulated in response to bacteria in the hemocoel. Midgut and fat body tissues had the strongest detectable transcript levels which concurs with the initial discovery of the Prolixin EST and the fact that the fat body is the major organ for AMP biosynthesis in insects.1: Egg; 2: Salivary glands; 3: Thorax; 4: Fat body; 5: Fat body 24h post bacterial injection; 6: Midguts; 7: midguts 24h post bacterial injection; 8: no template control.
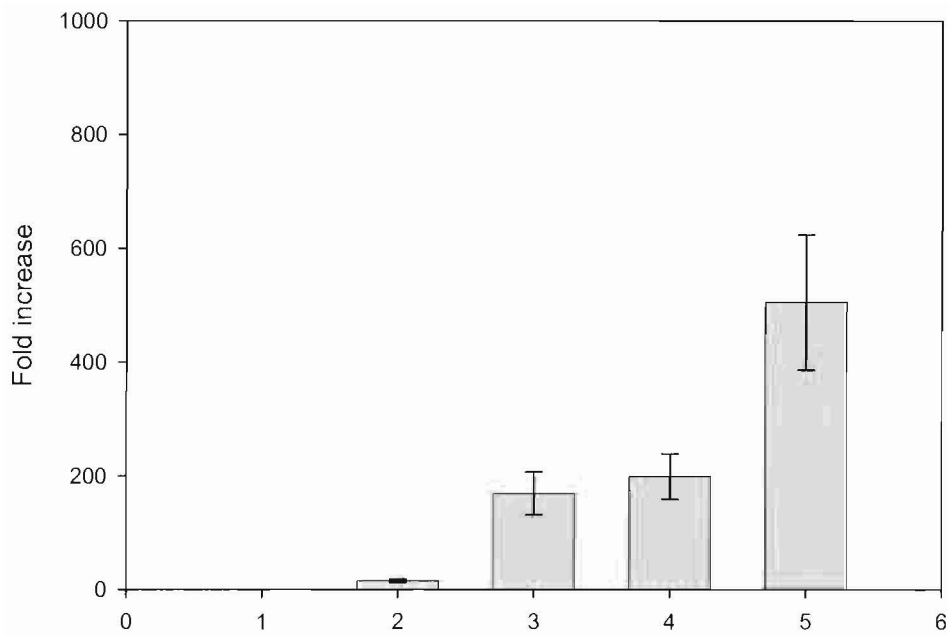
Figure 6-5 Prolixin's transcript levels in fat body after immune stimulation.

1: Naïve (normalizer); 2: Sterile injected; 3: Bacteria injected; 4: LIT injected; 5: LIT-*T. cruzi* injected. Data was quantified using the $2^{-(DDCt)}$ method. A Beta actin fragment was used to normalize amongst samples as described (Lopez et al., 2003). Error bars represent s.e. of three independent experiments with each sample in triplicate.
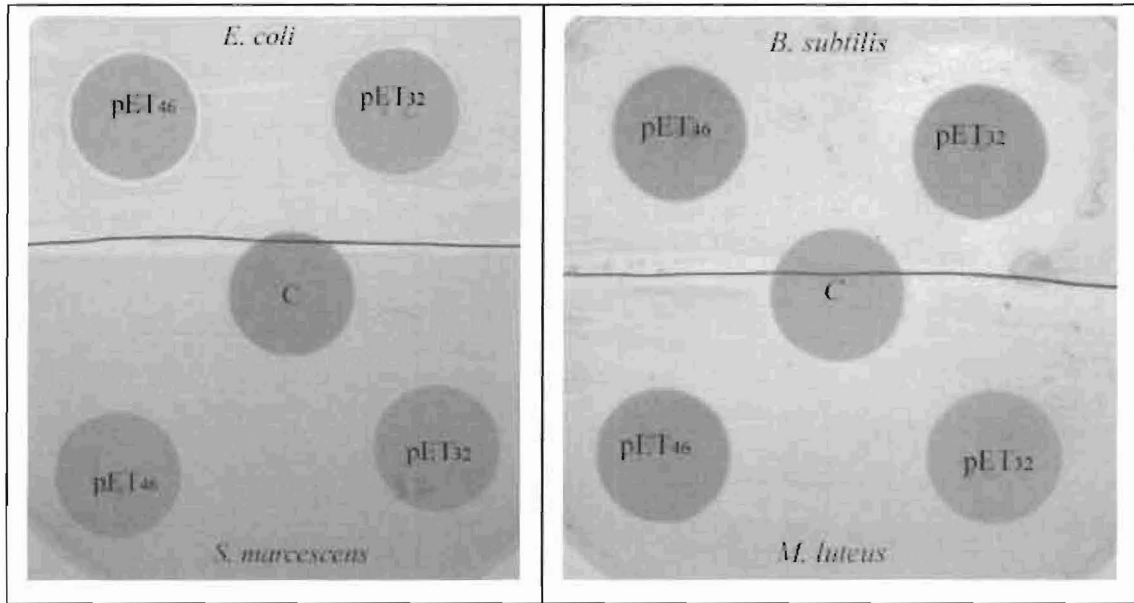
Figure 6-6 Recombinant Prolixin zone inhibition assays:

