The antibacterial immune response to Escherichia coli in the flea Xenopsylla cheopis

James David Driver

The University of Montana

Follow this and additional works at: http://scholarworks.umt.edu/etd

Recommended Citation


http://scholarworks.umt.edu/etd/9385

This Dissertation is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Theses, Dissertations, Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mail.lib.umt.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®
Permission is granted by the author to reproduce this material in its entirety, provided that this material is used for scholarly purposes and is properly cited in published works and reports.

**Please check "Yes" or "No" and provide signature**

Yes, I grant permission

No, I do not grant permission

Author's Signature: 

Date: 6/13/02

Any copying for commercial purposes or financial gain may be undertaken only with the author's explicit consent.
The Antibacterial Immune Response to *Escherichia coli* in the Flea

*Xenopsylla cheopis*

by

James David Driver, BS, MS

The University of Montana, Missoula, MT

Presented in partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

The University of Montana

May, 2002

Approved by

[Signature]

Chair

Dean, Graduate School

[Signature]

Date
The innate immune response in arthropods has previously been shown to be a complex process that is capable of mounting a multifaceted antimicrobial attack. The flea *X. cheopis* has been shown to produce an antibacterial response against the Gram-negative bacterium *Escherichia coli*. Inoculation or injection of *E. coli* into the flea hemocoel produces bactericidal peptides in the flea hemolymph. Following infection, analysis of the flea hemolymph by 1- and 2-dimensional SDS-PAGE indicated that some flea proteins are induced to higher levels of production when compared to naïve hemolymph. To further characterize this antibacterial response, cDNA libraries from uninfected and infected fleas were subtractively hybridized to selectively remove common sequences. The remaining cDNA sequences, enriched for genes induced by bacterial exposure, were sequenced and their similarities to other known arthropod immune genes were determined using a BLAST search. The results of the search identified two antibacterial genes in the flea with sequence similarity to other known antibacterial peptides. A flea defensin was identified by sequence similarity and and the presence of conserved cysteine residues that are a hallmark of the defensin family of antimicrobial peptides. A flea cecropin was also observed that aligned to other members of the cecropin family. Additionally, a cDNA sequence was obtained for a flea peptidoglycan recognition protein (PgRP). The PgRP’s from insects and mammals have been shown to induce the production of antibacterial peptides in those organisms following bacterial exposure. Also, a partial sequence was obtained for flea serine protease with similarity to mosquito and *Drosophila* proteases implicated in induction of the prophenyloxidase cascade. Other sequences obtained from the flea indicate that they are capable of producing additional antibacterial peptides with similarities to the attacins, coleoptericins, apidaecin, and diptericin B. These results indicate that *X. cheopis* is capable of producing a complex innate antibacterial immune response consisting of recognition factors and a variety of antibacterial peptides. The flea may produce other effector molecules that can act as regulators and/or inducers of the immune response.
# Table of Contents

Abstract ii  
Table of Contents iii  
List of Figures vi  
Acknowledgments ix  
Chapter 1. Introduction  
  1.1. Literature Review 9  
  1.2. Research Objectives 37  
2. Experimental Procedures 39  
  2.1. Flea Colony Care and Maintenance 39  
  2.2. Growth and Quantitation of Bacteria 39  
  2.3. Bacterial Challenge of *X. cheopis*. 40  
  2.4. LPS Inoculation of Fleas 41  
  2.5. Flea Protein Isolation 42  
  2.6. Solid Phase Extraction 43  
  2.7. Molecular Mass Centrifugation 44  
  2.8. Protein Concentration Assay 44  
  2.9. RP-HPLC 45  
  2.10. Ultracentrifugation 47  
  2.11. Antibacterial Bioassay 47  
  2.12. Determination of Specific Activity 48  
  2.13. Polyacrylamide Gel Electrophoresis 49  
  2.14. Densitometry 51  
  2.15. Protein Blotting and Immunoblotting 52  
  2.16. Amino Acid Sequencing 52
2.17. Extraction of Flea RNA 53
2.18. cDNA Synthesis 54
2.19. Subtractive Hybridization 55
2.20. TA Cloning of Subtracted Library 56
2.21. Sequencing of Flea Genes 57
2.22. Sequencing Data Analysis 57

3. Results 59
3.1. Antibacterial Bioassay Results from *X.cheopis* 59
3.2. SPE and Molecular Weight Cutoff Centrifugation of Flea Lysates 62
3.3. RP-HPLC of Whole Flea Lysates 68
3.4. SDS-PAGE of Fractionated Flea Lysates 69
3.5. Specific Activity Assays 76
3.6. Amino Acid Sequencing of RP-HPLC Fractions 83
3.7. Bioassay of Flea Hemolymph 83
3.8. Optimization of Antibacterial Response 83
3.9. Purification and 1D Gel Analysis of Hemolymph 86
3.10. 2D Gel Analysis of Hemolymph 89
3.11. Amino Acid Sequencing of 2D gel Spots 94
3.13. TA Cloning of Subtracted Library 98
3.14. DNA Sequencing and Analysis of cDNA 98
3.15. Identification of Flea Defensin 99
3.16. Identification of Flea Cecropin 107
3.17. Identification of a Flea PgRP 117
3.18. Identification of Flea Serine Protease 124
3.19. Partial Identification of Flea Sequences 132
# List of Figures

## Figure

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Infected <em>X. cheopis</em> Bioassay</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>Infected Flea Solid Phase Extraction Bioassay</td>
<td>64</td>
</tr>
<tr>
<td>3</td>
<td>Uninfected Flea Solid Phase Extraction Bioassay</td>
<td>66</td>
</tr>
<tr>
<td>4</td>
<td>SDS-PAGE of Flea Proteins</td>
<td>72</td>
</tr>
<tr>
<td>5</td>
<td>SDS-PAGE of HPLC-Separated Flea Proteins</td>
<td>74</td>
</tr>
<tr>
<td>6</td>
<td>Cecropin Bioassay</td>
<td>77</td>
</tr>
<tr>
<td>7</td>
<td>Cecropin Solid Phase Extraction Bioassay</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>SDS-PAGE of Induced Flea Hemolymph Proteins</td>
<td>87</td>
</tr>
<tr>
<td>9</td>
<td>2D Gel of Induced Flea Hemolymph Proteins</td>
<td>90</td>
</tr>
<tr>
<td>10</td>
<td>2D Gel Comparison of Induced vs Uninduced Flea Proteins</td>
<td>92</td>
</tr>
<tr>
<td>11</td>
<td>Agarose Gel of Subtracted Flea cDNA Library</td>
<td>96</td>
</tr>
<tr>
<td>12</td>
<td>Flea Defensin Protein Sequence</td>
<td>101</td>
</tr>
<tr>
<td>13</td>
<td>Defensin Sequence Alignments</td>
<td>103</td>
</tr>
<tr>
<td>14</td>
<td>Defensin Sequence Dendrogram</td>
<td>105</td>
</tr>
<tr>
<td>15</td>
<td>Flea Cecropin Protein Sequence</td>
<td>109</td>
</tr>
<tr>
<td>16</td>
<td>Cecropin Sequence Alignments</td>
<td>111</td>
</tr>
<tr>
<td>17</td>
<td>Cecropin Sequence Dendrogram</td>
<td>113</td>
</tr>
<tr>
<td>18</td>
<td>Secondary Structure Analysis of Cecropins</td>
<td>115</td>
</tr>
<tr>
<td>19</td>
<td>Flea PgRP Protein Sequence</td>
<td>118</td>
</tr>
<tr>
<td>20</td>
<td>PgRP Sequence Alignments</td>
<td>120</td>
</tr>
<tr>
<td>21</td>
<td>PgRP Sequence Dendrogram</td>
<td>122</td>
</tr>
<tr>
<td>22</td>
<td>Flea Serine Protease-like Protein Sequence</td>
<td>126</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Acknowledgements

First I must thank my advisor Bill Granath for not only his support but also for the freedom he gave me to pursue this project. I must also thank the other members of my committee, Mike Minnick and Scott Samuels, for always having an open door, good advice and the use of their facilities and materials whenever needed. Also, Tom Schwan and Joe Hinnebusch for the opportunity to work at RML. Additionally, the Division of Biological Sciences faculty and staff have always been very supportive of my educational goals. Finally, I would like to dedicate this thesis to my wife Jennifer Stephens and to my children Sydney Hale and James McKinley Driver
Chapter 1

Introduction

The study of the immune system in animals has tended to focus on the roles that the humoral and/or cellular components play in protecting an organism against infection. Although the adaptive immune system has an advantage in its ability to develop a specific response against a pathogen, it is clear that it is not triggered quickly enough to immediately protect the host against the initial exposure. More recently there has been a greater appreciation of the innate immune system and its ability to respond quickly to microbial invasion. Also, research undertaken since the early 1980’s has uncovered a variety of antimicrobial peptides common to the innate immune system that has been described in many animal species that have a central role in the immune response against microbes (Hultmark et al., 1980). The innate immune system, whose components have been described in many organisms, is a complex system with a variety of effectors. It has the ability to attack invading pathogens based on recognition of the common motifs or pathogen-associated molecular patterns (PAMP’s) found on bacterial and fungal surfaces. Recognition molecules, also called pattern recognition receptors (PRR’s) bind to specific PAMPs and some examples are an LPS-binding protein (LBP) and peptidoglycan recognition protein (PgRP) that
have been described in several *Drosophila* spp. (Kim et al., 2000; Werner et al., 2000). There are also similar signal transduction pathways that activate the transcription of genes encoding host defense molecules with the best described being the Toll and Toll-like receptor pathways found in *Drosophila* spp. and humans (Imler and Hoffmann, 2001). Finally, there are groups of antibacterial peptides and proteins capable of killing or inhibiting bacterial, fungal, or parasitic invaders.

Antimicrobial proteins and peptides have been found in plants and animals including arthropods, molluscs, crustaceans, amphibians, birds, fish, and mammals. The antibacterial peptides contain between 12 and 50 amino acids and most have a folded size that is approximately the thickness of the bacterial membrane (Huang, 2000). Most of the peptides and proteins isolated (752 as of January, 2002) are cationic in nature with at least two excess positive charges from extra lysine or arginine residues (Boman, 2000). They can be grouped into at least four structural classes: linear peptides forming amphipathic α-helices (e.g. cecropins), β-sheet structures linked by disulfide bonds (e.g. defensins), peptides with an irregular amino acid composition with an over-representation of proline and/or glycine (e.g. indolicidin, attacin), and peptides containing loop structures (e.g. bactenicin) (Epand and Vogel, 1999). The following chart illustrates representative
groups of antibacterial peptides, the diversity of animals that produce them, the sizes and targets of these peptides, and the year of their discovery.

<table>
<thead>
<tr>
<th>Representative peptides</th>
<th>Origin</th>
<th>Targets</th>
<th>Size (kDa)</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>α-helical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cecropin A</td>
<td>Silk moth</td>
<td>G+ and G-</td>
<td>4.2</td>
<td>1982</td>
</tr>
<tr>
<td>Magainin 2</td>
<td>Frog</td>
<td>G+ and G-</td>
<td>2.4</td>
<td>1987</td>
</tr>
<tr>
<td>Pexiganan</td>
<td>Synthetic</td>
<td>G+ and G-</td>
<td>3.0</td>
<td>1999</td>
</tr>
<tr>
<td>Dermaseptin 1</td>
<td>Frog</td>
<td>G+ and G-</td>
<td>4.7</td>
<td>1991</td>
</tr>
<tr>
<td>LL-37</td>
<td>Human</td>
<td>G+ and G-</td>
<td>4.4</td>
<td>1995</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>One disulphide bond</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thanatin</td>
<td>Insect</td>
<td>G+ and G-, fungi</td>
<td>2.4</td>
<td>1996</td>
</tr>
<tr>
<td>Brevinin 1</td>
<td>Rana frogs</td>
<td>G+ and G-</td>
<td>2.7</td>
<td>1994</td>
</tr>
<tr>
<td>Ranalexin</td>
<td>Rana frogs</td>
<td>G+ and G-, fungi</td>
<td>2.1</td>
<td>1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Two disulphide bonds</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tachyplesin</td>
<td>Crab</td>
<td>G+ and G-, fungi</td>
<td>2.3</td>
<td>1990</td>
</tr>
<tr>
<td>Androctonin</td>
<td>Scorpion</td>
<td>G+ and G-, fungi</td>
<td>3.1</td>
<td>1996</td>
</tr>
<tr>
<td>Protegrin 1</td>
<td>Pig</td>
<td>G+ and G-, fungi</td>
<td>2.2</td>
<td>1993</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Three disulphide bonds</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-defensin</td>
<td>Human</td>
<td>G+, G-, fungi, &amp; virus</td>
<td>3.4</td>
<td>1988</td>
</tr>
<tr>
<td>ß-defensin</td>
<td>Cow</td>
<td>G-, fungi</td>
<td>4.3</td>
<td>1993</td>
</tr>
<tr>
<td>ù-defensin</td>
<td>Monkey</td>
<td>G+ and G-</td>
<td>3.0</td>
<td>1999</td>
</tr>
<tr>
<td>Defensin</td>
<td>Insect</td>
<td>G+</td>
<td>4.4</td>
<td>1993</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Linear, not α-helical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attacin</td>
<td>Insect</td>
<td>G+ and G-</td>
<td>20.0</td>
<td>1984</td>
</tr>
<tr>
<td>PR-39</td>
<td>Pig</td>
<td>G+, G-, LPS Binding</td>
<td>4.5</td>
<td>1995</td>
</tr>
<tr>
<td>Indolicidin</td>
<td>Cow</td>
<td>G+, G- and fungi</td>
<td>1.9</td>
<td>1992</td>
</tr>
<tr>
<td>Apidaecin</td>
<td>Honeybee</td>
<td>G-</td>
<td>32.7</td>
<td>1993</td>
</tr>
<tr>
<td>Sarcotoxin</td>
<td>Insect</td>
<td>G+, and G-</td>
<td>4.2</td>
<td>1986</td>
</tr>
</tbody>
</table>

Table adapted from Zasloff, M., 2002, Hoffman et al, 1994, and the Antimicrobial Sequences Database, University of Trieste, Italy.
Antibacterial peptides were first discovered by two independent lines of research: (1) studies into the ability of cell-free hemolymph from immunized insects to kill bacteria, and (2) studies into the mechanisms used by mammalian phagocytic cells to kill bacteria in vacuoles. Research into the insect immune system yielded the first known antibacterial peptide that was isolated from the cell-free hemolymph of the moth *Hyalophora cecropia*, (Hultmark et al., 1980). This antibacterial, named cecropin, is a cationic peptide with bacteriocidal activity against both Gram positive and Gram negative bacteria. It is the first in a class of small linear peptides (approximately 4000 daltons) with amphipathic α-helical structures.

Concurrently, researchers isolated the first of a group of defensins from rabbit macrophages and subsequently in rabbit, rat, and human neutrophils (Selsted et al., 1984). Defensins are largest group of antibacterial peptides identified with approximately 200 known sequences with 32 coming from insects (cataloged by the Antimicrobial Sequences Database at the University of Trieste, Italy). These defensins were the first members of the group of peptides with β-sheet structures linked by disulfide bonds to be characterized, and this group contains not only those found in intracellular granules but also secreted defensins found in vertebrates and in arthropods.
Arthropods are an extremely diverse group of organisms. More than one million species have been identified and they inhabit a wide range of ecological niches. Their success can be partially attributed to their ability to adapt to and resist the variety of infectious organisms and parasites that also inhabit these niches. To combat these pathogens, arthropods have evolved complex innate immune systems capable of recognizing and reacting to foreign invaders. Arthropod immunity consists of innate components such as a cuticular exoskeleton as well as cellular and humoral constituents. Cellular immunity is based upon the various types of hemocytes found circulating in the hemolymph. Bacteria entering the hemocoel are first recognized and removed by hemocytes either by phagocytosis or nodule formation. Bacterial invasion has also been shown to induce the production of a variety of opsonins, agglutinins and antibacterial proteins in a number of insect species. The antibacterial peptides are evolutionarily ancient components and their widespread distribution throughout the plant and animal kingdoms suggests that they played a key role in the success of many species in resisting microbes. Antimicrobial peptides have remained effective even though the dogma suggests that bacteria, fungi, and viruses are capable of developing resistance to any substance (Zasloff, 2002). These humoral factors, acting in concert with the hemocytes or alone, are the organism's
internal defense against bacteria, fungus or parasites thereby maintaining a sterile environment inside the arthropod (Gillespie et al., 1997). The best characterized arthropod innate response is the antibacterial immune response of *Drosophila* spp. which includes cecropins, defensins, antifungal peptides, peptidoglycan recognition proteins, LPS-binding proteins, protease cascades, and transcription signaling cascades (Irving et al., 2001). Based on this evidence it is likely that many arthropods have a multifaceted immune response to microorganisms similar to that found in the fruit fly.

The oriental rat flea, *Xenopsylla cheopis*, is considered to be the most important vector for the plague bacillus, *Yersinia pestis*, as well as for murine typhus (Roberts and Janovy, 2000). This flea is found associated with Rattus species worldwide except in cold climates. In the United States it ranges as far north as New Hampshire, Minnesota, and Washington. *X. cheopis* ingests *Y. pestis* when it takes a blood meal from an infected mammal. The bacteria can colonize the flea midgut and are capable of multiplying in the midgut to a high enough level to block the passage of the blood through the esophageal sphincter or proventriculus. This block causes the flea to attempt to feed constantly since the new blood cannot enter the midgut through the bacterial blockage. The fresh blood comes into contact with the blockage and can be contaminated with bacteria and regurgitated.
into the bite wound thereby transmitting the agent. *X. cheopis* is a good
vector for the bacterium because it can easily become blocked by *Y. pestis*,
feeds on both rats and humans, and is found near human habitations.