Cell lysates from induced or un-induced recombinant bacteria (≈1mg of total protein) were spotted on sterile 8mm diameter chromatography paper (Whatman, USA) circles and then placed on the bacterial lawns. Antibacterial activity was assessed the next morning after incubation at 25°C. The level of inhibition in these assays, as measured by the radius of the inhibition was compared with zones of inhibition of known concentrations of ampicillin. Left panel: Activity against Gram negative bacteria (*E. coli* and *S. marcescens*) Right panel: activity against Gram positive bacteria (*B. subtilis* and *M. luteus*). C: control containing Tris buffer. *R. prolixus* prolixin was expressed in *E. coli* BL21 cells using expression vectors pET_{46} or pET_{32}
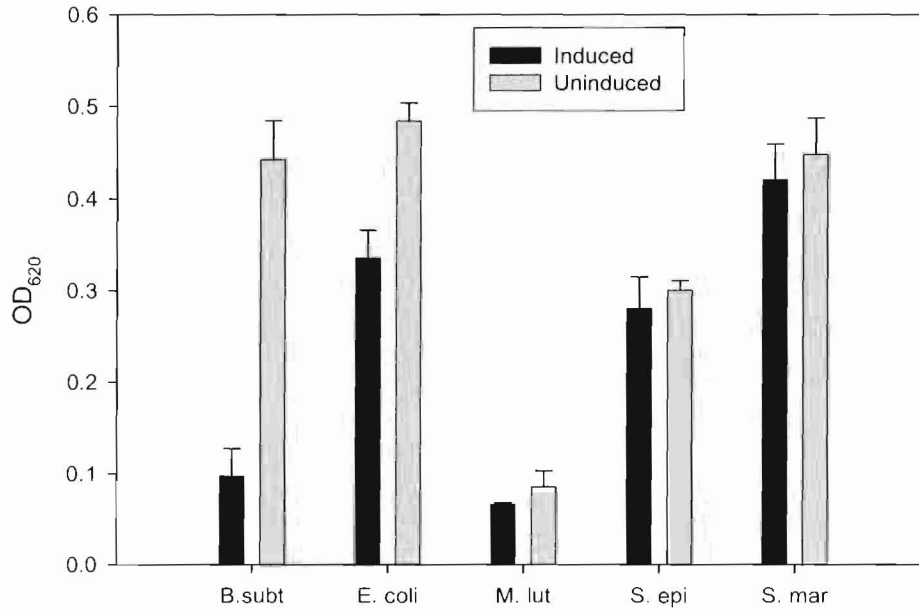
Figure 6-7 Recombinant Prolixin's activity spectrum.

Cell lysates from Induced or Uninduced *E. coli* BL21 carrying Prolixin-pET$_{46}$ were incubated with a diluted bacterial suspension (logarithmic growth phase) of each strain at 37°C for 6 hours. OD$_{620}$ was recorded. Bacterial strains used, Gram negative: *E. coli*, *S. marcescens* and Gram positive: *B. subtilis*, *S. epidermis* and *M. luteus*. Error bars represent s.d. of two independent experiments with each sample assayed in six different wells.
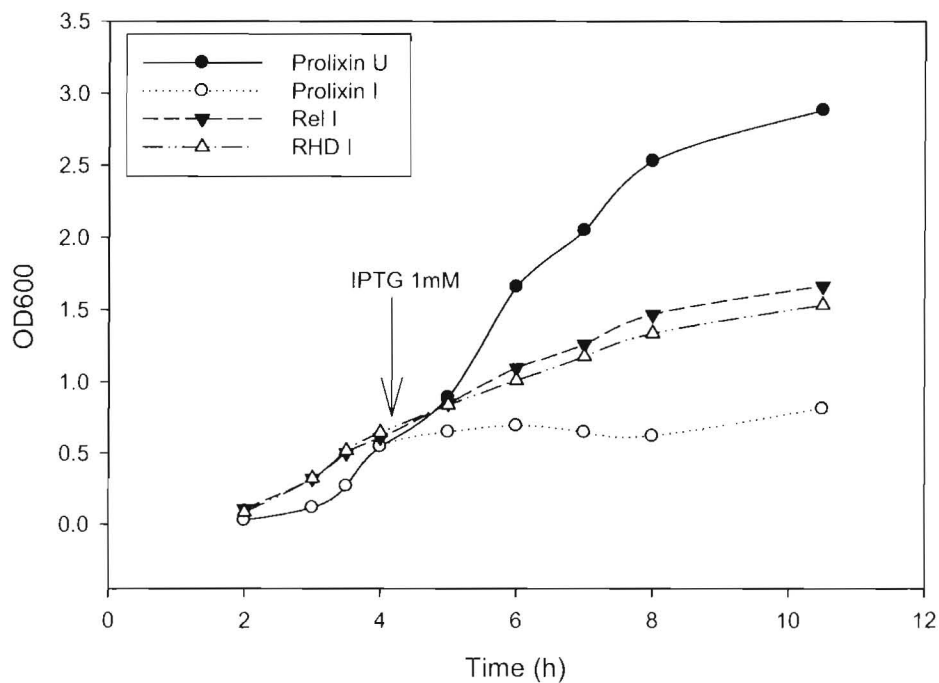
Figure 6-8 Prolixin's activity in liquid culture against *E. coli*.

Bacterial growth (room temperature) was monitored every hour by measuring OD_{600} and recombinant protein production was induced with 1mM IPTG at OD_{600}=0.8.
RHD: Rel homology domain; Rel I: Dorsal 1C; I: induced; U: Un-induced.

Tables

Genomic clone	GATA sequence	Location	Origin
NAAX-adg52b08.b1	GAGATAGATA	-488	Trace archive
NAAX-ado90e10.b1	AAGCTGATAAAA	-380	
	CTTATCTCGTG	-292	
	CGGAGATAGAT	-249	

Table 6-1 Prolixin's promoter putative transcription factor binding sites.

Putative transcription binding sites were identified using Alibaba 2.1 software, which predicts transcription factor binding sites by context dependent matrices generated from TRANSFAC public sites. Putative binding sites location is given relative to the start methionine codon.

References

Araujo, C.A., Waniek, P.J., Stock, P., Mayer, C., Jansen, A.M. & Schaub, G.A., 2006. Sequence characterization and expression patterns of defensin and lysozyme encoding genes from the gut of the reduviid bug *Triatoma brasiliensis*. *Insect biochemistry and molecular biology*, 36, 547-60.

Araujo, R.N., Santos, A., Pinto, F.S., Gontijo, N.F., Lehane, M.J. & Pereira, M.H., 2006b. RNA interference of the salivary gland nitrophorin 2 in the triatomine bug *Rhodnius prolixus* (Hemiptera: Reduviidae) by dsRNA ingestion or injection. *Insect biochemistry and molecular biology*, 36, 683-93.

Asling, B., Dushay, M.S. & Hultmark, D., 1995. Identification of early genes in the *Drosophila* immune response by PCR-based differential display: the Attacin A gene and the evolution of attacin-like proteins. *Insect biochemistry and molecular biology*, 25, 511-8.

Azambuja, P., Feder, D. & Garcia, E.S., 2004. Isolation of *Serratia marcescens* in the midgut of *Rhodnius prolixus*: impact on the establishment of the parasite *Trypanosoma cruzi* in the vector. *Experimental parasitology*, 107, 89-96.

Balczun, C., Knorr, E., Topal, H., Meiser, C.K., Kollien, A.H. & Schaub, G.A., 2008. Sequence characterization of an unusual lysozyme gene expressed in the intestinal tract of the reduviid bug *Triatoma infestans* (Insecta). *Parasitology research*, 102, 229-32.

Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S. & Brunak, S., 2004. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, 4, 1633-49.

Borges, A.R., Santos, P.N., Furtado, A.F. & Figueiredo, R.C., 2007. Phagocytosis of latex beads and bacteria by hemocytes of the triatomine bug *Rhodnius prolixus* (Hemiptera: Reduvidae). *Micron*.