Evidence from our laboratory has shown that *X. cheopis* produces an
antibacterial protein following bacterial challenge to its hemocoel. This
antibacterial compound can be found in the hemolymph and is likely
produced in the fat body or by hemocytes as has been found in other insects
(Bulet et al., 1999). This antibacterial compound is likely to have no effect
on the transmission of *Y. pestis* since the bacteria are confined to the midgut
and have no contact with the hemolymph. However, there is evidence that
different species of *Yersinia* may have different levels of susceptibility to
cationic antimicrobial peptides (Bengoechea et al., 1998). *Y. pestis* may
contact a flea antimicrobial and may be immune to its effects. Researchers
have isolated antibacterials from the midgut tissues in other organisms
(Jones and Bevins, 1992). It is possible that the flea midgut may produce a
type of antibacterial that interacts with *Y. pestis* and that the bacterium may
have developed resistance in order to successfully colonize the midgut.

Insect antibacterial peptides are being examined as additions to
commonly used antibiotics (Hancock and Lehrer, 1998). The widespread use
of antibiotics has led to the development of antibiotic resistance in a variety
of human pathogens. Much effort is going into the search for new antibiotics with novel modes of action. Cationic antimicrobial peptides have several attributes that make them attractive candidates as new antibiotics. They have a broad spectrum of activity, can kill bacteria rapidly, and appear to use mechanisms that circumvent common antibiotic resistance mutations. They also work synergistically with common antibiotics, appear to neutralize endotoxins, and have activity in animal models (Hancock, 2000). There are drawbacks to the use of antimicrobial peptides in humans such as toxicity and the size difference between the larger peptides and traditional antibiotics. However, there are a number of antimicrobial peptides of animal origin undergoing trials as of 2002. These new compounds include those for topical, oral, and systemic usage (Zasloff, 2002). The induction of antimicrobial peptides in insects has been shown to be both rapid and specific. For example, *Drosophila* spp. can produce more than ten different antimicrobials following challenge with various microorganisms. Some are effective against a broad range of targets while others are specific for Gram-positive and/or Gram-negative bacterial, or fungal targets (LeMaitre, 1997). Characterization of the *X. cheopis* antibacterial compound(s) will help to elucidate the innate immune response in the flea and could possibly provide
other examples of novel bioactive peptides suitable for human health applications.

1.1. Literature Review

Research into the immune response in animals has most often focused on the adaptive response. The ability of vertebrates to identify a microbial invader, isolate specific antigens from the invader, and produce antibodies against these antigens that will either kill the invader or facilitate the activities of cellular immune components has long been the hallmark of an effective immune system. Due to its specificity it has been considered more advanced than the innate or non-adaptive response. However, recent advances in the understanding of the components and mechanisms of induction for the innate response have led to a greater appreciation for its abilities to respond quickly and to a variety of microbial pathogens even though it lacks the antigenic specificity of the adaptive immune response. Exposure to a novel pathogen would require weeks for the adaptive response to produce an antibody. Bacteria have an extremely rapid growth rate when compared to even the secondary adaptive response time of three days. It is becoming more evident that the innate response is essential for preventing the onset of infection (Janeway, 1998). The inducible components of the
innate response are different from the adaptive response in that they are relatively non-specific. These components recognize conserved molecular patterns from microorganisms such as the LPS of Gram negative bacteria, the glycolipids of mycobacteria, the lipoteichoic acids and peptidoglycan of Gram-positive bacteria, the β-1,3-glucans of fungi, and even the double-stranded RNA of viruses. Also, they are rapidly induced in minutes or hours following contact unlike the days necessary to mount an antibody response. Although the commonly recognized components of the innate response are the phagocytic cells such as neutrophils and macrophages as well as serum proteins such as complement, there is growing evidence that the group of induced cationic antimicrobial peptides play an important role. Not only do they have broad antimicrobial activities they also have been implicated as inflammatory response effector molecules (Hancock and Scott, 2000). Additionally, there is evidence that the innate response uses recognition and signaling molecules to induce the production of these antibacterials. In *Drosophila* spp. the innate immune response is activated by membrane receptors of the Toll family and by transcription factors related to NF-κB found in mammals (Werner et al., 2000). Toll and other membrane receptors can mediate immune responses in *Drosophila* spp. and mammals through interactions with recognition molecules such as the Gram-negative binding
protein and a peptidoglycan recognition protein found in insects and their homologs found in mammals (Kim et al., 2000). This evidence indicates that the innate system is capable of pathogen recognition, self vs non-self discrimination, signaling and induction of effector molecules, and killing of the microbial invaders by the effector molecules.

The history of research on the innate immune system of arthropods goes back to the beginnings of microbiology. By the 1920’s it was known that injection of bacterial cultures into insects would induce the production of bacteriolytic substances in the hemocytes and the hemolymph. Due to the technical difficulties in isolating, purifying and identifying these substances it was not until the mid-1970’s that the first inducible insect antibacterial protein was identified, when lysozyme was isolated from the wax moth Galleria mellonella and the silkworm Bombyx mori (Powning and Davidson, 1973). By 1980, Hans Boman at the University of Stockholm, Sweden had isolated and determined the amino acid sequences of two antibacterial peptides from the hemolymph of the moth H. cecropia following bacterial challenge in the hemocoel. These peptides, named cecropin A and B were small, 35 to 39 amino acids long, and specifically bactericidal for only certain gram-positive and gram-negative bacteria, especially Escherichia coli (Boman and Steiner, 1981).
The initial isolation of antibacterials was a painstaking process involving the induction of an immune response by bacterial inoculation followed by collection of the insect’s hemolymph. Once enough hemolymph was collected a series of purification steps involving both solid phase extraction (SPE) and high pressure liquid chromatography (HPLC) were used to isolate single proteins/peptides for automated sequencing (Steiner et al., 1981). The initial attempts to identify the active antibacterials induced following bacterial challenge indicated that there were a number of inducible proteins but that only three had antibacterial activity, lysozyme and cecropin A and B (Hultmark et al., 1980). When tested against Gram-negative bacteria the purified cecropins were bacteriolytic against *E. coli*, *Enterobacter cloacae*, and *Pseudomonas aeruginosa*, while some strains of *Serratia marcescens* were resistant. The cecropins were also variably effective against Gram-positive bacteria, with *Micrococcus luteus* and *Bacillus megaterium* being highly susceptible, *B. subtilis* being somewhat susceptible, and *B. cereus* and *B. thuringiensis* fully resistant (Hultmark et al., 1980).

Following their initial discovery of inducible cationic antibacterial peptides in *H. cecropia*, Boman’s group investigated another member of the Lepidoptera, *G. mellonella*, to determine if this type of humoral immunity
was widespread within the Order. They observed two inducible peptides following bacterial challenge and showed that they were bacteriolytic. However, due to the difficulty of obtaining sufficient amounts of purified protein for sequencing they did not further characterize the peptides (Boman, 1981). Using similar techniques an additional antibacterial peptide was isolated from *H. cecropia* and characterized by sequence and bacteriolytic activity. Cecropin D, a 36 amino acid antibacterial peptide had a narrow range of bacteriolytic activity against only *E. coli* and two other bacterial species (Hultmark et al., 1982). Attempts to discover other antibacterial peptides lead to the discovery in the mid-1980’s of the defensin family of antibacterial peptides in mammalian cells (Lehrer et al., 1991). Defensins were subsequently discovered and further characterized in a number of insect Orders including Dipterans and Coleopterans (Hoffmann and Hetru, 1992).

By 1990 research into the induction of the innate immune response in *H. cecropia* had yielded the previously mentioned lysozyme and cecropin A, B, and D, as well as a new group of antibacterials, the attacins. These antibacterials were large, approximately 20 kD proteins, and their activity was limited to growing bacteria (Hultmark et al., 1983). Initially it was observed that there were six forms of attacins (attacins A through F) found
in *H. cecropia*, after being separated by their isoelectric points. N-terminal sequencing indicated that the attacins could be divided into a basic and acidic forms with three basic forms having similar sequences and the two acidic forms having identical sequences that were slightly different from the basic. Sequencing of cDNA from the two main forms, the acidic and the basic, showed that the mature proteins are very similar with 79% identity of amino acids while the preprotein and the proprotein regions of the complete protein share only 43% identity of their amino acid sequence. It was determined that processing of the pro-protein to the mature form is what accounted for all six forms of attacin observed (Engstrom, 1984; Fischer and Spiess, 1987). Attacins were shown to have a limited spectrum of activity with good activity only against *E. coli* and two other bacteria found in the gut of larval Chinese oak silk moth larvae (Hultmark et al., 1983). For the two main types of attacins, the acidic and the basic forms, their antibacterial activity was shown to be against the outer membrane of *E. coli* and it was also shown to act in concert with lysozyme and cecropin which increased their effectiveness (Engstrom et al., 1984). These results indicated that not only were insects capable of producing a variety of antimicrobial agents but that they appeared to have some specificity against bacteria.
During the 1990's and into the beginning of the 21st century, research into the innate immune system and the isolation of cationic antibacterial proteins/peptides has continued. One reason for the continued interest in these antibacterials is the increase in bacterial antibiotic resistance and the need for new therapeutic agents. Among the desirable features they exhibit are their broad spectrum of activity and their ability to rapidly kill microorganisms. Also they are unaffected by the commonly induced antibiotic resistance mutations, do not easily select for resistance in bacteria due to their targets, show synergy with some common antibiotics, and are active in animal models (Hancock and Lehrer, 1998). The Antimicrobial Sequences Database lists peptides and proteins that have been isolated from insects, plants, amphibians, mammals, spiders/scorpions/ticks, fish, crabs, molds, birds, nematodes, tunicates, and amoeba. The antimicrobials are broken down into three main groups of proteins, s-s bridged peptides and linear peptides. These groups are further broken down to subgroups of s-s bridged peptides with a single bridge (e.g. dodecapeptide), two s-s bridges (e.g. protegrins), and three or more s-s bridges (e.g. defensins). In addition, the linear peptides contain the alpha-helical peptides (e.g. cecropins) and the peptides/proteins with an over-representation of certain amino acids such as proline-rich (e.g., bactenecins), glycine-rich (e.g., attacins), histidine-rich...
(e.g., histatins), tryptophan-rich (e.g., indolicins), and glutamic acid (e.g., enkelytin). The antimicrobials have been found to have bacteriolytic activity against Gram-positive and Gram-negative bacteria and fungus, and can also be hemolytic.

Cecropins from various other insects had also been isolated by 1990 with amino acid sequences reported from the flesh fly *Sarcophaga peregrina* (Matsumoto et al., 1986), the silkworm *B. mori* (Morishima et al., 1990), and the tobacco hornworm *Manduca sexta* (Dickinson et al., 1988), and cDNA sequences from *Drosophila melanogaster* (Kylsten et al., 1990), as well as a cecropin isolated from porcine intestine (Lee et al., 1989). The structure of these earliest isolated cecropins was similar in each having a strongly basic N-terminal region and a long hydrophobic stretch in the C-terminus. The cecropins are synthesized as preproproteins of 62-64 residues with a 24-26 amino acid prepro region at the N-terminus. The signal peptide encompasses the first 22 amino acids in *C. hyalophora*. The first 12-13 residues are conserved in the three forms of cecropin produced by the moth and these are conserved to a lesser degree in four other insects, *D. melanogaster, S. peregrina, M. sexta*, and *B. mori*. The pre region following the conserved section is quite variable among insect groups, however it is conserved for a number of frog skin peptides (Bevins and Zasloff, 1990) and
for the defensins (Jones and Bevins, 1992). Following the signal sequence is a pro-sequence composed of one or two dipeptides that must be cleaved to activate the peptide. It was speculated that this sequence would be cleaved by a dipeptidyl aminopeptidase similar to the one found in the bee venom toxin melittin (Kreil et al., 1980). This enzyme was subsequently isolated from C. hyalophora hemolymph and successfully tested for in vitro processing of cecropin A and B (Boman et al., 1989).

Subsequent research has isolated 35 individual cecropins from the Lepidoptera and Diptera, a Tunicate spp. and one from mammals, a porcine cecropin. More recent additions have include representatives from mosquitos. An amino acid sequence comparison of the mature cecropin A from Aedes aegypti to other Dipteran cecropins showed similarities of 82.4% to Aedes albopictus, 88.2% to Anopheles gambiae, and 35.3% to D. melanogaster (Lowenberger et al., 1999). A phylogenetic analysis of the insect cecropin family at the amino acid level showed that mosquito cecropins form one branch while other Lepidopteran and other Dipteran cecropins form a second branch. Additionally, the Dipteran cecropins, except for those from mosquito, form a single root while the Lepidopteran cecropins separate into two root groups (Lowenberger et al., 1999).
The structure of insect cecropin consists of two amphipathic alpha helices connected by a short hinge region with the N-terminal helix having more amphipathic character and the C-terminal region having more hydrophobic character (Holak et al., 1988). Their mechanism of killing against Gram-negative and Gram-positive bacteria is still a matter of debate. It has been demonstrated that cecropin is bactericidal to a variety of bacteria and, in some cases, at lower molar concentrations than commonly used antibiotics when tested in an inhibition zone bioassay. Cecropins will lyse bacteria but not eukaryotic cells, possibly due to the lack of cholesterol in bacterial membranes (Steiner et al., 1981). It has been postulated that antibacterial peptides are attracted to the negatively charged bacterial surface due to their cationic nature. Once bound to the surface they are able to insert themselves into the lipid bilayer due to their hydrophobic properties where they disrupt the bacterial membrane leading to cell death. Aminoarabinose modification of lipid A leads to the alteration of bacterial surface net charge and this has been implicated in resistance to cationic antimicrobial peptides (Gunn et al., 1998). This observation is consistent with the hypothesis that surface charge plays an important role in antibacterial function. Additionally, increased acylation of lipid A has been observed in Salmonella spp. following induction of the PhoP-PhoQ signal transduction system (Guo...
et al., 1998). This is believed to alter the fluidity of the outer membrane by increasing the hydrophobic interactions between the increased number of lipid A acyl tails. This would retard the insertion of the hydrophobic moiety of an antimicrobial peptide, altering or abolishing their ability to disrupt the outer membrane.

Although the evidence is strong to support non-specific charge interactions and hydrophobic insertion of cationic antimicrobials into bacterial membranes the actual method of bacterial killing is not as clear. There is evidence that once the peptides have attached and inserted themselves in the membrane they permeabilize the lipid bilayer through three possible mechanisms. The peptides may come together and reorient themselves into a barrel-stave pore (Christensen et al., 1988), or the aggregation of the peptides may cause the membrane to thin and form toroidal pores (Matsuzaki et al., 1996). In addition, the peptide molecules could cause disaggregation of the membrane in a detergent-like manner once they reach a critical concentration (Gazit et al., 1995). Permeabilization of the membrane can then lead to membrane depolarization, loss of essential metabolites and translocation of the peptides to the cytoplasm. Cationic peptides in the cytoplasm may interfere with cellular mechanisms by binding to proteins or DNA and disrupting their function leading to cell death.
However, it is not clear at this time whether membrane permeabilization or secondary events are more important in microbial killing or if they act synergistically. Also different antibacterial cationic peptides may act in different manners or against different targets. Many studies on antimicrobial peptide have been performed on synthetic lipid vesicles rather than using bacteria. Research performed using cecropin A and *E. coli* indicated that permeabilization of the bacterial membrane lagged behind bactericidal activity as the peptide concentration increased (Silvestro et al., 2000). This may indicate that there are cytoplasmic targets for the peptides and that disruption of these targets leads to bacterial death. This and other research has shown that membrane depolarization also lags behind bactericidal activity thereby lowering the number of possible bactericidal mechanisms (Zhang et al., 2000). It is likely that cationic peptides such as cecropin either form stable or transient pores through the bacterial membrane leading to loss of essential metabolites and/or the translocation of the peptides to the cytoplasm causing disruption of cell functions (Tossi et al., 2000).

The Antimicrobial Sequences Database lists 200 individual defensin sequences isolated from a broad variety of animals and they constitute the largest group of antibacterial peptides isolated to date. The first pure peptides were originally isolated from rabbit lung macrophages and they
later yielded the first sequences for defensins (Selsted et al., 1983).

Defensins are found in phagocytes from a variety of animals including humans and have also been isolated from Paneth cells in mouse and human small intestine (Boman, 1995). Defensins are abundant in human neutrophils where they constitute up to 15% of the total protein in these cells (Hancock and Diamond, 2000). Mammalian defensins have been observed and isolated from skin, lungs, trachea, kidney, salivary glands, small intestine, and liver. They are constitutively expressed in healthy epithelial cells and are also induced by pro-inflammatory cytokines, LPS, or bacteria (Schroder and Harder, 1999). In vivo studies indicate that defensin levels dramatically increase following infection and are protective in animals (Hancock and Scott, 2000). One study conducted with mice altered the gene for the activation of all mouse defensin types from the propeptide to the mature active form. Inactivation of this single gene leads to a tenfold increase in susceptibility in these mice to infection by orally introduced virulent bacteria (Wilson et al., 1999). In neutrophils defensins appear to be involved in the non-oxidative killing of bacteria since neutrophils lacking an oxidative response isolated from chronic granulomatous disease patients are still able to kill most bacteria (Scott et al., 1999). In addition to their role in bacterial killing there is evidence that defensins may also play a role in induction of
wound repair proteoglycans (Chan and Gallo, 1998), stimulation of nonopsonic phagocytosis (Sawyer et al., 1988), and chemoattraction of IL-8-stimulated neutrophils (Chertov et al., 1996).