Brener, Z., 1973. Biology of *Trypanosoma cruzi*. *Annual review of microbiology*, 27, 347-82.

Bulet, P. & Stocklin, R., 2005. Insect antimicrobial peptides: structures, properties and gene regulation. *Protein and peptide letters*, 12, 3-11.

Chen, Y., Weng, Z.-H. & Zheng, L., 2008. Innate immunity against malaria parasites in *Anopheles gambiae*. *Insect Science*, 15, 45-52.

Cudic, M., Bulet, P., Hoffmann, R., Craik, D.J. & Otvos, L., Jr., 1999. Chemical synthesis, antibacterial activity and conformation of diptericin, an 82-mer peptide originally isolated from insects. *European journal of biochemistry,* 266, 549-58.

Daffre, S., Kylsten, P., Samakovlis, C. & Hultmark, D., 1994. The lysozyme locus in *Drosophila melanogaster:* an expanded gene family adapted for expression in the digestive tract. *Molecular & general genetics,* 242, 152-62.

Dimarcq, J.L., Keppi, E., Dunbar, B., Lambert, J., Reichhart, J.M., Hoffmann, D., Rankine, S.M., Fothergill, J.E. & Hoffmann, J.A., 1988. Insect immunity. Purification and characterization of a family of novel inducible antibacterial proteins from immunized larvae of the dipteran *Phormia terranovae* and complete amino-acid sequence of the predominant member, diptericin A. *European journal of biochemistry,* 171, 17-22.

Durvasula, R.V., Gumbs, A., Panackal, A., Kruglov, O., Aksoy, S., Merrifield, R.B., Richards, F.F. & Beard, C.B., 1997. Prevention of insect-borne disease: an approach using transgenic symbiotic bacteria. *Proceedings of the National Academy of Sciences of the United States of America,* 94, 3274-8.

Eichler, S. & Schaub, G.A., 2002. Development of symbionts in triatomine bugs and the effects of infections with trypanosomatids. *Experimental parasitology,* 100, 17-27.

Emanuelsson, O., Brunak, S., Von Heijne, G. & Nielsen, H., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature protocols,* 2, 953-71.

Georgel, P., Meister, M., Kappler, C., Lemaitre, B., Reichhart, J.M. & Hoffmann, J.A., 1993. Insect immunity: the diptericin promoter contains multiple functional regulatory sequences homologous to mammalian acute-phase response elements. *Biochem Biophys Res Commun,* 197, 508-17.

Grabe, N., 2002. AliBaba2: context specific identification of transcription factor binding sites. *In silico biology,* 2, S1-15.

Hansen, J.E., Lund, O., Tolstrup, N., Gooley, A.A., Williams, K.L. & Brunak, S., 1998. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconjugate journal,* 15, 115-30.

Hao, Z., Kasumba, I., Lehane, M.J., Gibson, W.C., Kwon, J. & 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, 12648-53.

Hedengren, M., Borge, K. & Hultmark, D., 2000. Expression and evolution of the *Drosophila* attacin/diptericin gene family. *Biochemical and biophysical research communications,* 279, 574-81.

Hein, J., 1990. Unified approach to alignment and phylogenies. *Methods in enzymology,* 183, 626-45.

Hetru, C. & Bulet, P., 1997. Strategies for the isolation and characterization of antimicrobial peptides of invertebrates. *In* W.M. Shafer (ed.) *Methods in molecular biology.* Totowa: Humana press, 35-49.

Hu, Y. & Aksoy, S., 2005. An antimicrobial peptide with trypanocidal activity characterized from *Glossina morsitans morsitans. Insect biochemistry and molecular biology,* 35, 105-15.

Imler, J.L. & Bulet, P., 2005. Antimicrobial peptides in *Drosophila*: structures, activities and gene regulation. *Chemical immunology and allergy,* 86, 1-21.

Kadalayil, L., Petersen, U.M. & Engstrom, Y., 1997. Adjacent GATA and kappa B-like motifs regulate the expression of a *Drosophila* immune gene. *Nucleic acids research,* 25, 1233-9.

Keppi, E., Pugsley, A.P., Lambert, J., Wicker, C., Dimarcq, J.L., Hoffmann, J. & Hoffmann, D., 1989. Mode of action of diptericin A, a bactericidal peptide induced in the hemolymph of *Phormia terranovae* larvae. *Archives of Insect Biochemistry and Physiology,* 10, 229-239.

Kollien, A.H., Fechner, S., Waniek, P.J. & Schaub, G.A., 2003. Isolation and characterization of a cDNA encoding for a lysozyme from the gut of the reduviid bug *Triatoma infestans. Archives of insect biochemistry and physiology,* 53, 134-45.

Kylsten, P., Kimbrell, D.A., Daffre, S., Samakovlis, C. & Hultmark, D., 1992. The lysozyme locus in *Drosophila melanogaster*: different genes are expressed in midgut and salivary glands. *Molecular & general genetics,* 232, 335-43.

Lavine, M.D. & Strand, M.R., 2002. Insect hemocytes and their role in immunity. *Insect biochemistry and molecular biology,* 32, 1295-309.

Lemaitre, B. & Hoffmann, J., 2007. The host defense of *Drosophila melanogaster. Annual review of immunology,* 25, 697-743.

Livak, K.J. & Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* 25, 402-8.

Lopez, L., Morales, G., Ursic, R., Wolff, M. & Lowenberger, C., 2003. Isolation and characterization of a novel insect defensin from *Rhodnius prolixus*, a vector of Chagas disease. *Insect biochemistry and molecular biology,* 33, 439-47.

Lowenberger, C., Charlet, M., Vizioli, J., Kamal, S., Richman, A., Christensen, B.M. & Bulet, P., 1999. Antimicrobial activity spectrum, cDNA cloning, and mRNA expression of a newly isolated member of the cecropin family from the mosquito vector Aedes *aegypti. The Journal of biological chemistry,* 274, 20092-7.

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., Deweese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Yamashita, R.A., Yin, J.J., Zhang, D. & Bryant, S.H., 2005. CDD: a Conserved Domain Database for protein classification. *Nucleic acids research,* 33, D192-6.

Meister, M., Braun, A., Kappler, C., Reichhart, J.M. & Hoffmann, J.A., 1994. Insect immunity. A transgenic analysis in *Drosophila* defines several functional domains in the diptericin promoter. *Embo J,* 13, 5958-66.

Moore, A.J., Beazley, W.D., Bibby, M.C. & Devine, D.A., 1996. Antimicrobial activity of cecropins. *The Journal of antimicrobial chemotherapy,* 37, 1077-89.

Novagen, 2005. pET System Manual. 11 ed.

Schaub, G.A. & Eichler, S., 1998. The effects of aposymbiosis and of an infection with Blastocrithidia triatomae (Trypanosomatidae) on the tracheal system of the reduviid bugs *Rhodnius prolixus* and *Triatoma infestans. Journal of insect physiology,* 44, 131-140.

Senger, K., Harris, K. & Levine, M., 2006. GATA factors participate in tissue-specific immune responses in *Drosophila* larvae. *Proceedings of the National Academy of Sciences of the United States of America,* 103, 15957-62.

Strand, M.R., 2008. The insect cellular immune response. *Insect Science,* 15, 1-14.

Tingvall, T.O., Roos, E. & Engstrom, Y., 2001. The GATA factor Serpent is required for the onset of the humoral immune response in *Drosophila* embryos. *Proceedings of the National Academy of Sciences of the United States of America,* 98, 3884-8.

Ursic-Bedoya, R.J. & Lowenberger, C.A., 2007. *Rhodnius prolixus*: identification of immune-related genes up-regulated in response to pathogens and parasites using suppressive subtractive hybridization. *Developmental and comparative immunology,* 31, 109-20.

Ursic-Bedoya, R.J., Nazzari, H., Cooper, D., Triana, O., Wolff, M. & Lowenberger, C., 2008. Identification and characterization of two novel lysozymes from *Rhodnius prolixus*, a vector of Chagas disease. *Journal of insect physiology,* 54, 593-603.

Waterhouse, R.M., Kriventseva, E.V., Meister, S., Xi, Z., Alvarez, K.S., Bartholomay, L.C., Barillas-Mury, C., Bian, G., Blandin, S., Christensen, B.M., Dong, Y., Jiang, H., Kanost, M.R., Koutsos, A.C., Levashina, E.A., Li, J., Ligoxygakis, P., Maccallum, R.M., Mayhew, G.F., Mendes, A., Michel, K., Osta, M.A., Paskewitz, S., Shin, S.W., Vlachou, D., Wang, L., Wei, W., Zheng, L., Zou, Z., Severson, D.W., Raikhel, A.S., Kafatos, F.C., Dimopoulos, G., Zdobnov, E.M. & Christophides, G.K., 2007. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science,* 316, 1738-43.