The insect defensins were originally isolated in 1988 from the supernatant of a cell line that originated in *S. peregrina* (Matsuyama and Natori, 1988), and in the hemolymph of the flesh fly *Phormia terranovae* (Lambert et al., 1989). Insect defensins, like the cecropins, are induced and secreted into the hemolymph as a response to bacterial challenge, and are not contained in phagocytes as found with mammalian defensins. To date arthropod defensins have been discovered in seven insect Orders as well as in scorpions, and insect defensin-like molecules have been found in mussels and in the nematode, *Ascaris suum*.

Defensins are cationic peptides ranging in length from 29 to 46 amino acids and are characterized by six cysteine residues that form three intramolecular bridges. Mammalian defensins are variably cationic, relatively arginine-rich non-glycosylated peptides and the molecule consists of a three-stranded beta sheet stabilized by three disulfide bonds (Pardi et al., 1992). The classical or alpha defensins have cysteine linkages at the 1-6, 2-4, and 3-5 positions while the beta defensins have linkages at 1-5, 2-4, and 3-6, and the insect defensins are linked 1-4, 2-5, and 3-6 (Martin et al.,
Mammalian alpha defensins form a triple stranded beta sheet structure connected by a loop with a beta hairpin hydrophobic finger. The structure of insect defensin contains an amino terminal loop, an amphipathic alpha helix and a carboxyl-terminal antiparallel beta sheet. Details of the structure of *P. terranovae* defensin, using NMR studies, indicated that the amino terminal loop is linked to the first strand of the beta sheet by a disulfide bridge, while the alpha helix is stabilized by the two other disulfide bridges linked to the second strand of the beta sheet (Bonmatin et al., 1992). It was initially thought that insect and mammalian defensins were similar in structure based on the sequence similarity between them. However, based on the structural data and the different linkages of the three intramolecular disulfide bonds it is apparent that insect and mammalian defensins are markedly different.

Alpha defensins are active against Gram positive and Gram negative bacteria, fungi, and some enveloped viruses, and a mouse intestinal defensin is active against the intestinal parasite *Giardia lamblia* (Lehrer et al., 1993; Aley et al., 1994). The beta-defensins are active against Gram-positive and Gram-negative bacteria and fungi (Selsted et al., 1993). In general the bactericidal activity of defensins is increased in proportion to their net positive charge. However their bactericidal activity against Gram negative
bacteria is diminished in the presence of salts and divalent cations and this may contribute to the prevalence of chronic bacterial infections in the lungs of cystic fibrosis patients (Goldman et al., 1997). Insect defensins are primarily active against Gram-positive bacteria and only a few Gram-negative bacteria, fungi, and yeasts in contrast to the activities of the mammalian defensins against Gram-negative bacterial, fungal, and viral targets (Hoffmann and Hetru, 1992). Insect defensins work very rapidly such that a 1 minute exposure of *M. luteus* to 0.5 μM of defensin A will kill growing or resting bacteria (Bulet et al., 1991). This bactericidal concentration can be compared to the endogenous defensin level of 1 to 5 μM found in *P. terranovae* 24 hours after inoculation and indicates that the induction of this peptide in vivo is to levels sufficient for bacterial killing.

Based on their cationic and amphipathic nature it has been proposed that defensins act in a manner similar to the cecropins. Their overall positive charge likely attracts them to the negatively charged surface of bacteria. Gram-positive bacteria have peptidoglycan as part of the outer cell wall which is rich in teichoic acid. Studies performed on *Staphylococcus aureus* having mutations that increased the negative surface charge showed increased sensitivity to cationic antimicrobial peptides including defensin (Peschel et al., 1999). Once attached to the outer membrane defensins can

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
insert themselves in the lipid bilayer and form channels. This has been demonstrated in artificial membranes, where both mammalian and insect defensins formed voltage-dependent channels (Kagan et al., 1990). Insect defensin was also observed forming voltage dependent channels in intact *M. luteus* indicating a correlation between the activity seen in artificial membranes and a live bacterium (Cociancich et al., 1993). This study showed that the insect defensin disrupted the permeability barrier of the cytoplasmic membrane resulting in the loss of cytoplasmic potassium, a partial depolarization of the inner membrane, a decrease in ATP in the cytoplasm, and an inhibition of respiration. Due to the rapidity of these events in bacteria as seen with defensins, cecropins, and maginins it is difficult to determine if there are other targets for these peptides. However, pretreatment of *S. aureus* with novobiocin, an inhibitor of bacterial DNA gyrase subunit B, or with 50S ribosomal subunit protein synthesis inhibitors, such as daltopristin, blocked bacterial killing by the mammalian defensin HNP-1 (Xiong, 1999). Studies using all D-analogs of antimicrobial peptides showed the same antimicrobial activities as their natural L-analogs and this indicates that the mechanisms are not receptor-mediated or chiral in nature but are aimed at the general structure of the bacterial lipid bilayer. All-L and all-D enantiomers of antimicrobial peptides may not all be of equal activity.
and the activities may be species-dependent and for some bacterial species there may be a "receptor" molecule (Epand and Vogel, 1999). Thus, there may be an intracellular target for these peptides once they have traversed the bacterial membrane. Defensins may also act as signaling molecules for the recruitment of macrophages, granulocytes and lymphocytes to the site of an infection (Welling et al., 1998). Defensins may also act as a link between the innate and adaptive immune systems by enhancing systemic IgG involving Th-1 and Th-2 cytokines (Lillard et al., 1999).

Drosophila spp. use three broad mechanisms to maintain resistance to microbial infections: (a) phagocytosis of the invading microorganisms by circulating hemocytes, (b) a proteolytic cascade that causes localized hemolymph clotting, melanization and opsonization, and (c) transient, inducible synthesis of antimicrobial peptides. There is now a considerable body of information on the structure, function, regulation, and expression of antimicrobial peptides in a variety of insects and especially in Drosophila ispp. Drosophila spp. are capable of the induced production of seven antimicrobial peptides (andropin, cecropin, defensin, diptericin, drosomycin, drosocin, and metchnikowin) and one antimicrobial protein (attacin). The peptide drosomycin acts potently against fungi but not against bacteria (Fehlbaum et al., 1994) while the other peptides primarily act against
bacteria although they may also target Gram-positive and/or Gram-negative bacteria as well as some fungi. The antimicrobial peptides are mostly produced in the insect’s fat body, the functional equivalent of the mammalian liver, and are secreted into the hemolymph as a systemic response. Additionally, *Drosophila* spp. have been shown to produce antimicrobial peptides as a local response in barrier epithelial cells (Ferrandon et al., 1998).

*Drosophila* spp. are able to recognize different classes of invading organisms and will respond by producing peptides that will target that specific microbe. An example of that response is found in flies that are naturally infected with entomopathogenic fungi where they will selectively activate a pathway to produce antifungal peptides. This study also indicated that infection with bacteria is the best inducer for the antibacterial peptides while infection with fungi was the best inducer of the antifungal peptides. These data suggest that discrimination between invading microorganisms can lead to higher levels of activation of the appropriate genes. Also it was shown that metchnikowin, a peptide with activity against bacteria and fungi, was effectively activated by both types of organisms (Lemaitre et al., 1997). The likely explanation for the specificity of these responses is that *Drosophila* spp. and other insects produce extracellular pattern recognition
molecules capable of initiating the pathway that leads to induction of the necessary peptides (Medzhitov and Janeway, 1997).

Mammals produce several extracellular pattern recognition molecules that act to limit an infectious challenge. Lipopolysaccharide-binding protein (LBP) recognizes and binds to the lipid A moiety of LPS and allows effector cells to be triggered by subpicomolar concentrations of endotoxin (Hoffmann et al., 1999). The activated LBP molecule will then bind with immune effector cells through CD14 binding in association with the mammalian Toll-like receptor. Mannose-binding protein (MBP), a member of the collectin family of first line host defense molecules, selectively recognizes the carbohydrate patterns found on various microorganisms. Once MBP binds to the appropriate ligand on a foreign invader it is capable of activating complement through the lectin pathway which ultimately leads to the activation of C3 convertase (Thiel et al., 1997).

In *Drosophila* spp. the Toll and 18-wheeler membrane receptors can mediate an immune response. However it is unclear whether they interact directly with microbial products in a manner similar to that found with the mammalian Toll-like receptors. In the fly the activation of Toll takes place via the binding of the endogenous ligand Spaetzle that is generated by a proteolytic cascade following its activation by a specific microbe (Lemaitre...
et al., 1996). Recently several proteins have been identified in *Drosophila* spp. that may serve as recognition molecules capable of inducing an antimicrobial response. A Gram-negative bacteria-binding protein (GNBP) has been observed in a soluble form and as a glycosylphosphatidylinositol-anchored membrane form on the surface of immunocompetent cells (Kim et al., 2000). *Drosophila* spp. GNBP has a high binding affinity for LPS, and \( \beta-1,3\)-glucan, but no binding affinity for peptidoglycan, \( \beta-1,4\)-glucan, or chitin. The overexpression of the GNBP enhances the inducibility of antimicrobial peptide genes in response to LPS and \( \beta-1,3\)-glucan, and blocking of GNBP with antibody inhibited this inducibility. These data suggest that GNBP is a functional pattern recognition molecule that plays a role in immune signaling.

A second type of signaling molecule is the peptidoglycan recognition protein (PgRP) that had been initially isolated from the hemolymph of the silkworm *B. mori*. Silkworm PgRP was able to bind peptidoglycan but not \( \beta-1,3\)-glucan, or chitin, and it was required to trigger the prophenyloxidase cascade in the plasma fraction of hemolymph in the presence of peptidoglycan indicating that PgRP/peptidoglycan binding was required to initiate the reaction (Yoshida et al., 1996). Subsequently a number of PgRP sequences have been found in mouse and humans where they are expressed.
in hematopoietic tissue (Kang et al., 1998) and several other insects where they are produced in the fat body and hemocytes (Michel et al., 2001). Data from the *Drosophila* spp. genome project indicates that PgRP constitutes a highly diversified family with at least 12 members (Werner et al., 2000). This family was grouped into two classes with the first class containing members with predicted structures for extracellular proteins and the second class containing intracellular and membrane spanning proteins. Transcripts from these genes were found in the immunity related organs such as the fat body, gut, and hemocytes. Several of the proteins could bind peptidoglycan, and a mutation in one of the PgRP genes (dPgRP-SA) was sufficient to stop Toll activation by Gram-positive bacteria (Michel et al., 2001). Thus PgRP is also capable of acting as a foreign pattern recognition and signaling molecule for the induction of an antibacterial response.

An additional class of putative recognition proteins in insects may be the lectins. Lectins have been implicated in recognition of microorganisms in vertebrates and invertebrates (Vasta et al., 1999). In mammals the mannose binding lectin has been implicated in activating complement to either opsonize microbes or form membrane attack complexes on their surfaces. Lectins have been isolated from a variety of insects including from *M. sexta* following bacterial infection (Minnick et al., 1986) and from *S.*
peregrina following larval injury (Komano et al., 1980). A single lectin has been purified from D. melanogaster (Haq et al., 1996) while at least 19 others have been identified from the Drosophila spp. genome based on their carbohydrate recognition domain (Theopold et al., 1999). Insect lectins also act in recognition and induction of various innate immune functions.

In the insect, the pathway from infection and recognition to induction of cellular receptors is not well mapped out. The Toll receptor is activated in Drosophila spp. by a proteolytically cleaved form of the polypeptide Spaetzle. What is not understood is the pathway and the components necessary to cleave Spaetzle although it is likely under the control of a serine protease cascade and their inhibitors the serpins. The mosquito A. gambiae has several serine protease genes, named the division 14 proteases, that showed changes in transcript abundance following various types of immune challenge (Gorman and Paskewitz, 2001). Phylogenetic analysis of these proteases indicates that they are most similar to the prophenyloxidase activating enzymes (PPAE), and the protease Easter from Drosophila spp. These similarities suggest that one or more of them may act as a PPAE or an activator of Spaetzle-like ligands. In Drosophila spp., the involvement of a proteolytic cascade in the cleavage of Spaetzle following immune challenge has only been indirectly substantiated. A hemolymph serine protease
inhibitor (serpin), Spn43Ac, negatively regulates the Toll pathway leading to
the production of drosomycin. A loss-of-function mutation in Spn43Ac leads
to spontaneous cleavage of the Spaetzle protein and continuous expression
of the drosomycin gene (Levashina et al., 1999). The genome-wide
microarray results from *D. melanogaster* showed upregulation of serine
proteases and that several genes encoding easter and snake-like proteases
were induced, especially in response to Gram-positive bacteria (De Gregorio
et al., 2001; Irving et al., 2001). The *snake* and *easter* gene products are
involved in Spaetzle cleavage in early embryos during dorso-ventral
patterning and no up-regulation of these genes was observed following
bacterial infection. However the upregulation of the *easter*- and *snake*-like
protease genes indicates that their induction by bacterial challenge is part of
the protease cascade necessary for Toll activation.

The final step from infection to production of antimicrobial peptides is
the Toll signaling pathway. The Toll gene was identified in the early 1980’s
as an essential component in embryogenesis in *Drosophila* spp. The Toll
receptor is a transmembrane protein with an extracellular portion with
leucine-rich repeats and an intracytoplasmic portion with shared sequence
similarity to the IL-1-receptor family, the so-called Toll/IL-1-receptor
domain (TIR) (Means et al., 2000). In addition to Toll and 18-Wheeler
(Toll2) the *Drosophila* spp. genome contains eight other related genes that have Toll-like domains (Tauszig et al., 2000). The current model for the activation of the gene for the antifungal drosomycin requires the cleavage of the Spaetzle protein, a secreted protein from the cysteine-knot family of growth factors. Proteolytic cleavage of Spaetzle is required for activation of Toll although direct interaction between the two proteins has yet to be reported (Medzhitov and Janeway, 2000). Once activated the Toll signal is transduced through the cytoplasm in a complex series of steps starting with the adaptor molecule Tube and the Serine/Threonine kinase Pelle, which leads to the phosphorylation and degradation of the inhibitor Cactus. Degradation of Cactus releases Dif, a member of the Rel family of transcription factors, which translocates to the nucleus and is believed to bind to and activate the drosomycin promoter (Imler and Hoffmann, 2001). In *Drosophila* spp., the Rel proteins are members of the Nuclear Factor-κB (NF-κB) family of transcription factors and are believed to be major regulators of antimicrobial gene expression through their interactions with the κB binding sites on the promoters of these genes (Khush et al., 2001).

The discovery of the Toll receptor in *D. melanogaster*, and the characterization of the key role it plays in host immune defense, lead to the description of the first mammalian homolog, the Toll-like receptor 4 (TLR-
4) in 1997 (Medzhitov et al., 1997) followed by the discovery of nine other homologs (Imler and Hoffmann, 2001). The mammalian TLRs are similar in structure to the Toll receptors and contain the intracytoplasmic TIR domain as well as the extracellular leucine-rich repeat regions flanked by cysteine-rich regions. In contrast, the mammalian IL-1-receptor’s extracellular domain is characterized by three Ig regions. Studies into the TLR’s have established that they are involved in the recognition of such microbial molecules as LPS, lipoteichoic acid, peptidoglycan, lipoproteins, and unmethylated bacterial CpG (Means et al., 2000). Also, TLRs direct the immune response through a pathway leading to activation of NF-κB. Additionally, unlike the activation of Drosophila spp. Toll, TLR-4 is activated by the direct binding of LPS and not through the binding of an activated accessory protein (Lien et al., 2000). Various TLR’s are found on the surfaces of host defense cells such as macrophages, neutrophils, and epithelial cells. Also a large variety of TLR’s are expressed by dendritic cells which may reflect their role in detection of pathogens and activation of the adaptive immune system (Visintin et al., 2001). Mammalian TLR’s play a central role in the detection of pathogenic microorganisms and the induction of the innate response. There is now much evidence that Toll and
other related receptors should play a similar role in the *Drosophila* spp. innate response.

Although the innate immune response in *Drosophila* spp. is the best characterized to date there is also considerable research being performed on the mosquito. Synthesized cecropin-like molecules can kill *Plasmodium* spp., and insect defensins or the general immune response can reduce the development of parasites in mosquitoes (Lowenberger et al., 1999). In the mosquito *A. aegypti*, the antibacterial peptides cecropin (Lowenberger et al., 1999) and defensin (Chalk et al., 1994) have been identified and the induction of a transferrin molecule was observed in this mosquito following wounding or bacterial inoculation (Yoshiga et al., 1997). Additionally, several serine proteases found in *A. gambiae* with similarity to enzymes with presumed roles in melanization or antimicrobial peptide synthesis in other insects have been identified and observed undergoing upregulation in transcript abundance following infection (Gorman and Paskewitz, 2001). Mosquitos have a fairly complex innate response but its effect on infection and transfer of microorganisms from vector to host remains unknown.

*X. cheopis*, a flea vector of the plague bacillus *Y. pestis*, also likely mounts a complex innate immune response. Although most of the effector molecules will likely be in the hemocoel there may be others produced and
secreted from epithelial cells in the midgut as has been observed in *Drosophila* spp. (Ferrandon, 1998). *Y. pestis* is taken into the flea with the blood meal and remains in the gut where infection of the next host occurs through mechanical passage of bacteria from the gut to the bite site. *Y. pestis* and *Y. pseudotuberculosis* are more resistant than *Y. enterocolytica* to the cationic antibacterial Polymixin B (Bengoechea et al., 1998) and *Y. pestis* appears to have some resistance to insect cecropin A (J. Hinnebusch, personal communication). *Y. pestis* may have adapted to the innate immune response in the flea in order to be maintained in the flea gut.