Who, 2002a. *Control of Chagas disease: second report of a WHO expert committee* Geneva: World Health Organization.

Who, 2002b. *The world health report 2002: Reducing risks, promoting healthy life.* Geneva: World Health Organization.

Wigglesworth, V.B., 1936. Symbiotic bacteria in a blood-sucking insect, *Rhodnius prolixus* Stahl (Hemiptera, Triatomidae). *Parasitology,* 284-289.

Winans, K.A., King, D.S., Rao, V.R. & Bertozzi, C.R., 1999. A chemically synthesized version of the insect antibacterial glycopeptide, diptericin, disrupts bacterial membrane integrity. *Biochemistry,* 38, 11700-10.

Zou, Z., Evans, J.D., Lu, Z., Zhao, P., Williams, M., Sumathipala, N., Hetru, C., Hultmark, D. & Jiang, H., 2007. Comparative genomic analysis of the *Tribolium* immune system. *Genome biology,* 8, R177.

CHAPTER 7: CONCLUSIONS

URSIC BEDOYA Raul
Simon Fraser University
Department of Biological Sciences

Conclusion and future perspectives

Much of the work done by molecular biologists since the 1970's has been focused on describing genes and their functions and to link these genes into pathways and networks to explain the molecular processes that govern living cells and organisms. However, with the development of new technologies, we have experienced a change of approach from reductionism where genes and proteins are studied one at a time to a holistic approach where many (or ideally, all) genes are studied simultaneously. This global analysis of gene function is the basis of functional genomics. An example of this is high-throughput analysis by microarrays, where simple gene-by-gene hybridization techniques like Northern blots have been replaced with DNA chips where thousands of transcripts are probed enabling global expression profiling. Microarrays have been developed for a variety of insects, including *D. melanogaster* (White et al., 1999), *Anopheles gambiae* (David et al., 2005), *Aedes aegypti* (Sanders et al., 2003) and *Bombyx mori* (Noji et al., 2003) and have been used to study different biological processes. DNA chips are not yet available for our study insect *Rhodnius prolixus.*

Results presented in this dissertation, describe the use of functional genomics to identify and describe genes involved in the immune response of *R. prolixus*. As an ancient hemipteran insect of major medical importance, characterizing its immune response has two major points of relevance. Firstly, from an evolutionary perspective as insect immunology studies have focused on higher insect orders like Lepidoptera and Diptera.
Secondly, understanding the molecular interactions between this insect and *Trypanosoma cruzi*, the parasite it transmits (cause of Chagas disease) from an immune perspective. We hope that the discovery of the biological determinants

that permit such interactions will allow us to identify molecules or processes that limit, or ideally stop the transmission of human pathogens.

The basis for the identification of the genes described in this study relied on bioinformatics analysis of expressed sequence tags (ESTs) using similarity searches with sequences (protein and nucleic acids) deposited in remote databases. Because homologous genes with similar sequences tend to be derived from a common evolutionary ancestor, the proteins they encode for are related not only in structure, but also in function. Thus, the simplest way to assign a putative function to the newly discovered genes was to look for related sequences that have already been described, annotated and deposited in public databases like that of the National Center for Biotechnology Information (NCBI) (Jenuth, 2000). Sequence comparisons throughout these studies were carried out at the protein level because amino acid sequences vary less than nucleotide sequences, so protein sequence comparisons were assumed to be more sensitive.

Many insect immune genes have been identified and characterized either independently or on the basis of their relationships to known immune genes of the model organism, *Drosophila melanogaster*. Our work has also taken a similar approach. Because of this, the identification of genes (or similar genes) that have already been described was biased against completely novel genes with no similarities in the database. In consideration of this, a proteomic study of *R. prolixus* hemolymph was undertaken, where pure hemolymph from bacterial challenged or naïve insects was resolved by bi-dimensional gel electrophoresis, and presumed differential protein spots were excised and sequenced by Edman degradation.

Although bi-dimensional gel electrophoresis of proteins can be a powerful technique, it has its limitations. The difficulty in obtaining reproducible runs with similar spots profiles, a requirement for comparative analysis, made data analysis much more difficult and unreliable than expected. As a result, the

approach using Suppressive Subtractive Hybridization (SSH) was chosen over the proteomic approach for the remainder of the study.

Initially, we identified immune-related molecules from the fat body, and intestine of adult *R. prolixus*. Insects were challenged by introducing bacteria or *T. cruzi* into the hemocoel, and 24h after stimulation tissues were dissected to extract RNA. Based on SSH, we generated three subtracted libraries, sequenced the expressed sequence tags (ESTs) and compared them to DNA and protein sequences in the databases. The functional annotation revealed ESTs belonging to the three fundamental pillars of immunity: recognition, activation and effector mechanisms. Following EST sequencing we obtained the full length of several of these genes and continued with functional and transcriptional studies adressing the questions presented in the introduction.

A key feature of the immune response it its inducibility, in many organisms including insects, transcriptional regulation involves transcription factors of the Rel/NF-κB family. We cloned and characterized three members of this transcription factor family in this insect and studied their expression profiles and their DNA binding activity. Our results suggest that these transcription factors play a dual role in development and immunity. Their immune role is partially related to their ability to regulate the expression of several antimicrobial peptides as indicated by several putative NF-κB binding sites in the promoter regions of these genes known to be involved in the innate immune response of insects.

One of the antimicrobial peptides whose gene promoter contains this regulatory DNA sequence belongs to the defensin family of small cationic peptides. We isolated Defensin A from the hemolymph of *R. prolixus*, by HPLC it and sequenced it. Molecular cloning of the gene showed that the active peptide contains 43 amino acid residues and that there are three isoforms exist. Transcriptional analysis by Northern blot and Real-time quantitative PCR indicate that there is a very low baseline transcription of this peptide in naive insects, and

that transcription increases significantly in the fat body of immune activated insects while there is a delayed induction of transcription in the intestine. This suggests that the midgut/intestine of this species is active in the immune response against non self.

A second type of antimicrobial peptide under NF-κB regulation is lysozyme. We reported the characterization of two c-type lysozymes, RpLys-A and RpLys-B, isolated from the fat body and digestive tract of immune stimulated adult insects. A different transcriptional profile from that of Defensin emerges indicating that the temporal and spatial expression patterns of these two peptides are very different. RpLys-A is expressed predominantly in the midgut after ingestion of *T. cruzi* in a blood meal, or after injection of bacteria into the hemocoel. Phylogenetic analysis of these lysozymes suggests a different physiological role for each molecule. RpLys-A aligns best with molecules from other hemipterans whose major expression is found in the intestinal tract whereas RpLys-B aligns best with mosquito and tick molecules whose expression is found principally in hemocytes and fat body and whose role has been described as immune-related. These data suggest a differential compartmentalized role of two closely-related molecules; one for immunity in the hemocoel and the other for digestion in the midgut.

In the final data chapter of this dissertation, we characterize a novel antimicrobial peptide isolated from midgut tissues of *R. Prolixus.* Prolixin is a small, glycine rich peptide which belongs to the Diptericin family of antimicrobial peptides and it is the first member to be isolated outside the Diptera. This inducible peptide is produced by both fat body and midgut tissues of nymphs or adults in response to bacterial or *T. cruzi* infection of the hemolymph or the midgut. What is striking is that the *Prolixin* gene does not seem to be under the control of NF-κB binding sites like the other two antibacterial peptides we described. However, its promoter region contains several GATA sites, suggesting the presence of a different regulatory pathway. The recombinant protein has strong bactericidal

activity against *B. subtilis* but also bacteriostatic activity against *E. coli* in liquid culture.

In summary, our findings seem to indicate that invasion of the hemocoel by *T. cruzi* elicits an immune response in this reduviid insect. This response is mediated by components of the pro phenol oxidase cascade, and antimicrobial peptides; some of which are transcriptionally regulated by transcription factors of the Rel/NF-κB family. This supports previous findings where hemolymph agglutination activity for erythrocytes and parasites was evidenced against *T. cruzi* (Mello *et al.*, 1995) and *Trypanosoma rangeli* (Takle, 1988) but for which no molecular data was available. Furthermore, it also shows similarities with the defense mechanism observed in the Tse Tse fly *Glossina morsitans morsitans* against *Trypanosoma brucei* spp. In this fly, the pathogen-induced expression profile of the antimicrobial peptides attacin and cecropin is under the regulation of the Imd pathway transcriptional activator Relish (GmmRel) (Hu and Aksoy, 2006), which shows high amino acid identity and structural similarity to its *Drosophila melanogaster* orthologue.