Beginning with the isolation of the cationic antibacterial peptide cecropin from *H. cecropia*, research into innate immunity has yielded a picture of a complex and many-faceted system. Studies on the innate response in *Drosophila* spp. have been the best understood due to the ability to mutate single genes and characterize that effect on a specific response. These results have lead to breakthroughs in the understanding of not only insect immunity and the evolution of the immune system but also the mammalian innate response, and its importance during the initial exposure to a pathogen as well as its ability to recruit components of the adaptive response. Many antibacterial peptides and proteins have been isolated from eukaryotic organisms (Lehrer and Ganz, 1999). Some have shown promise
as new types of antibiotics to help counter the problem of antibiotic resistance that is increasing in strains of bacterial pathogens (Cole and Ganz, 2000). Further research into the innate immune systems of vertebrates and arthropods will likely uncover novel innate mechanisms and immune effector molecules as well as increase the understanding of the evolution of the immune response.

1.2. Research Objectives

The experiments conducted and the results reported in this dissertation were designed to identify the initial innate antibacterial immune response in a member of the Order Siphonoptera. Our research has shown that the flea *X. cheopis* produces an antibacterial protein or proteins following an internal bacterial challenge. This protein is likely a cationic peptide and a member of one of the families of induced cationic peptides found in insects and other animals. Bacterial challenge may also induce the production of additional immune effector proteins that may have similarity to those found in other organisms. Preliminary research for this project indicated that inoculation or injection of bacteria into the flea hemocoel produced a protein capable of killing *E. coli*. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The specific goals of this research were:

1. Determine if bacterial challenge can induce the production of antibacterials and other proteins and use protein methods to observe the pattern of induction.

2. Isolate and purify an antibacterial peptide found in *X. cheopis* fleas and determine its amino acid sequence. Using the amino acid sequence, make a probe for the isolation of the gene responsible for producing this antibacterial peptide.

3. Use mRNA isolated from bacterially-challenged and unchallenged fleas in a subtractive hybridization process to isolate cDNA transcripts induced by bacterial infection. The cDNA transcripts will be sequenced and the sequences compared to known cationic antimicrobial peptides and to other known innate immune effectors to determine sequence similarities.
Chapter 2
Experimental Procedures

2.1. Flea Colony Care and Maintenance

Adult female *X. cheopis* fleas were used as the source of whole organism and hemolymph proteins. Flea colonies were kept at 23° C with 65% humidity in the dark. The flea colonies were fed twice weekly with neonatal mice.

2.2. Growth and Quantitation of Bacteria

Liquid cultures of *Escherichia coli* D31 (streptomycin resistant) and *Micrococcus luteus* were used to provide bacteria for injection of a known number into the flea hemocoel. They were grown in Luria broth (pH 7.2) at 37° C with shaking to an OD$_{600}$ of 0.5. Cultures were pelleted at 3000g for 5 minutes, rinsed with 50 mM sodium acetate buffer (pH 7.0), re-pelleted and resuspended in fresh sodium acetate to an OD$_{600}$ of 0.5. The bacteria *E. coli* D31 and *M. luteus* were inoculated onto LB agar plates and incubated overnight at 37° C to form colonies for making the bacterial paste used in flea inoculations. To determine the number of viable bacteria present in a
culture for injection of individual fleas, the washed *E. coli* cultures (OD<sub>600</sub> = 0.5) were counted with a Petroff/Hausser counting chamber under a light microscope. The stock culture was serially diluted to the desired number of bacteria per microliter. The dilutions were confirmed by plating on selective LB agar plates and counting the number of colonies following overnight incubation.

2.3. Bacterial Challenge of *X. cheopis*.

Fleas were chilled on ice and then separated in a glass dish on a chilled table. Only females were used for bacterial challenge. Fleas were embedded head down into grooves marked in a clay surface placed on a glass slide on the chilled table. Fleas were inoculated with a finely drawn glass needle produced in the lab from 50 ul borosilicate glass pipettes that were heated and drawn in a Narishige pipette puller. The pulled pipettes were dipped into a viscous mixture of bacterial colonies and the fleas were inoculated by inserting the coated needle between two of the hind tergites, or tegumental plates and into the hemocoel. This process inoculated a mass of bacteria into the open circulatory system of the flea. Once inoculated, the fleas were kept at 23° C with 65% humidity in the dark for time periods ranging from 2-24 hours in 2 hour increments. After the incubation period
the fleas were chilled on ice and collected for storage or immobilized in clay for collection of hemolymph. Fleas were also inoculated with a specific number of exponential phase *E. coli* suspended in 50 mM sodium acetate buffer (pH 7.0). The *E. coli* were grown to an OD$_{600}$ of 0.5 at 37° C in Luria broth. This culture was diluted in 50 mM sodium acetate buffer to the desired number of CFU/ ml. Approximately 0.1 ul of the suspension was injected into the flea hemocoel using a pulled pipette needle. The needles were produced by grinding the pulled pipettes (previously described) on a Narishige grinder.

2.4. Lipopolysaccharide Inoculation of Fleas

Various concentrations of LPS were used to inoculate *X. cheopis* fleas to induce an antibacterial response. LPS from *E. coli* serotype 0111:B4 (Sigma Chemicals Co., St Louis, MO) was reconstituted from a lyophilized powder with 50 mM sodium acetate buffer (pH 7.0). Concentrations of 1 mg/ml and 10 mg/ml were used for inoculation of fleas using a needle dipped in the desired solution. Concentrations of 10 mg/ml, 50 mg/ml, and 100 mg/ml were used for the injection of approximately 0.1 ul into the hemocoel of each flea.
2.5. **Arthropod Protein Isolation**

Whole fleas were ground in an acidic extraction buffer. The buffer consisted of a 10% acetic acid solution to help extract cationic proteins or peptides, 1 mg/ml phenylthiourea to inhibit the clotting mechanism and 10 ug/ml aprotinin to inhibit protease activation. The fleas were ground by hand in a microfuge tube using a small pestle. The debris was spun down by centrifugation at 12,000g at 4° C for ten minutes and the first supernatant was removed. This was followed by a second grinding step in fresh buffer, a second centrifugation at 12,000g at 4° C for ten minutes, and removal of a second supernatant fraction. Supernatants were pooled and a final centrifugation step at 18,500g at 4° C for 20 minutes was performed to remove any fine debris remaining in the mixture. All steps were carried out on ice or at 4° C. The supernatants were heat-treated by immersion in boiling water for 5 minutes to inactivate proteases and then centrifuged at 18,500g for 5 minutes to remove any coagulated proteins. In the initial stages of this research the heat-treated supernatant was lyophilized and resuspended in 0.2M sodium acetate buffer (pH 5.2). The resuspended product was tested for bioactivity on antibacterial bioassay plates and compared to the heat-treated supernatant.
Hemolymph was collected from individual fleas by puncturing the tegument with a thin beveled pipette that contained a small amount of ice cold acid extraction buffer (10% acetic acid, 1 μg/ml phenylthiourea, 10 μg/ml aprotinin) or ice cold extraction buffer (50 mM sodium acetate, pH 7.0, 1 μg/ml phenylthiourea, 10 μg/ml aprotinin). The flea body was squeezed with forceps to produce a drop of hemolymph, which was collected in the pipette by capillary action. Infected fleas and hemolymph were then stored at -70° C until used.

2.6. Solid Phase Extraction

To further separate the proteins from either whole organisms or hemolymph, solid phase extraction (SPE) was used. The extracts were diluted with a solution of 2% acetonitrile and 0.05% trifluoroacetic acid in MilliQ water. This solution was put through a solid phase extraction step using a Sep-Pak Plus C18 column (Waters Corp., Milford, MA) with isocratic gradients of 2%, 25%, 40% and 100% acetonitrile containing 0.05% trifluoroacetic acid. The fractions were vacuum-dried with heating using a Speed-Vac (Savant Instruments, Holbrook, NY), resuspended in MilliQ water, and tested for bioactivity.
2.7. Molecular Weight Centrifugation

The active fractions obtained from SPE were further purified by passage through Microcon Centrifugal Filters (Millipore Corp., Bedford, MA) with 30 kDa and 10 kDa molecular weight cutoffs. Samples were placed in the upper filter chamber and spun, using the manufacturer's recommendations, to collect the higher molecular weight proteins. The flow through from the 30 kDa filter was placed on a 10 kDa filter and the flow through from the 10 kDa filter was pooled, vacuum-dried with heating, and resuspended in sodium acetate buffer (pH 7.0).

2.8. Protein Concentration Assay

Protein concentrations for all samples used in column, centrifugation, and SDS-PAGE separation steps were determined using a Bio-Rad Protein Assay Kit (BioRad Laboratories, Hercules, CA). Protein samples were compared to the bovine serum albumin standard provided in the kit. Samples and standards were placed in a microtiter plate which included a distilled water blank as well as a sample solvent blank. The plates were placed in a Thermo Labsystems multiskan Plus microplate reader (fisher Scientific Co., Pittsburgh, PA) and read at an absorbance of 595 nm.
2.9. **RP-HPLC**

The resulting active fractions obtained from initial protein extraction of whole organisms, SPE separation, and/or molecular weight centrifugal filtration, were separated by reverse phase high-performance liquid chromatography (RP-HPLC). An ACTA Explorer 10 fast protein liquid chromatograph (FPLC) with a Sephasil Peptide C18 column (4.6 x 100 mm) (Amersham Biosciences, Piscataway, NJ) was used for the separations. To produce the solvent gradient, solvent A contained 2% acetonitrile with 0.1% trifluoroacetic acid in MilliQ water and solvent B contained 80% acetonitrile with 0.1% trifluoroacetic acid in MilliQ water. Samples containing up to 100 ug of protein were diluted in solvent A and then loaded onto the column. Several gradients were used to obtain desired separations of the proteins in the samples. For centrifuged and heat-treated whole flea lysate the initial gradient was 0% - 100% solvent B in 100 minutes at a flow rate of 1ml/min. Samples were collected in 1 ml aliquots, vacuum-dried, resuspended in MilliQ water, and bioassayed to identify antibacterial activity. Bioactive samples were run in a second round of RP-HPLC that employed a steep initial gradient of 0% - 30% solvent B over 15 minutes followed by a shallow gradient from 30% - 40% solvent B over 50 minutes (0.2%/minute) at a flow rate of 0.8 ml/minute. Samples were collected in 0.8 ml aliquots,
vacuum-dried, resuspended in MilliQ water, and bioassayed to identify fractions with antibacterial activity. Samples that had been separated by molecular weight cutoff centrifugation and/or solid phase extraction were separated using a steep initial gradient of 0% - 30% solvent B over 15 minutes followed by a shallow gradient from 30% - 40% solvent B over 50 minutes (0.2%/minute) at a flow rate of 1 ml/minute. Samples were collected in 1 ml aliquots, vacuum-dried with heating, resuspended in MilliQ water, and bioassayed to identify fractions with antibacterial activity. The second round of RP-HPLC would employ a steep initial gradient of 0% - 30% solvent B over 15 minutes followed by a shallow gradient from 30% - 40% solvent B over 50 minutes (0.2%/minute) at a flow rate of 0.8 ml/minute. Samples were collected in 0.8 ml aliquots, vacuum-dried, resuspended in MilliQ water, and then bioassayed for antibacterial activity. The pattern of the peaks on the chromatograms from the various RP-HPLC steps was used to determine how well the fractionation steps had performed. The samples were run on SDS-PAGE gels to confirm the results of the RP-HPLC fractionation steps.
2.10. Ultracentrifugation

Hemolymph from uninduced and E. coli-induced fleas was collected into chilled 50 mM sodium acetate buffer (pH 7.0) with phenylthiourea and aprotinin and centrifuged at low speed (3000g) for 15 minutes at 4° C. The supernatant from the low speed centrifugation was removed and placed in clean tubes and this cleared hemolymph was ultracentrifuged in a Beckman TL-100 Ultracentrifuge with a TLA 100.2 rotor at 200,000g for 1 hour at 2° C. The ultracentrifuged hemolymph was used for 1-dimensional (1D) and 2-dimensional (2D) SDS-PAGE analysis.

2.11. Antibacterial Bioassay

Determination of the antibacterial properties of the extracted protein fractions was performed following the protocol of Hultmark, et al. (1980). Briefly, the assay for antibacterial activity was carried out on a lawn of E. coli D31 (streptomycin resistant) grown in Luria-Bertani broth with 7% agarose (LB-agarose). First, the E. coli were grown to mid-exponential phase (OD\text{600} = 0.5) in Luria broth at 37° C with shaking. Seven ml of LB-agarose was melted and cooled to 45° C followed by the addition of 35 mg of lysozyme, 7 ul of 100 mg/ml streptomycin, and 10 ul of mid-log E. coli. The solution was spread on a 100 mm diameter sterile Petri dish and cooled
to room temperature. 2.5 mm or 3.5 mm wells were drilled in the LB-agarose and aliquots of samples to be tested for activity were placed in the holes and incubated overnight at 37° C. The plates were observed for zones of clearing in the bacterial lawn around the wells indicating bacterial killing. Any observable clear zone around the well was an indication that the sample has antibacterial activity. Samples of the agar were removed from the clear zone and were re-plated on to fresh LB plates to determine if the compound was bacteriostatic or bacteriocidal. Following the determination of antibacterial activity, samples were treated with proteinase K to determine if the bioactivity was protein-based. Lyophilized proteinase K was resuspended in distilled water to a concentration of 200 ugs/ul (2.6 U/ul) and added to samples at a concentration of approximately 1 U/ug of protein to be digested. Samples were incubated in a waterbath at 37° C for 2 hours and then placed in bioassay wells which included untreated samples for comparison.

2.12. Determination of Specific Activity

To improve the purification steps and to determine the enrichment of the antibacterial protein in each step, a specific activity curve was produced using purified cecropin A (Sigma Chemical Co., St. Louis, MO.) in an
antibacterial bioassay. Decreasing amounts of cecropin A were diluted in 6 ul of distilled water and placed in 3.5 mm diameter wells on an antibacterial bioassay plate, incubated overnight, and the diameter of the clear zones was measured. One microgram of cecropin was arbitrarily designated as having 10,000 Units of activity and dilutions were made down to 0.001 ug of cecropin with 10 Units of activity. By correlating the diameter of a zone of clearing from a known quantity of sample protein with the amount of cecropin required to produce a similar zone it could be determined how many units of activity per microgram (U/ug) a specific protein fraction contained. This also gave a rough estimate of how many micrograms of the active protein was contained in a mixture of proteins allowing the estimate of the total protein amount that was necessary to visualize the bioactive peptide in a silver-stained SDS-PAGE gel and/or on a Western blot.

2.13. Polyacrylamide Gel Electrophoresis

Fractions from all purification steps were tested for antimicrobial activity by using the antibacterial bioassay and those with antibacterial activity were further characterized using polyacrylamide gel electrophoresis. Bioactive samples and uninduced control samples were separated and compared by SDS-PAGE. Samples from whole flea lysates or hemolymph
were either run as collected or further separated by SPE and/or molecular weight centrifugation. These samples were run on 12.5% tris-glycine, 16.5% acid-urea, 16.5% tris-tricine, and 4-20% tris-tricine gradient polyacrylamide gels followed by silver staining. The purified active HPLC fraction(s) were run on a 16.5% acid-urea polyacrylamide gel and either silver stained or blotted on to a polyvinylidifluoride (PVDF) membrane in a Trans Blot SD semi-dry blotting system (Bio-Rad Labs). Alternatively, the HPLC fraction(s) were run on 16.5% Tricine-SDS-PAGE gels and either silver stained or blotted to a PVDF membrane in either a Mini Trans-Blot Cell system or a larger format Trans-Blot Cell System (Bio-Rad Labs).

Flea hemolymph from bacterially challenged and unchallenged fleas was separated by 2D gel electrophoresis using first dimension isoelectric focusing. The first dimension was run using the IPGphor Isoelectric Focusing System (Amersham Biosciences) with the protein placed on Immobiline Drystrip gels containing an immobilized pH gradient of pH 3-10. The ultracentrifuged hemolymph samples were solubilized in rehydration buffer containing 8 M urea, 2% (w/v) CHAPS detergent, 1% (v/v) IPG pH 3-10 buffer and 2.8 mg/ml DTT. The samples were incubated in rehydration buffer at room temperature for 4 hours, centrifuged at 100,000g at 22°C for 1 hour then placed on the pH 3-10 Drystrip in a
ceramic strip holder. The strip holder was placed on the IPGphor isoelectric focusing system and allowed to rehydrate and absorb the protein sample for 12 hours at 20° C. The strips were focused with a step and hold gradient at 250 volts for 1 hour, 500 volts for 1 hour, 1000 volts for 1 hour, and 8000 volts until approximately 18,000 volt-hours was reached. The strips were either frozen at -80° C or immediately run in the second dimension using SDS-PAGE gel. For the second dimension the strips were incubated in SDS equilibration buffer for 10 minutes at room temperature with shaking then placed on the surface of a 12.5% tris-glycine SDS-PAGE gel. The gel was run at 80 mA constant current at 14° C until the dye front had completely come off the gel. The gels were immediately immersed in a methanol/acetic acid fixative followed by silver staining.

2.14. Densitometry

Densitometry was performed on silver-stained 1D and 2D 12.5% SDS-PAGE gels of ultracentrifuged flea hemolymph from naïve and bacterially challenged fleas. The gels were scanned using a BioRad Gel Doc scanner (BioRad Labs). The 1D gel band densities were read and compared using the QuantOne software program and the 2D gel spot densities were read and compared using the PDQuest software program. The band and spot
comparisons were used to determine if proteins from bacterially challenged fleas were up- or down-regulated as compared to unchallenged fleas.

2.15. Protein Blotting

Proteins separated on Tris-tricine SDS-PAGE gels were blotted onto PVDF membranes. The proteins were transferred using a 10mM CAPS buffer, pH 11.0 with 10% methanol in a BioRad gel blotting apparatus. The proteins were transferred using 250 mA current for 3 hours at 16° C. The blots were removed from the gel surface, rinsed in distilled water, and stained with 0.1% amido black stain in 40% methanol/10% acetic acid for 5 minutes. The PVDF blots were destained in 3 washes of 40% methanol/10% acetic acid followed by several rinses in distilled water then air dried and stored at -20° C.