Our results also indicate the activation of several pathogen specific genes in response to bacterial or parasitic invasion of the hemocoel. Some of these are homologous to genes described in other insect-parasite systems but the large number of unidentified genes suggests the possibility of unique immune genes in hemimetabolous insects. Although, we have focused on homologous genes here; future studies will characterize these novel immune related genes in terms of their biological activity and their effects on parasite development and transmission.

The much anticipated release of the annotated genome sequence of *R. prolixus* later this year should, coupled with biochemical analyses, greatly increase the rate of novel discoveries in this disease model insect.

Disease perspective

The most recent report on Chagas disease portraits an optimistic picture, where the number of annual deaths has decreased from over 45,000 to 12,500 from 1990 to 2006 (WHO-TDR, 2007). Similarly, the cases of human infection and new cases per year have decreased from 30 million and 700,000 to 15 million and 41,200 respectively. Most of the progress made was the result of efforts targeted to fight and eliminate vectorial transmission. However, these strategies do nothing for those individual already infected, thus the disease is far from being eradicated. Despite a renewed interest by non for profit organizations like the Drug for Neglected Diseases initiative (DNDi), composed of an international consortium of five public sector institutions and one international research organization, no new drug has yet been brought to the market.

Additionally, political instability and the economic disparities between most Latin-American countries and western European and North American countries has forced a large numbers of Latino nationals to emigrate legally or illegally to these countries, bringing with them *Trypanosoma cruzi*. A recent study by the Pan American Health Organization suggests that countries, that receive significant numbers of immigrants, like Canada, Spain, and the United States have thousands of individuals infected without knowing it (Schmunis, 2007). Estimates report that 1218 of the 131,135 Latin American immigrants received by Canada in 2001 may be infected with *T. cruzi*. This trend highlights the fact that Chagas disease is not a health concern only for Latin-American countries anymore but a global health issue. As such, primary health care workers in developed countries should be well informed of the disease (Just ask your general practitioner whether he/she knows about Chagas disease, the answer might surprise you!). Research initiatives at the basic and applied level should be encouraged to develop novel strategies to combat Chagas disease, and to prevent from spreading even more.

References

David, J.P., Strode, C., Vontas, J., Nikou, D., Vaughan, A., Pignatelli, P.M., Louis, C., Hemingway, J. & Ranson, H., 2005. The *Anopheles gambiae* detoxification chip: a highly specific microarray to study metabolic-based insecticide resistance in malaria vectors. *Proceedings of the National Academy of Sciences of the United States of America,* 102, 4080-4.

Hu, C. & Aksoy, S., 2006. Innate immune responses regulate trypanosome parasite infection of the tsetse fly *Glossina morsitans morsitans. Molecular microbiology,* 60, 1194-204.

Jenuth, J.P., 2000. The NCBI. Publicly available tools and resources on the Web. *Methods in molecular biology,* 132, 301-12.

Mello, C.B., Garcia, E.S., Ratcliffe, N.A. & Azambuja, P., 1995. *Trypanosoma cruzi* and *Trypanosoma rangeli:* interplay with hemolymph components of Rhodnius prolixus. *Journal of invertebrate pathology,* 65, 261-8.

Noji, T., Ote, M., Takeda, M., Mita, K., Shimada, T. & Kawasaki, H., 2003. Isolation and comparison of different ecdysone-responsive cuticle protein genes in wing discs of *Bombyx mori. Insect biochemistry and molecular biology,* 33, 671-9.

Sanders, H.R., Evans, A.M., Ross, L.S. & Gill, S.S., 2003. Blood meal induces global changes in midgut gene expression in the disease vector, *Aedes aegypti. Insect biochemistry and molecular biology,* 33, 1105-22.

Schmunis, G.A., 2007. Epidemiology of Chagas disease in non-endemic countries: the role of international migration. *Memorias do Instituto Oswaldo Cruz,* 102 Suppl 1, 75-85.

Takle, G.B., 1988. Studies on the cellular immune responses of insects toward the insect pathogen Trypanosoma rangeli. *Journal of invertebrate pathology,* 51, 64-72.

White, K.P., Rifkin, S.A., Hurban, P. & Hogness, D.S., 1999. Microarray analysis of Drosophila development during metamorphosis. *Science,* 286, 2179-84.

Who-Tdr, 2007. *Reporte sobre la enfermedad de Chagas.* Geneva.