2.16. Amino Acid Sequencing

Specific bands or spots blotted onto a PVDF membrane were cut out, placed in microfuge tubes and frozen at -20° C. The proteins were sequenced using automated Edman sequencing at the Research Technologies Branch of the National Institute of Allergy and Infectious Diseases at the National Institutes of Health, Rockville, MD. The amino acid sequences obtained
were aligned by BLAST and a translated BLAST search to identify similarities with other immune effector genes.

2.17. **Extraction of Flea RNA and mRNA**

Semi-gravid *X. cheopis* fleas (gravid but with no obvious egg cases or embryos in the hemocoel when observed under a dissecting scope) were divided into two equal groups based on approximate size. The infected group was injected with approximately 1000 *E. coli*, as determined by direct count, into each flea. Each group was kept on a chilled table until the injections were completed and then they were placed in a 27°C incubator for 5 hours. Both groups were collected, chilled on ice, and placed in a -80°C freezer. Frozen fleas were placed in a Fast RNA green lysing matrix tube (Q-Bio Gene, Carlsbad, CA) with 1 ml of cold Trizol (Sigma Chemical Co.) and shaken in a Fastprep FP 120 shaker (Q-Bio Gene) for 25 seconds at setting #6, placed on ice for 30 seconds, and shaken for another 25 seconds at setting #6. Total flea RNA was extracted from the Trizol following the standard protocol, pelleted, dried, and resuspended in sterile distilled water. Flea mRNA was purified from total RNA using a Clontech Nucleotrap mRNA Purification Kit (ClonTech Labs Inc., Palo Alto, CA) following the standard kit protocol. The concentration of the flea RNA and mRNA
solutions were determined by spectrophotometry at 260 nm and purity determined by the 260/280 nm ratio. The mRNA pellet was resuspended in sterile distilled water and stored at -80° C until used for cDNA synthesis.

2.18. cDNA Synthesis

Purified mRNA from uninjected and E. coli–injected fleas was used to produce representative groups of cDNA using the Clontech SMART PCR cDNA Synthesis Kit (Clontech Labs). The SMART PCR technique uses a poly(dT) primer and MMLV reverse transcriptase to produce the first strand cDNA. During first strand synthesis the MMLV-RT will add a few additional nucleotides, usually deoxycytidines to the 3’ end of the cDNA. The SMART oligonucleotide primer contains an oligo(dG) sequence at its 3’ end and base pairs with the deoxycytidine stretch. This allows the RT to switch templates and replicate the end of the oligonucleotide. The full length single-stranded cDNA will have a poly (dT)-primed tail and an end complementary to the SMART oligonucleotide primer. The single stranded cDNA is then amplified using long distance PCR to produce a library enriched for full-length cDNA strands. The uninduced and induced cDNA’s were purified by column chromatography followed by Rsa I digestion. The digested cDNA was ethanol precipitated, resuspended in buffer, quantified
by UV spectrophotometry, and diluted to the final concentration of 300 ng/ul required for use in the subtractive hybridization step.

2.19. Subtractive Hybridization

Subtractive hybridization was performed on cDNA libraries produced from uninduced (naïve) *X. cheopis* and fleas injected with *E. coli* following the protocol from the Clontech Subtractive Hybridization Kit. A forward subtraction of the flea cDNA was performed with the induced cDNA being ligated to adaptors and being used as the tester cDNA. The adaptor-ligated cDNA’s were mixed with an excess of unligated, uninduced flea cDNA, denatured, and allowed to hybridize. This yielded hybridization of common sequences from the uninduced and induced pools without the necessary adaptors for efficient PCR amplification, and left a group of induced sequences with the correct adaptors for exponential amplification by PCR. The subtracted library was subjected to two rounds of PCR amplification using adaptor-specific primers supplied with the kit and the resulting products were run on a 2% agarose/EtBr gel to observe the preferential amplification of cDNA transcripts from the induced library.
2.20. TA Cloning of Subtracted Library

The PCR-amplified subtracted library was cloned and amplified using the TOPO TA Cloning Kit (Invitrogen Life Technologies, Carlsbad, CA). The transformation was performed using the pCR 2.1-TOPO vector and TOP10 One Shot Chemically Competent *E. coli* cells. Two cloning reactions were performed with one containing 1 ul of the PCR product and a second with 4 ul of the PCR product. The vector ligations were incubated for 30 minutes on ice to maximize the number of transformants. A volume of 2 ul of each reaction was mixed in separate vials of chemically-competent *E. coli* and incubated on ice for 15 minutes. The bacteria were heat-shocked for 30 seconds at 42° C, incubated at 37° C with shaking for 1 hour, and plated on LB agar plates containing 50 ug/ml kanamycin and then spread with a solution of 40 mg/ml X-gal in dimethylformamide. The plates were incubated at 37° C overnight. All white and light blue colonies (transformants) from both reactions and several dark blue colonies were picked and transferred to 96-well PCR plates as well as to LB agar/kanamycin plates. The vector inserts were amplified by colony PCR using the M13 Forward and M13 Reverse vector primers supplied with the TOPO TA Cloning Kit. The PCR products were run on 2% agarose/EtBr gels to determine the presence and size of the inserted genes.
2.21. Sequencing of Flea Genes

DNA samples from all the white and light blue colonies obtained from cloning of the subtracted flea cDNA library were transferred from LB agar/kanamycin plates to 96-well PCR plates. Sequencing reactions were performed using the TOPO TA Cloning Kit M13 forward and reverse primers and the Perkin Elmer Big Dye Terminator Sequencing Reaction Kit. Samples from the PCR reactions were placed in 96 well sequencing plates and run on a sequencing gel in a Perkin Elmer ABI Prism 7700 Sequence Detector (Perkin Elmer Biosystems, Boston, MA).

2.22. Sequencing Data Analysis

Sequencing data from the cloned flea genes, in the form of chromatogram files, were assembled using the Finch Suite DNA Sequence Assembly Program (Geospiza Inc., Seattle, WA). The chromatogram files generated from the samples run on the sequencer were transferred to the Finch Server database. After transfer the chromatogram files were processed and parsed and quality values and reads were obtained from the chromatograms using the Phred algorithm to determine the quality of the sequence. Regions of high quality were identified and reads were screened.
for vector and *E. coli* contamination. The sequencing data were placed in reports and files for further analysis. The reports contained contigs assembled from multiple reads of a cloned gene or gene region. The contigs obtained from the Finch program were aligned against known gene sequences using a translated BLAST search (BLASTX) to identify genes with known immune function. The contigs for the tentatively identified immune genes were further analyzed using the Sequencher 4.1 program (Gene Codes Corp., Ann Arbor, MI.). Sequencher was used to further identify the correct bases by using the base-calling feature to compare chromatograms from each sequence in an alignment. The cleaned consensus sequence for the putative immune genes was again BLAST-searched (BLASTX) and also used to search for conserved domain architecture (DART). The best matches from the BLAST search were compiled in MacVector 6.5.3 (Oxford Molecular Ltd., Madison, WI), and the ClustalW algorithm was used to align these matches to the flea consensus sequence. The aligned sequences were also used to produce dendrograms to observe the relationships between known and putative sequences.
Chapter 3

Results

3.1. Antibacterial Bioassay Results from *X.cheopis*

Clean whole flea lysates from uninoculated fleas were consistently negative with no indication of a clear zone on the bioassay plates while the inoculated flea lysate produced clear zones in the agar following incubation of the fleas for either 12 or 24 hours (Fig. 1). A sterile inoculation followed by a 24 hour incubation did not induce an antibacterial response in the fleas that could be seen at the level of detection in the antibacterial bioassay (Fig. 1). Lysate from fleas inoculated with a paste of *E.coli* only and *M. luteus* only were combined and the bioassay showed a zone of clearing for both of these treatments (Fig. 1).
Figure 1. Bioassay performed on whole flea lysates from 60 *X. cheopis* adult females. Aliquots of centrifuged lysates from naïve, sterile inoculated, and bacterially-inoculated fleas were compared for antibacterial activity in a bacterial zone of clearing bioassay. When added to the wells, the uninfected and sterile inoculated fractions produced no clear zones while the fractions from bacterially-inoculated fleas produced antibacterial zones of clearing.
3.2. **SPE and Molecular Weight Cutoff Centrifugation of Flea Lysates**

Following an SPE isocratic fractionation of inoculated whole flea lysates the 25-40% fraction consistently showed the majority of the antibacterial activity based on the size of the clear zone in the bacterial lawn while the 40-100% fraction would occasionally show a small clear zone (Fig. 2). SPE was performed on uninoculated whole flea lysate and no bioactivity was observed for any eluted fraction (Fig. 3). The 25-40% SPE fraction was further separated on Centricon molecular weight cutoff centrifugation columns beginning with the 30kD column followed by the 10 kD column. Running the fractions through the two columns yielded a >30kD protein fraction, a <30kD but >10kD fraction (10-30kD), and a <10kD fraction. When these fractions were run on bioassay plates the results showed that the >30 kD fraction usually produced a small clear zone, the 10-30kD fraction would produce a larger zone from less total protein in the well, and the <10kd fraction would produce the largest clear zone from the smallest amount of protein per well. Protein gel analysis of the various fractions containing antibacterial activity was compared to similar fractions from uninoculated fleas in an attempt to observe a band or bands containing the antibacterial activity. Tris glycine SDS-PAGE with 12.5% and 16.5% acrylamide, 15% acid urea PAGE, and tris tricine SDS-PAGE using 16.5%
and a 10-20% gradient gel were used and no obvious induced bands were observed that could be attributed to antibacterials. A silver-stained 10-20% gradient Tris-tricine gel analysis of whole flea lysate and the SPE and molecular weight centrifugation purification steps was used to observe the results of these fractionation and purification procedures on the complexity of the lysate (Fig. 4). Whole organism lysates are a very complex mixture of proteins and it was necessary to further separate out the inactive proteins from the active samples.
Solid Phase Extraction Bioassay
Wash - . . Wash

40-80% . . Wash . . 40-80%

80-100% . . 0-2 . . 80-100%

25-40% . . 25-40%

Whole Flea Extracts Infected
Figure 2. Bioassay performed on whole flea lysates from 60 infected *X. cheopis* adult females. The lysates were separated by solid phase extraction using isocratic gradients of acetonitrile with 0.05% trifluoroacetic acid (TFA). The wash fractions were eluted from the column with distilled water containing 0.05% TFA. Aliquots of flea lysates from bacterial-inoculated fleas yielded an antibacterial zone of clearing in the bioassay plate from the fraction eluted with the 25-40% acetonitrile gradient. No antibacterial activity (zone of clearing) was observed from any other inoculated flea lysate fractions.
Figure 3.  Bioassay performed on whole flea lysates from 60 uninfected

*X. cheopis* adult females. The lysates were separated by solid phase

extraction using isocratic gradients of acetonitrile with 0.05% trifluoroacetic

acid (TFA). The wash fractions were eluted from the column with distilled

water containing 0.05% TFA. No antibacterial zone of clearing was

observed with aliquots of flea lysates from naive fleas.
3.3. **RP-HPLC of Whole Flea Lysates**

Using the whole flea lysate in a single round of RP-HPLC yielded several fractions with bioactivity that eluted around 31% acetonitrile concentration. However the chromatogram did not show a distinct peak but rather showed a broad area with multiple peaks in that region indicating that the fractions were still a complex mixture. Another round of SPE and molecular weight cutoff centrifugation steps were performed using a starting lysate derived from 600 individually inoculated fleas. The first round of RP-HPLC on the 25-40% SPE/10-30 kD fraction was performed using a steep gradient of 1%/min up to the 30% acetonitrile level followed by a shallow gradient of 0.2%/min from 30% to 35.2% acetonitrile. This yielded two fractions with antibacterial activity that eluted from the column at 31.2% acetonitrile concentration. The bioactive fractions were associated with a peak on the chromatogram that eluted at the end of the first fraction and a small portion of the peak carried over into the second fraction. The first fraction contained a significantly higher amount of bioactivity. A second round of RP-HPLC using the same gradient but a slower flow rate of 0.8 ml/min further separated this fraction. Bioassay results from this separation identified two fractions associated with a single peak that eluted at 31.2%
acetonitrile. After two rounds of HPLC the amount of protein left in each fraction was approximately 2-4 ug.

3.4. SDS-PAGE of Fractionated Whole Flea Lysates

Gel analysis of the bioactive fractions obtained by SPE followed by molecular weight centrifugation, and RP-HPLC was performed using acid urea PAGE and Tris-tricine SDS-PAGE. A comparison was made between the various separations performed on whole flea lysates before the use of RP-HPLC using a silver stained 10-20% Tricine SDS-PAGE gradient gel (Fig. 4). The results showed that even following several separation steps the resulting fractions were still very complex mixtures of proteins. SPE separation of the whole lysate produced samples with possibly fewer protein bands and a higher concentration of some lower molecular weight proteins. Molecular weight cutoff centrifugation using Microcon filters are designed to concentrate protein samples with molecular weights above the filter’s cutoff limit or to remove lower molecular weight contaminants. These filters were used to remove the higher molecular weight proteins from SPE separations and concentrate the lower molecular weight proteins and peptides in an attempt to collect enough of the active antibacterials to observe on a gel. The >30 kD Microcon fraction was a complex mixture of proteins similar in molecular weight range to the 25-40% SPE fraction. The
10-30 kD Microcon fraction obtained from the 30 kD filter flow-through contained several protein bands at or slightly above 30 kD with the majority of the bands between 6 and 30 kD in weight. The <10 kD Microcon fraction contained bands from approximately 4 to 10 kD. Attempts to blot proteins from the 10-30 kD and <10 kD samples from tricine gels to a PVDF membrane were unsuccessful. The blots did not yield discrete bands that could be individually excised for sequencing. Another sample of whole flea lysate from 600 *E. coli* and *M. luteus* inoculated fleas was separated by SPE and the 25-40% fraction was fractionated with Microcon molecular weight filters. The 10-30 kD fraction was further separated by two rounds of RP-HPLC and the bioactive fractions were run on a 10-20% Tricine SDS-PAGE gradient gel followed by silver staining (Fig. 5). Although the HPLC fractions were concentrated in the lower molecular weight region, they still contained a complex mixture of proteins that had many close bands and some smeared areas. The comparison of two samples without bioactivity that eluted from the column right before the bioactive fraction following the first round of HPLC showed similar banding patterns to the bioactive fraction, with a number of bands and no distinctly different band or bands that could be attributed to an antibacterial. A second round of RP-HPLC was performed on the bioactive sample from the first round and a similar result
was obtained. A comparison between the fractions eluted before and after the two bioactive fractions eluted showed no obvious discrete bands that could be attributed to the antibacterial protein(s) contained in the bioactive fractions. Additionally, the second round of HPLC did not improve the resolution of the protein bands to a point where they could be easily excised individually for sequencing.
Figure 4. Silver stained 10-20% Tricine SDS-PAGE gradient gel of various flea protein extracts. MW- Polypeptide molecular mass standards. Lane 1, untreated whole flea extract. Lane 2, heat-treated whole flea extract. Lane 3, 25-40% acetonitrile eluted SPE fraction. Lane 4, >30 kD molecular weight cutoff centrifugation fraction. Lane 5, 10-30 kD molecular weight cutoff centrifugation fraction. Lane 6, <10 kD molecular weight cutoff centrifugation fraction. Lane 7- cecropin A. Lane Bl- blank. Protein load in lanes 1-6 was 2 micrograms per lane. Lane 7 contained one microgram of purified recombinant cecropin A.
Figure 5. Silver stained 10-20% Tricine SDS-PAGE gradient gel of whole flea lysate HPLC fractions. MW, Polypeptide molecular mass standards. Lanes 1 through 3 were from HPLC run #1. Lanes 1 and 2 contained HPLC fractions 42 & 43 (no bioactivity). Lane 3 contained fraction 44 (Specific activity = 90U/ug). Lane 4 contained cecropin A. Lanes 5 through 8 were from HPLC run #2. Lane 5 contained fraction 31 (no bioactivity). Lane 6 contained fraction 32 (specific activity = 125 u/ug). Lane 7 contained fraction 33 (specific activity = 63 U/ug). Lane 8 contained fraction 34 (no bioactivity). All HPLC lanes loaded with three micrograms of protein per lane. Cecropin lane contained one microgram of purified recombinant cecropin A.
3.5. Specific Activity Assays of Antibacterial Fractions

A specific activity assay was used to determine the efficiency of the various methods used for separating out unwanted proteins and enriching the fractions for the desired antibacterial components. Table 1 shows the results from bioassays using various amounts of cecropin A and the average sizes of the clear zones that cecropin makes in a lawn of *E. coli* in LB agarose (Fig. 6). This table was used as a comparison to determine the Units of activity per microgram of sample after it was tested in a bioassay.

Table 1.

<table>
<thead>
<tr>
<th>Cecropin (ug/well)</th>
<th>Clear zone (mm)</th>
<th>Units (1 ug = 10,000U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>16.25</td>
<td>10,000</td>
</tr>
<tr>
<td>0.5</td>
<td>15.0</td>
<td>5,000</td>
</tr>
<tr>
<td>0.1</td>
<td>13.25</td>
<td>1,000</td>
</tr>
<tr>
<td>0.05</td>
<td>12.25</td>
<td>500</td>
</tr>
<tr>
<td>0.01</td>
<td>9.75</td>
<td>100</td>
</tr>
<tr>
<td>0.005</td>
<td>8.25</td>
<td>50</td>
</tr>
<tr>
<td>0.001</td>
<td>4.25</td>
<td>10</td>
</tr>
</tbody>
</table>
Figure 6. Bioassay performed with various amounts of the recombinant antibacterial peptide cecropin A. Dilutions of cecropin A were used to produce a specific activity graph for comparison with the antibacterials isolated from the arthropods tested during this research.
Insect cecropin A was also tested by SPE to determine its hydrophobicity based on its elution from the SPE column. Similarity in the hydrophobicity of the two samples based on the SPE results could give an indication of the similarity between the flea lysate-derived antibacterial and cecropin. The bioassay results indicated that the majority of this bioactive peptide would elute from the column in the 25-40% acetonitrile fraction while a smaller amount would elute in the 40-80% fraction (Fig. 7). Table 2 lists the best specific activity results obtained by each treatment of the whole flea lysates from *E. coli* and *M. luteus* inoculated fleas. The whole flea lysate was usually obtained from groups of 600 inoculated fleas. The best results were obtained from an initial SPE of the heat-treated lysate followed by the molecular weight cutoff centrifugation step. The RP-HPLC results shown came from elution of the proteins in the 10-30 kD fraction. HPLC fractionation of the <10 kD samples had inconsistent results and the amount of protein recovered from these bioactive fractions was usually the minimum necessary for observable clear zones on the bioassay.
Table 2.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Protein (ug) in 6ul well-loading volume</th>
<th>Clear Zone (mm)</th>
<th>Units/ug</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heat-treated</td>
<td>2.88</td>
<td>5.0</td>
<td>6.1</td>
</tr>
<tr>
<td>40% fraction</td>
<td>2.9</td>
<td>6.0</td>
<td>9.5</td>
</tr>
<tr>
<td>&gt;30 kD</td>
<td>9.0</td>
<td>4.0</td>
<td>0.8</td>
</tr>
<tr>
<td>10-30 kD</td>
<td>4.1</td>
<td>4.5</td>
<td>3.1</td>
</tr>
<tr>
<td>&lt;10 kD</td>
<td>1.5</td>
<td>8.0</td>
<td>32</td>
</tr>
<tr>
<td>HPLC #1</td>
<td>0.42</td>
<td>7.0</td>
<td>89</td>
</tr>
<tr>
<td>HPLC #2</td>
<td>0.18</td>
<td>5.5</td>
<td>125</td>
</tr>
</tbody>
</table>
Figure 7. Bioassay performed on cecropin A to determine its hydrophobicity based on the percentage of acetonitrile required to elute it from an SPE column. The peptide was separated by solid phase extraction using isocratic gradients of acetonitrile containing 0.05% trifluoroacetic acid (TFA). The wash fractions were eluted from the column with distilled water containing 2% acetonitrile and 0.05% TFA. Antibacterial activity based on a zone of clearing in a bioassay was observed in the 25-40% and the 40-80% acetonitrile eluted fractions, although the majority of the activity was found in the 25-40% fraction.
3.6. Amino Acid Sequencing of RP-HPLC Fractions

*E. coli* and *M. luteus* inoculated whole flea lysate was purified by RP-HPLC, separated by SDS-PAGE and blotted to a PVDF membrane. Several diffuse bands from the lower molecular weight section of the blot were sent for amino acid N-terminal sequencing. Only one band with a molecular weight of approximately 4 kD returned a partial sequence of six amino acids in length. No sequence similarity was found from a BLAST search of the partial sequence.

3.7. Bioassay of Flea Hemolymph

Uninoculated flea hemolymph did not produce a clear zone in any bioassay plate test even at high protein concentrations of up to 8 ug/well. Inoculated flea hemolymph produced variable zone sizes and specific activities from a low of 6 U/ug up to 35 U/ug depending on the length of incubation of the inoculated fleas. An observable zone of clearing could be produced in the bioassay plate with as little as 0.3 ug of infected flea hemolymph.

3.8. Optimization of Antibacterial response

Due to the variability of the specific activity results a series of experiments were run to optimize the yield of the antibacterials from hemolymph. A comparison between inoculations with a paste made from *E.*
coli and M. luteus and inoculations made with either E. coli or M. luteus was performed using bioassay plates having a lawn of lysozyme-treated E. coli. Results indicated that hemolymph collected from fleas inoculated with both types of bacteria together and E. coli alone produced clear zones but differed in their specific activities. Hemolymph from fleas inoculated with a mixture of both bacteria had a lower specific activity than hemolymph from fleas inoculated with E. coli alone. Hemolymph from fleas inoculated with a paste of M. luteus alone would not produce clearing in a lawn of E. coli.

When the three types of induced hemolymph were tested on the surface of a bioassay plate having a lawn of M. luteus, all three types produced zones of clearing. Lipopolysaccharide was injected into fleas to determine if it could induce the flea's immune system into producing an antibacterial response similar to that from whole bacteria. Tests of concentrations ranging from a needle dipped into 1 mg/ml LPS and injections of 0.1 ul of 1 mg/ml to 100 mg/ml LPS did not induce any observable clear zone of any size in a bioassay plate.

Tests were performed to optimize the length of the incubations following needle inoculation and bioassays were performed on hemolymph from groups of 20 fleas incubated at two-hour intervals starting at 0 hours until 24 hours and also at 36 hour and 48 hour intervals. Results from
bioassays of replicates from the time intervals for *E. coli* -inoculated fleas showed a high level of variability. To determine if the variability was due to differences in inoculum amount, a test was performed comparing needle-inoculation of the bacterial paste to the needle injection of a known number of bacteria into the hemocoel. Hemolymph from individual fleas either inoculated or injected with *E. coli* and incubated for 12 hours was collected into separate tubes containing a small amount of ice cold buffer. The relative amount of hemolymph from each flea was determined and the samples were placed in individual wells on a bioassay plate. The results showed that injection of fleas produced hemolymph with an antibacterial response from more individual fleas when compared with needle-inoculation with a bacterial paste. Also, higher specific activities were observed in the hemolymph from groups of fleas when compared to the inoculated fleas. Also it appeared that small fleas or those producing only a small amount of hemolymph were more likely to produce negative results. The best results were obtained from hemolymph from semi-gravid fleas (enlarged abdomen but no egg cases observed) or gravid fleas. Using semi-gravid fleas the *E. coli* culture was optimized at 1.2 x 10^7/ml with an injection of 0.1 ul/flea (approximately 1200 bacteria per flea as determined by direct plate count).
3.9. Purification and 1D Gel Analysis of Hemolymph

Ultracentrifuged hemolymph from uninjected and *E. coli*-injected fleas was separated by SDS-PAGE. Densitometry of the silver-stained bands observed indicated that 2 bands were induced to higher levels of production following bacterial challenge of fleas compared to naïve flea hemolymph controls. Band #1 was induced by a level of 1.7:1 although it was significant only at p > 0.10 (p=0.06). Band #2 was induced by 2.1:1 (p=0.04) and band #3 was induced by 1.3:1 (p=0.02). However the two significantly induced bands were of a higher molecular weight than you would expect from the commonly identified antibacterials (Fig. 8).
Figure 8. 12.5% SDS-PAGE of naïve and *E. coli*–injected flea hemolymph. Lane 1, Bio-Rad low range SDS-PAGE molecular mass standards. Lane 2, naïve flea hemolymph. Lane 3, *E. coli*-injected flea hemolymph. The proteins in bands 2 and 3 were determined to have been induced to a significantly higher level of production in flea hemolymph following bacterial injection based on densitometry analysis.
3.10. 2D Gel Analysis of Hemolymph

2D gel analysis of *E. coli*-injected and naïve flea hemolymph was performed using a pH 3-10 range first dimension separation followed by SDS-PAGE second dimension separation. Several spots were identified having been induced to a significantly higher level of production following bacterial challenge of the flea. However, no new spots were seen in the induced gels that were absent from the uninduced gels (Fig. 9) indicating that there were no proteins of bacterial origin in the induced gel profile. A comparison of two matched 2D pH 3-10 gels showed that very few spots could be visually identified as possibly being induced and the densitometry analysis was used to confirm that only three spots (#’s 2, 3, and 4) were significantly induced following bacterial injection (Fig. 10). The induced spots were again found only in the higher molecular weight regions of the gels. Spot #2 was induced by 2.1:1, spot#3 was induced by 1.6:1, and spot#4 was induced by 2.0:1 compared with the uninjected hemolymph controls.
Figure 9. 12.5% SDS-PAGE second dimension of pH 3-10 2D gel of *E. coli*-injected flea hemolymph. The proteins in spots 2, 3, and 4 were determined to have been induced to a higher level of production in the flea following bacterial injection based on densitometry analysis. Spots 6 and 7 were used as internal controls and spot 11 was an external control using the protease inhibitor aprotinin.
Figure 10. 12.5% SDS-PAGE second dimension runs of pH 3-10 2D gels, stained with silver, comparing naïve and *E. coli*-injected flea hemolymph. This figure shows a visual comparison of the bacterially-induced proteins #2, #3, and #4 that were confirmed by densitometry.
3.11. Amino Acid Sequencing of 2D gel Spots

Densitometry analysis was performed on five pairs of 3-10 pH flea hemolymph gels and the results showed that three proteins had been induced to significantly higher levels ranging from 1.6:1 to 2.1:1. A pair of 3-10 pH 2D gels were run using 45 ug of hemolymph from bacterial injected fleas and one was silver stained while the other was blotted on to a PVDF membrane (Fig. 9). The PVDF membrane was stained with amido black and the induced proteins and several control proteins were cut out of the membrane and sent to the Structural Biology Section of the NIH-NIAID for Edman amino acid sequencing. Sequence was obtained from two of the uninduced flea hemolymph spots but a Blast search did not find any alignments. The sequence obtained from Spots #6 and #7 on the 2D PVDF blot was DAPAQNLEIQATTLAKETL. A third spot, #11, from the 2D PVDF blot was tentatively identified, based on sequence similarity, as aprotinin, part of the hemolymph collection buffer components.

3.12. Subtractive Hybridization of cDNA Libraries

The PCR amplification of the ss cDNA used 2 ul of template and was optimized at 18 cycles to produce a smear of cDNA products ranging in size from 400 to 4000 bp in length. Samples from the second round of PCR were run on a 2% agarose/EtBr gel. A small number of discrete bands were
observed from the subtracted library amplification when compared to the initial mRNA gel, indicating that the subtraction process had eliminated a large number of cDNA transcripts (Fig. 11).
Figure 11. Agarose/EtBr gel of subtracted cDNA library from uninduced and bacterial-induced fleas. Lane 1, Lambda/HindIII - φX174/Hae III DNA markers in base pairs. Lanes 2 and 3, final PCR products from subtracted library amplified using Taq polymerase. Lanes 4 and 5, final PCR products from subtracted cDNA library amplified using Clontech PCR kit polymerase. Bright bands in gel are preferentially amplified bacterially-induced cDNA’s from the *E. coli* -injected fleas.
3.13. TA Cloning of Subtracted Library

The PCR-amplified subtracted library was cloned using the Invitrogen TA Cloning Kit. Three reactions were run using 1 ul, 2 uls, and 4 uls of the PCR product. After overnight incubation 384 clones containing vector inserts were picked from the plates, inoculated into 96 well PCR plates for colony PCR, and patched onto grids on new selective media agar plates for storage prior to sequencing. Colony PCR was used to determine the size of the inserts in the TA cloning vector. Insert size was determined following standard colony PCR protocols from the TA Cloning Kit using the M13 forward and reverse primers provided. The vector inserts ranged in size from approximately 2000 bp to 250 bp with most inserts being in the 500-750 bp range.

3.14. DNA Sequencing and Analysis of cDNA Library Clones

All 384 clones obtained were sequenced in both directions using the TOPO TA M13 reverse and T7 promoter forward primers. The sequences obtained were analyzed and assembled using the Finch Suite Assembler program and 168 contigs were determined by the program. A BLAST search using the program blastx 2.2.1 of all 168 contigs determined that 151 sequences had similarities to known protein sequences in the database. Of
the 151 sequences, 35 were chosen for further analysis due to their similarities to other known antibacterial or immune response genes.

3.15. Identification of Flea Defensin

Following Finch assembly a Sequencher analysis was performed on nine contigs that had sequence similarity to the antibacterial defensin. The consensus sequence for the flea defensin consisted of 552 bases and the amino acid sequence following translation of the DNA in the +1 frame produced a putative defensin peptide 102 amino acids in length (Fig. 12). The flea defensin mature peptide has a molecular weight of 4774.21 daltons and an isoelectric point (P.I.) of 12.31. The mature peptide sequence contains 8 basic amino acids and 2 acidic amino acids. The putative flea defensin was aligned to several other insect defensins using MacVector and the ClustalW alignment tool (Fig. 13). A comparison of the mature peptides (approximately 47 amino acids) from *D. melanogaster*, *Tenebrio molitor*, *Orcytes rhinoceros*, *Holotrichia diomphalia*, *Bombus pascorum*, and *Allomyrina dichotoma* showed significant areas of alignment as well as the 6 conserved cysteine residues that are the identifying hallmarks of a defensin. The best alignment was to the mature defensin peptide from the beetle *A. dichotoma* which had a 57% identity and an expect value of $2 \times 10^{-6}$. A conserved domain database search using RPS-BLAST was performed on the
flea amino acid sequence and the results indicated that it was similar to the family of arthropod defensins. A dendrogram using the best BLAST alignments showed that the flea defensin was most closely related to the *B. pascuorum* sequence (Fig. 14).
**FLEA DEFENSIN TRANSLATED SEQUENCES**

<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGGCCGCCAGTGATGATCTGCAAGATTGCCCTTAGCGTGGTCGCGGCCGAGGTAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCGGGATACACACACGTTAGTGACTAGAAGAAGACATTTACATATAGTGAATATAAAAAAAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>130</th>
<th>140</th>
<th>150</th>
<th>160</th>
<th>170</th>
<th>180</th>
</tr>
</thead>
<tbody>
<tr>
<td>AACACTTTTCATCACAGATGAGATTTGCCTATATAGTGCATCGATAGCCTTTTTTT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRFVIVVCIASIAFF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>190</th>
<th>200</th>
<th>210</th>
<th>220</th>
<th>230</th>
<th>240</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGGTTGTCTAGTCACTCTGTAGATCTATTCGAGATCCGTCACTTTACGGAGAG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVVQSVPSVEDLFEDPSLYGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>250</th>
<th>260</th>
<th>270</th>
<th>280</th>
<th>290</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCAGCTTATGGACTTTGAGGTTGGATGGGTTTTGATAATACCTCTGTAAGACAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAYGLVDELQGAFTDNTRGLVRQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>310</th>
<th>320</th>
<th>330</th>
<th>340</th>
<th>350</th>
<th>360</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAACGGGTTACATGATCTTTTAAATGTTAACCCCATGGCTCTAGATTCATGCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KRVTCDLSSLVNTPIGSNLHAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>370</th>
<th>380</th>
<th>390</th>
<th>400</th>
<th>410</th>
<th>420</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCTTGGCAGCTCATGCTTTAGTTACGTCGAGGATTTAAGAGGTCGCTGTGAANAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACAAHCCLSLRRGFKKGGRCEN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>430</th>
<th>440</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGAGTTGTCAACTGCCCAGAGA</td>
<td></td>
</tr>
<tr>
<td>GVCCNCRR</td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 12. *X. cheopis* defensin-like translated sequence derived from cloning of a subtracted library. Translated from the prepro region with the mature peptide in bold.
Figure 13. ClustalX alignment of defensin sequences from various insects to the defensin-like sequence of *X. cheopis*.
Figure 14. Dendrogram of alignment of defensin from various insects to the defensin-like sequence of *X. cheopis*.
3.16. Identification of Flea Cecropin

Sequencher was used to analyze the sequences that made up the three Finch contigs with sequence similarities to the antibacterial cecropin, Contigs 6, 8, and 113. A consensus sequence was obtained from each of these three putative flea cecropin contigs. The contigs were then aligned using Sequencher to produce the final consensus sequence (Fig. 15). Flea cecropin has a molecular weight of 4337.71 daltons and a P.I. of 10.49. Its amino acid sequence contains 10 basic residues and 3 acidic residues out of the 63 amino acids in the mature peptide. The flea cecropin consensus was used in a BLAST search using blastx and the best results were used in a ClustalW alignment with sequences from the Drosophila family, D. melanogaster, and Ceratitis capitata (Fig. 16). A conserved domain database search using RPS-BLAST was performed on the flea amino acid sequence and the results indicated that it has similarity to the family of arthropod cecropins but only at an expect value of 0.44. The best alignment for the putative flea cecropin was to the 63 amino acid prepropeptides cecropin A1 from D. mauritiana, and cecropin A2 from D. melanogaster. The alignments encompassed the entire 63 amino acid regions with 39% identities and an expect value of 0.11. When aligned against the 43 amino acid mature peptide region of D. melanogaster cecropin B the identity value
increased to 51%. A dendrogram using the best BLAST alignments showed that the flea cecropin was most closely related to the *C. capitata* sequence (Fig. 17). Identification of the flea peptide as a cecropin-like sequence is also based on the observation of α-helical structure within the sequence. Additional analysis of the amino acid sequence of flea cecropin, using the GORIV secondary structure prediction analysis program, determined that the mature peptide did contain a region of α-helical structure (Fig 18). The mature peptide region of *D. melanogaster* and *B. mori* analyzed by GORIV also showed α-helical structure.
FLEA CECROPIN TRANSLATED SEQUENCE

10 20 30 40 50 60
TCGAGCGGCCGCCGCCGCCAGGTACGCGGGGATCAGTTGATATCCGACTAACTTTAAAGCT

70 80 90 100 110 120
AATTTAAGTAATTCAATATTAGTGGTTGATATTTTAAAAATCTGCTAAGATGAAGTTGT

MKL

130 140 150 160 170 180
TCAATATTGTTTTTTTGTCCTGCTTGCTTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTCAGAAGCTA

FNIVFVLCLVFALLAVPS(EA

190 200 210 220 230 240
AATGGAAGTTCCGCAAGAAATGGGACGTATGGGAAAACGAGTATTGCACGCCACTGAA

KWFGKKLERMGKRLHATE

250 260 270 280 290
AAGTACTGCCAGCAGCTCAGGGGTGGCGGTGTAGCTACAGCTTTACAGAGGG

KVLPAAQGVAGVATAFRGG

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 15. *X. cheopis* cecropin-like translated sequence derived from cloning of a subtracted library. Translated from the prepro region with the mature peptide in bold.
Cecropin Alignments

Drosophila family
Drosophila melanogaster
Ceratitis capitata
Xenopsylla cheopis

KSAK | KKL | FVBFV

M. NIF. VAL. A

KSYV

NF NKV C I

FA T

M.

VAL.

A

GQ SE AG W L

Drosophila family
Drosophila melanogaster
Ceratitis capitata
Xenopsylla cheopis

FLM VRK LG QH

TT TRA

AV P

K

Q

A

N

A Q A A N

Drosophila family
Drosophila melanogaster
Ceratitis capitata
Xenopsylla cheopis

TAF

V A A T A R G

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 16. ClustalX alignment of cecropin sequences from several insects to the cecropin-like sequence of *X. cheopis*.
Cecropin Dendrogram

- Drosophila family
  - Drosophila melanogaster
    - Ceratitis capitata
      - Xenopsylla cheopis

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 17. Dendrogram of alignment of cecropin from several insects to the cecropin-like sequence of *X. cheopis*. 
a) *X. cheopis*

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq.</td>
<td>EAKWKFGKKLERMGKRVLHATEKVLPAAQGVAVGATAFRG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>cccceeeccccchhhhhhhhhhhhhhhhhhhheeeeeeeeec</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sequence length: 40

GOR4:
- Alpha helix (Hh): 19 is 47.50%
- Extended strand (Ee): 11 is 27.50%
- Random coil (Cc): 10 is 25.00%

b) *D. melanogaster*

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq.</td>
<td>EAGWLKKIGKKIERVGQHTRDATIQGLGVAQQAANVAATARG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>cccccccccccceeeccccchhhhhhhhhhhhhhhhhheeeneec</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sequence length: 42

GOR4:
- Alpha helix (Hh): 17 is 40.48%
- Extended strand (Ee): 8 is 19.05%
- Random coil (Cc): 17 is 40.48%

c) *M. sexta*

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq.</td>
<td>WNPFKELERAGQVRDAVISAAPAVATVQAAAIAARG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>ccccccchhhhhhhhhhhhhhhhheeeec</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sequence length: 37

GOR4:
- Alpha helix (Hh): 24 is 64.86%
- Extended strand (Ee): 3 is 8.11%
- Random coil (Cc): 10 is 27.03%
Fig. 18a. Secondary structure analysis of the *X. cheopis* cecropin mature peptide.

Fig 18b. Secondary structure analysis of the *D. melanogaster* cecropin mature peptide.

Fig 18c. Secondary structure analysis of the *M. sexta* cecropin mature peptide.
3.17. Identification of a Flea Peptidoglycan Recognition Protein

Contig 57 from the Finch assembly was shown to have amino acid sequence similarity to a group of peptidoglycan recognition proteins (PgRP's). The contig was 676 bases long both forward and reverse from the same clone with a 660 bp region at or above the Q20 value. A Sequencher analysis of the chromatograms was performed and the ambiguous bases were corrected based on the best read from the chromatograms to yield the consensus sequence (Fig. 18). The flea PgRP consensus sequence was used in a BLAST search using blastx and the best results were used in a ClustalW alignment. The sequences from the Drosophila family, Bombyx mori, Trichoplusia ni, and Homo sapiens were aligned to show areas of sequence similarity (Fig. 19). The BLAST sequences ranged in length from 182 to 203 amino acids in length while the sequence obtained from the flea was 226 amino acids long. The flea PgRP gene has a molecular weight of 21,889 daltons and a P.I. of 5.92. The best alignment was to the 203 amino acid long D. melanogaster PgRP protein which had an expect value of 3e\(^{-32}\) and 52% identity to a 187 amino acid region of the Drosophila PgRP. A dendrogram using the best BLAST alignments showed that the flea PgRP was most closely related to the D. melanogaster sequence (Fig. 20).
FLEA PgRP TRANSLATED SEQUENCE

10  20  30  40  50  60
GGGGATCTAAAATCATGATACGAAAAACTAGTATAATATTCTTTAATTTAATTTCT
GES K I M I R K T S I I L F L N L I S

70  80  90  100  110  120
TGCGCTTTAATGGAAAGAAATGCGAGAATTTGTGGCCAACCATTTGTAATTTCACTATGTT
CAL M E E M P D V C P N I V I K S M W

130  140  150  160  170  180
GGCGGACGCCTCTCCACGGAATTTGGAATAATATCGATATATCGTTCAAAT
G G R P S Q V E Y M I I P T K Y V I I

190  200  210  220  230  240
CAACATACAGTCATTTGATTTTGAATCTTCCAGATTGCTAAATATCGTTCAAAT
Q H T V T P S C D N L S D C A N I V S N

250  260  270  280  290  300
ATACAGATATCATATAGGGAATTTGGAATCTGACTAGATAGACTAATACTTCTAAT
I Q D Y H M R E L G T H D I G L N F L I

250  320  330  340  350  360
GGTAATGGCTACATTGTTATAGGGAATTTGGAATCTGACTAGATAGACTAATACTTCTAAT
G N D G N I Y E G V G W Y K I G A H T Y

370  380  390  400  410  420
GGCTACAATTCAAAGGCTATAGGGAATTTTATGGAACATTTGGAGACAAATTTGCCT
G Y N S K A I G I A F I G T F D K L P

430  440  450  460  470  480
TCTGAAAGGCCTTATATGCTGCTGCCCAGCTCCTGCTGCTGCGGTTGAGGACTCTGGAGCA
S E K A L N N A A Q A L L D C G V A L G A

490  500  510  520  530  540
CTTGATGACACATATTAAATTATTTTGCCAAGACAAAGTTGTTGCTACGGAAAGTCTCAGGG
L D D R Y K L F G Q R Q V V A T E S P G

550  560  570  580  590  600
CTTCGCCCTATACCAGGGAATATACAAATTTGCGCTCCTGCTGCTGCGGTTGAGGACTCTG
L R L Y Q E I Q N W P H W K N K P * S *

610  620  630  640  650  660
ACAGCTTGTGATTATGTATTATATATATACAGTACCTCGGCCCGCAACCAGCTAAGGCCG
T A C I L C I Y N T S T S A A T T L R A

AATT
N

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 19. *X. cheopis* PgRP-like translated sequence derived from cloning of a subtracted library.
## Peptidoglycan Recognition Protein Alignment

<table>
<thead>
<tr>
<th></th>
<th>Drosophila PBP</th>
<th>Bombyx mori PBP</th>
<th>Trichoplusia ni PBP</th>
<th>Homo sapiens PBP</th>
<th>X. cheopis PBP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Position: 20</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 30</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 40</strong></td>
<td>S R Q R S P A N P T K L R Q G E P T I G S E Q V</td>
<td>V V S - - - - - - - - - - K Q D I P V H A A A</td>
<td>V V T - - - - - - - - - - D E D D T P I H E E A</td>
<td>E D P - - - - A C S P V P N E K A A E C A O H S</td>
<td>L M E E M P D V P N V I S M G R P S Q E R M I</td>
</tr>
<tr>
<td><strong>Position: 50</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 60</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 70</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 80</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 90</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 100</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 110</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 120</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 130</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 140</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 150</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 160</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 170</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 180</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 190</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 200</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 210</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 220</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 230</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Position: 240</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 20. ClustalX alignment of PgRP sequences from various insects and *Homo sapiens* to the PgRP-like sequence of *X. cheopis*.
Drosophila PBP

Bombyx mori PBP

Trichoplusia ni PBP

Homo sapiens PBP

X. cheopis PBP

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 21. Dendrogram of alignment of PgRP from various insects and Homo sapiens to the PgRP-like sequence of X. cheopis.
3.18. Identification of Flea Serine Protease

Contig 110 from the Finch assembly was shown to have sequence similarity to a family of trypsin-like serine proteases. The flea sequence was compiled from four reads in the forward direction from four separate clones. The sequence lengths ranged from 612 bases to 702 bases in length. The four cDNA sequences were aligned in Sequencher and the ambiguous bases were called using the best peaks from the four chromatograms to yield the final consensus sequence (Fig. 21). The flea serine protease consensus sequence was BLAST searched using blastx and the best alignments were to a group of serine proteases ranging in length from 355 to 397 amino acids in length. The best alignment was to the *Anopheles gambiae* serine protease 14D2, however the alignment was made in 2 sections with different reading frames used on the flea cDNA sequence. From the flea, cDNA bases 29 to 496 were translated in the +2 frame and produced alignment values of $2 \times 10^{-26}$ and 40% amino acid identities to a 161 amino acid region of the *A. gambiae* serine protease. The flea sequence from base 499 to 555 was translated in the +1 frame and the expect value was $2 \times 10^{-26}$ with 57% amino acid identities to a 19 amino acid region of the *A. gambiae* serine protease. A conserved domain database search using RPS-BLAST was performed on the flea amino acid sequence and the results indicated that it has similarity to the
family of trypsin-like serine proteases with an expect value of $3e^{-34}$. The sequences from *D. melanogaster*, and the *A. gambiae* 14D and 14D2 serine proteases, were aligned to the flea serine protease to show areas of sequence similarity (Fig. 22). A dendrogram using the best BLAST alignments showed that the flea serine protease was most closely related to the *A. gambiae* 14D2 sequence (Fig. 23).
FLEA SERINE PROTEASE TRANSLATED SEQUENCE

10  20  30  40  50  60
AATTCGCCCTTAGGCTGCAGGCGGCCGAGGTACAAAATTATGTGCCGACCCTGTACAAGA
IRP*RGRGRGTKLCADPVQD

70  80  90  100  110  120
TAGAACCGTGAGCAAAATGGCTCCATACACAAAATCTACAAAGGGTACATTGTGCTGTG
RTVEQVITVPQYSKKKNVDI

130  140  150  160  170  180
CGCTTTAATTCGATTATCCACGCCCTGCTGGATCTCAGTTCAGTGCAGTCCAGTTTG
ALIRLSSPADLSSSERSVRPC

190  200  210  220  230  240
TTTGCCCTATATCAATGCCCCTGAGACTAGAACCTGCGGTCCGTGTGGTCGAC
LPISNALQTMNLGRTSALVVT

250  260  270  280  290  300
AGGTTGGGATACACAGAATGGGCGACTAGTCACCGATTTACTCAAGTGATGAGTACC
GWDYTEWGTSPDLKLKVGVVP

310  320  330  340  350  360
TGTAGTGGATAACCGCCATACATACATCTACGCTAGAAGGAAAATACAAATAATAGG
VVNDNAQCNVTYSRKTTIIIAK

370  380  390  400  410  420
ACAGTTATGCTGCTGGTGACAGTGCTACACGCAGCCTGCGGGCTGATCTGGAGGTCC
QLCAAGTTGGRGDRS(CGDSGQP)

430  440  450  460  470  480
TTTAAAGTACCAGGGAAATTATGCTGCTCAGACAGGCGAGGTACAGCGAGGCGATCTAC
LKVPNGYRGQTTRVVQGIVS

490  500  510  520  530  540
ATTGGACGATCAGCTTGGTGTCAGCCGAATGTCCAGGATGTCCTAGCTGCGAGTGGACAT
FGPVCAGCAEMYQECIREWDI

550  560  570  580  590  600
TGCAATGCTGATTAATATAAATATAAGAACCTGAGCAATTTTTAAAAACAGAAACATGAA
AWTGY*IIIEPEQF*NKNN

610  620  630  640  650  660
TTTACTGCAAAATTTTTAAAGCTTAATTTAAATATTGCTCATTAAATATACAAATTTAT
LLQLIKTMINCH**YNFI

CATACT

IL
Figure 22. *X. cheopis* serine protease-like translated sequence derived from cloning of a subtracted library. The two downstream active sites are in bold with the conserved residues underlined.
Figure 23. ClustalX alignment of serine protease sequences from various insects to the serine protease-like sequence of *X. cheopis*.
Figure 24. Dendrogram of alignment of type of serine protease from various insects to the partial serine protease-like sequence of *X. cheopis*. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
3.19. Partial Identification of Other Flea Immune Sequences

Additional sequences from the flea, compiled in Table 3, were found to have similarities to other immune response genes based on the BLAST search. Contig 50 from the Finch assembly aligned to a serine carboxypeptidase found in *Homo sapiens* that is believed to activate or inactivate bioactive proteins or peptides and may have a function in the inflammatory response (Mahoney et al., 2001). The flea amino acid sequence aligned to a region of 125 out of the 476 amino acids for the human protein which yielded a 53% protein identity for that region with an expect value of 4e\(^{-31}\). The flea sequence also aligned to a serine carboxypeptidase found in *A. aegypti* that is expressed in the insect’s fat body. The flea sequence aligned to a region of 103 out of a total of 471 amino acids which yielded a 53% identity with an expect value of 3e\(^{-29}\). A single read from the flea cDNA sequence aligned to the fat spondin gene from *D. melanogaster*. The fat spondin gene is highly expressed in the insect fat body and hemocytes and also plays a role in embryogenesis. The flea sequence aligned to a 191 amino acid region of the 763 amino acid protein with a 62% identity and an expect value of 4e\(^{-70}\). Several contigs aligned to sequences from the coleoptericin family of antibacterials that are active against Gram negative bacteria. The best alignments came from contig 55.
which aligned against the 71 amino acid antibacterial Acaloleptin with a 37% identity and an expect value of $8 \times 10^{-4}$, and contig 129 which aligned against the 71 amino acid antibacterial rhinocerosin with a 37% identity and an expect value of $8 \times 10^{-4}$. Flea contig 89 aligned to an attacin B1 precursor sequence from *D. melanogaster*, a glycine-rich antibacterial protein. The flea sequence aligned to a 77 amino acid region of the 218 amino acid protein with a 40% identity and an expect value of $4 \times 10^{-9}$. Contig 95 aligned to the attacin A sequence from *Glossina morsitans morsitans* in a 116 amino acid region from the 208 amino acid protein with a 38% identity and an expect value of $8 \times 10^{-10}$. Contig 102 aligned with sequences from diptercin B, a glycine-rich antibacterial peptide found in *Drosophila*. The best alignment was to the 76 amino acid diptercin B from *G. m. morsitans* where the flea sequence aligned to a 43 amino acid region with 46% identity and an expect value of $2 \times 10^{-6}$.
Table 3. Identification of best BLAST similarity to *X. cheopis* contigs with putative immune functions.

<table>
<thead>
<tr>
<th>Contig#</th>
<th>Protein/Peptide</th>
<th>Aligned to:</th>
<th>Identities</th>
<th>X. cheopis Sequence, Complete or Partial*</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>Serine carboxypeptidase</td>
<td><em>H. sapiens</em></td>
<td>53%</td>
<td>Partial</td>
</tr>
<tr>
<td>55</td>
<td>Rhinocerosin</td>
<td><em>O. rhinoceros</em></td>
<td>36%</td>
<td>Partial</td>
</tr>
<tr>
<td>57</td>
<td>PgRP</td>
<td><em>D. melanogaster</em></td>
<td>52%</td>
<td>Complete</td>
</tr>
<tr>
<td>95</td>
<td>Attacin B</td>
<td><em>D. melanogaster</em></td>
<td>41%</td>
<td>Partial</td>
</tr>
<tr>
<td>102</td>
<td>Diptericin B</td>
<td><em>D. melanogaster</em></td>
<td>43%</td>
<td>Complete</td>
</tr>
<tr>
<td>110</td>
<td>Serine protease</td>
<td><em>A. gambiae</em></td>
<td>40%</td>
<td>Partial</td>
</tr>
<tr>
<td>113</td>
<td>Cecropin A</td>
<td><em>D. melanogaster</em></td>
<td>39%</td>
<td>Complete</td>
</tr>
<tr>
<td>135</td>
<td>Insect defensin</td>
<td><em>O. rhinoceros</em></td>
<td>50%</td>
<td>Complete</td>
</tr>
<tr>
<td>138</td>
<td>Sarcotoxin II</td>
<td><em>S. peregrina</em></td>
<td>45%</td>
<td>Partial</td>
</tr>
</tbody>
</table>

* Indicates whether the *X. cheopis* amino acid sequence was aligned with the complete subject sequence to produce the percent identities listed.
The flea *X. cheopis* was chosen for characterization of its immune response partly due to the availability of large numbers of flea. This flea is also the vector for the plague bacillus, *Y. pestis* and there is some evidence that the bacteria may be resistant to cationic antibacterial peptides. This resistance may be due to exposure of the bacteria to the innate immune components of the flea. Understanding this putative resistance to cationic peptides and possibly to flea immune products may help in identifying agents that can circumvent the resistance. Based on gene and protein databank searches, antibacterial peptides and proteins have been isolated from a wide variety of insect orders including Lepidopterans, Dipterans, and Coleopterans, but not from members of the Order Siphonoptera. Given the success researchers have had in finding antibacterial peptides from a variety of arthropods it was likely that *X. cheopis* could mount at least an antibacterial response and possibly have a more complex response using a number of effector molecules. In *X. cheopis*, the naïve fleas produced no observable clear zone in the antibacterial bioassay even when high levels of protein from either whole flea lysates or hemolymph were applied. However, the inoculated fleas consistently produced at least a minimal
antibacterial response following inoculation or injection of bacteria but not after a sterile wounding of the insect. Previously several researchers had isolated individual antibacterial peptides from whole organisms using multiple rounds of SPE and RP-HPLC to separate whole insect lysates (Chalk et al., 1994). One type of antibacterial was isolated from *D. melanogaster* by grinding thousands of whole flies in buffer followed by an initial solid phase extraction based on hydrophobicity, followed by multiple rounds of RP-HPLC (Bulet et al., 1993). This bacterially-induced peptide was ultimately isolated in a single RP-HPLC fraction and identified by Edman N-terminal sequencing.

Initially, attempts to isolate and characterize a flea antibacterial protein/peptide was performed with whole fleas ground in an acidic buffer to enrich the supernatant with cationic peptides. Following the methods of Boman et al. (1981), bacterially inoculated whole flea lysate was cleaned and separated by centrifugation and hydrophobic interaction chromatography and each step was checked for maintenance of antibacterial activity with a bioassay and success of separation by SDS-PAGE. The results of separation based on hydrophobicity appeared to concentrate a group of proteins and peptides and maintain their grouping throughout the stages rather than further separate them after each successive step. Lee et al.
(1994) had included a molecular weight centrifugation separation step in their protocol to isolate an antibacterial from insect integument (Lee and Brey, 1994). Although the molecular weight centrifugation step did separate the bioactive fractions obtained following SPE as expected, the fractions did not separate any better in the RP-HPLC steps than the results seen previously using just the SPE procedure.

The results from separation of the various lysates and fractions of the lysates by RP-HPLC were similar. Separation of whole flea lysate by multiple rounds of RP-HPLC was ineffective in producing fractions with single or a small number of peaks and the PAGE gels had a very large number of bands. Separation by RP-HPLC of SPE fractions containing antibacterial activity yielded a less complex mixture but the gel results were still too complex to show the necessary separation for identification of induced bands appropriate for blotting and sequencing. It appeared that SPE followed by multiple rounds of RP-HPLC concentrated the fractions that contained not only the antibacterial but also a number of unwanted proteins and peptides and did not provide the separation necessary to remove these unwanted components from the mixture. The addition of a molecular weight cutoff centrifugation step did not improve the end result and there were still too many bands grouped in the antibacterial-containing fraction even after
three rounds of RP-HPLC. Also each fraction invariably contained a smear of peptides, especially in the lower portion of the gel, that tended to obscure the lowest bands.

The inability to properly separate the whole flea lysate may be due to the fact that they are blood-feeding insects. Although when collecting fleas for inoculation care was taken to avoid those with recent blood meals evident in the midgut, the hindgut invariably contained the remnants of the blood meal which became a part of the whole flea lysate. The whole insects other researchers have previously used to isolate antibacterials were not obligate blood feeders and this may have contributed to their ability to obtain a less complex mixture of proteins. This may have been a particular problem for the lower molecular weight peptides where the flea’s gut contents, likely containing a large amount of digested protein, may have contributed to the protein smear found at the bottom of the gels. Also, given the success in using ultracentrifugation for purifying the flea hemolymph used for 1D and 2D gel analysis, a similar step may have improved the results using whole flea lysates.

Other groups of researchers had used insect hemolymph as the starting material for the isolation of antibacterials (Hultmark et al., 1980). Bioassay results again indicated that naïve flea hemolymph contained no observable
antibacterial activity at the limit of detection for the assay even when high levels of protein were added to the assay wells. Hemolymph from fleas inoculated with a paste of *E. coli* and *M. luteus* produced various levels of antibacterial activity on a bioassay plate with either a lawn of *E. coli* or *M. luteus*. This indicated that these bacteria induced production of either an antibacterial capable of killing Gram-positive and Gram-negative bacteria or induced multiple specific antibacterials. Interestingly, hemolymph from fleas inoculated or injected with *E. coli* alone had antibacterial activity in bioassays with a lawn of *E. coli* and with a lawn of *M. luteus*. Hemolymph from fleas inoculated with *M. luteus* had antibacterial activity only in the bioassay using a lawn of *M. luteus*. These results are similar to those from *D. melanogaster* that produce antibacterials with specificities to Gram-positive or Gram-negative bacteria as well as an antifungal that is effective to some fungal organisms but not others (Lemaitre et al., 1997). This indicates that the immune response in fleas is similar in some respects to *D. melanogaster* in being able to produce a specific response, and possibly a unique antibacterial, against a Gram-positive bacteria. Interestingly, pure insect cecropin A did not produce a clear zone in a bioassay with a lawn of *M. luteus* but the flea hemolymph obtained following *E. coli* inoculation/injection did produce a clear zone. This indicates that the flea
hemolymph contains a component or components that act in a manner different from that of cecropin A derived from \textit{H. cecropia}.

The flea hemolymph proved to be a much simpler mixture of proteins and yielded evidence of the induction of several protein bands on 1D SDS-PAGE, although only at a relatively high molecular weight. The critical step in collection of hemolymph was ultracentrifugation, which yielded much sharper bands especially at the low molecular weights. This same method was used to produce hemolymph samples for 2D gel analysis with a pH 3-10 range dimension. A 2D-gel analysis followed by blotting of the protein spots to a membrane is a commonly used method for obtaining single proteins or peptides for Edman N-terminal amino acid sequencing. The 2D gels from naïve and infected hemolymph were run in tandem and compared using densitometry analysis. The densitometry data showed that there were several individual proteins that had been induced to a higher level following bacterial injection. Again the induced proteins were found only at the higher molecular weights. The induced proteins and several others were sent for Edman amino acid sequencing but only the uninduced internal control proteins, which were the largest individual spots observed on the gels following staining, produced a sequence. To obtain enough protein for blotting and sequencing the induced proteins it may be necessary to overload
the 2D gel and hopefully not lose the resolution necessary to be able to isolate induced proteins/peptides. Or, it may be possible to use a larger 2D gel first dimension strip capable of holding a greater amount of protein so that the induced spots may contain enough protein for sequencing. However, the lack of clearly induced spots at the lower molecular weights indicates that further experimentation on the hemolymph may be necessary to enrich it to the levels necessary to observe the induced antibacterials by 1D or 2D gel analysis.

LPS directly injected into the flea hemocoel would not induce an antibacterial response. The injection was tested using a concentration of up to 100 mg/ml LPS with a total injected volume of 0.1 ul so that an individual flea would have a maximum internal exposure of 10 ug of purified LPS. Injection of this amount of LPS did not increase the mortality of the fleas and had no effect on them based on observations and tests run for this project. Experiments performed for this research indicated that an injection of as few as 200 \textit{E. coli} does produce an antibacterial response. These results indicate that there is no immune mechanism capable of recognizing LPS and/or inducing a response to it although the flea immune system is capable of recognizing and responding to a whole bacterium. This differs from the
results from other insects that shows the induction of cecropin following LPS inoculation (Wittwer et al., 1997; Roos et al., 1998).

Inoculation of fleas with a bacterial paste produced inconsistent results when tested by bioassay. Inoculation with *E. coli* alone produced a response to both Gram-positive and Gram-negative bacteria so it was decided to focus on isolation of the immune response products from the inoculation and injection of *E. coli* only. Due to the possibility that inconsistent bioassay results were caused by the loss of bacteria from the surface of the pipette as it passed through the overlapping body plates of the flea, experiments were performed using the injection of a known number of *E. coli* suspended in buffer. Even with the obvious positive injection of bacteria the bioassay results were still somewhat inconsistent. Bioassay’s performed using hemolymph collected from individual fleas indicated that some fleas produced a robust response to injection while others produced no observable antibacterial response. These results were somewhat, but not totally, dependent on flea size and amount of hemolymph collected. It became apparent that given the limit to the amount of protein that could be run on a 2D gel, optimizing the process so that it was more likely to obtain proteins from those fleas capable of producing an immune response with a high specific activity was a necessary step. An experiment showed that
semi-gravid fleas were more likely to produce a larger amount of hemolymph with high specific activity than the generally smaller, non-gravid fleas. The optimization steps did not yield the anticipated for results from protein isolation. The 1D gels did show induced bands but they were higher molecular weight proteins and none in the sizes expected from the antibacterial peptides that were likely produced following bacterial injection. The 2D gels also showed a few induced proteins but they were again at the higher molecular weights. Perhaps it is necessary to further manipulate the hemolymph samples to isolate and concentrate the desired peptides in amounts sufficient for blotting and sequencing.

The first antibacterial peptide was isolated from the hemolymph of C. hyalophora larvae using the standard techniques of inoculation with bacterial paste, collection of hemolymph, and separation by protein hydrophobicity using SPE and multiple rounds of RP-HPLC (Steiner et al., 1981). This succeeded in part due to the large amount of hemolymph collected from each larva. Other insects that had antibacterials isolated from their hemolymph or total proteins sometimes used thousands of individually inoculated organisms to obtain the necessary amounts (Robertson et al., 1986: Chalk et al., 1994). More recently characterization of an insect immune response has been based on cDNA identification rather than protein
isolation (Bulet et al., 1999). Antibacterial peptides and other immune
effectors fall into families that share structural or sequence similarities such
as the conserved cysteine residues in defensins or the alpha helical structure
of the cecropins. A number of the antibacterials and a variety of immune
response genes from Drosophila spp. have come from cDNA analysis and
these discoveries have lead to the isolation and characterization of a variety
of immune genes responsible for the innate response in other animals
including humans (Cole et al., 2000). In part these results have been due to
whole genome sequencing allowing researchers to identify homologs to
known genes in the organism under study. Improvements in DNA
sequencing technology have also lead to the ability to easily sequence a
large number of genes from either a cDNA library or from a large number of
induced cDNA's isolated by subtractive hybridization. Induced genes from
members of insect Orders such as the Lepidopterans and Dipterans have
been identified due to their sequence similarities by using PCR techniques
on cDNA libraries (Kim et al., 1998; Sun et al., 1999). More recently
techniques like differential display, subtractive hybridization, and
differential hybridization have been used to identify induced genes in
organisms where less is known about the sequences. Given the dearth of
sequence information in the Order Siphonoptera in general and in X. cheopis
specifically, the subtractive hybridization technique seemed likely to identify components of the induced immune response.

The ability of the flea to produce an easily observable antibacterial bioassay response following bacterial challenge but no observable response from naïve fleas made it likely that the constitutive level of transcription was either very low or absent in the uninfected flea. The process of subtractive hybridization is designed to remove cDNA sequences common to the induced and the uninduced cDNA pools and enrich the final subtracted library with only induced sequences. Following bacterial inoculation or injection the flea was observed to produce an antibacterial response in as little as two hours indicating that production of mRNA transcripts is very rapid. Collection of induced and uninduced mRNA from X. cheopis was performed following a 5 hour incubation period. The hybridization reaction between the two cDNA pools yielded a small number of discrete bands within a light smear when viewed in an agarose/EtBr gel. This indicated that although there were still a number of possible common sequences in the final subtracted library there should still be at least several induced genes that would make up a large portion of the cloned products.

After cloning, the E. coli-induced versus the uninduced subtractive hybridization library produced 384 individual clones and all were
sequenced. The DNA sequencing yielded 768 individual single stranded sequences that were analyzed for multiple copies of the same gene that could be compiled into contiguous sequences or for single DNA sequences. A BLAST search identified 36 of the 156 contigs and 3 of the 92 single regions as aligning to known immune response genes although some were in multiple copies. Antibacterial sequences identified from the flea included genes with similarity to the defensins, cecropins, attacins, coleoptericins, apidaecin, and a diptericin B. One sequence had similarity to the family of peptidoglycan recognition proteins that have been implicated in the innate response in *Drosophila* spp. There were sequences from two serine proteases, a serine carboxypeptidase, and two serine protease inhibitors that may play a part in activation of the wound response or in activation of zymogens. Also a flea sequence was found with similarity to a fat spondin gene that is induced in fat body and hemocytes following bacterial challenge that may possibly play an as yet unknown role in the immune response of *Drosophila* spp. (De Gregorio et al., 2001). In addition to the putative immune response genes sequenced from the flea, there were a number of genes identified that were likely common to both the induced and uninduced libraries. The list includes sequences with similarity to insect troponin T (82% identity), cytochrome P450 (51% identity), actin (94% identity),
transferrin (59% identity), a 40S ribosomal protein (94% identity), and a 90kD heat shock protein (86% identity).

The BLAST search identified 14 clones with sequence similarity to the defensin family of antibacterial peptides. Defensin is a small cationic peptide ranging from 29 to 46 amino acids in length with a molecular mass of approximately 4 kD. The identifying features of defensin sequences are the six conserved cysteine residues that provide specific linkages in the mature peptide. The putative flea defensin identified by this research consisted of a prepropeptide sequence of 57 amino acids and a mature peptide length of 45 amino acids. A conserved domain database search using RPS-BLAST was performed and the putative flea defensin was identified as a member of the arthropod defensin family. The mature peptide contained the six conserved cysteine residues and the sequence aligned best to that from the beetle *A. dichotoma* with 57% identity. The sequence similarity between the prepro regions between the flea and other insects was poor but that is commonly the case for these regions in defensins (Martin et al. 1995).

There were three clones identified with sequence similarity to the antibacterial cecropin in the flea. Cecropin is also a small cationic antibacterial peptide with usually 62-64 amino acid residues and the mature peptide has a molecular weight of approximately 4 kD. It is formed as a
prepropeptide and the secreted propeptide contains either one or two
dipeptides that must be cleaved by a dipeptidyl aminopeptidase to produce
the active antibacterial. The prepropeptide form of the putative flea cecropin
is 61 residues in length and the mature peptide is 38 residues in length with a
2 residue pro-region. The putative flea cecropin sequence was best aligned
to the sequence from *Drosophila* spp. cecropin A1 with 39% identity value
to its sequence. A conserved domain database search using RPS-BLAST
was performed and the putative flea cecropin was identified as a member of
the cecropin family of proteins.

A single cloned sequence was putatively identified as a peptidoglycan
recognition protein (PgRP). This family of proteins is implicated in binding
to peptidoglycan following bacterial invasion and in activation of the pro-
phenoloxidase system in the silkworm *B. mori* (Yoshida et al., 1996). PgRP
has also been shown to bind strongly to Gram positive bacteria and less or
not at all to Gram negative bacteria in the moth *T. ni* (Kang et al., 1998).
These proteins have a molecular weight of approximately 19 kD and they
range from 182 to 203 residues in length. The flea PgRP sequence is
approximately 197 residues in length and this indicates that the sequence
obtained by this research is likely the complete sequence for a putative flea
PgRP. The conserved domain database search identified the flea sequence as
belonging to the family of N-acetylmuramyl-L-alanine amidases that includes the arthropod and mammalian PgRP’s. The putative flea PgRP sequence was best aligned to the PgRP sequence from *D. melanogaster* with a 52% identity value to its sequence.

A trypsin-like serine protease sequence was obtained from a single clone from the flea cDNA subtracted library and this sequence aligned most closely to the serine protease 14D2 from *A. gambiae* mosquitos (Gorman and Paskewitz, 2001). The mosquito serine protease has been implicated in initiating part of the innate immune response of either melanization or antibacterial synthesis (Gorman and Paskewitz, 2001) and serine proteases have been induced following bacterial infection in *Drosophila* spp. (De Gregorio et al., 2001). The flea sequence obtained is not a complete serine protease gene, but it does contain putative regions for two of the three catalytic sites. The conserved domain database search identified the flea sequence as belonging to a protein from the family of trypsin-like serine proteases and the sequence contains the two downstream catalytic site residues (histidine and serine) common to this family of proteins. The flea sequence from the +2 frame translation is approximately 166 residues in length that produced the best alignment to the mosquito and other insect serine proteases. The complete insect sequences for this type of serine protease remain to be determined.
protease used for the alignment ranged from 360 to 390 residues in length and the flea sequence aligns to within a few residues of the end of the last half of these insect sequences. Inexplicably, the last few residues of the flea sequence do not align to other insect sequences unless they were translated in a +1 frame at which point they align with a 57% identity. Analysis of the chromatograms from the contiguous sequences used for this alignment did not identify any bases that were improperly read and that should be changed to produce a complete sequence in the +2 frame that would align the flea sequence to the total last half of the insect proteins. It was determined that a single base substitution at base 510 in the DNA consensus sequence would allow the translation of the complete sequence to the +2 frame and would improve the alignment to other insect sequences. It will be necessary to resequence this area of the gene to determine if that base substitution is warranted or if the original sequence is correct.

Other antibacterial peptide and protein sequences have tentatively been identified from the subtracted flea cDNA library. Coleoptericins are linear peptides with a molecular weight of approximately 8 kD that are active against Gram-negative and Gram-positive bacteria. There were four flea contiguous sequences identified that aligned to coleoptericins and the best alignment was a 37% identity to the peptide acaloleptin isolated from
the beetle *Acalolepta luxuriosa*. There were nine contiguous sequences found that aligned to members of the attacin protein family. The attacins range in molecular weight from 20-24 kD and are active against Gram-negative bacteria. The best results from the flea aligned a 77 residue section with 40% identity to the 218 residue attacin B1 precursor. Diptericin B is an attacin-like peptide with a molecular weight of 8.9 kD found in *Drosophila* spp. A flea sequence was identified that aligned to diptericin B with 46% identity in a 43 amino acid section of the 76 amino acid peptide. Although the flea sequences identified with similarities to the above antibacterials are not conclusive it is possible that further research will identify flea genes for these peptides and a protein that play a part in the innate immune response. This would indicate that the flea has a multifacteted immune response to bacterial challenge possibly on par with that found in *Drosophila* spp.

In addition to genes implicated in the innate immune response, this research has identified several other complete genes or gene segments with sequence similarity to other insect genes. A flea sequence was found that aligned with 88% identity to a 174 amino acid region of the 216 residue GTPase RAN gene from *Drosophila* spp. Also a flea sequence was found that aligned with 94% identity to a 120 amino acid region of the 130 amino acid 40S ribosomal protein S15A from *Drosophila* spp. It is possible that
these flea sequences contain the entire gene sequences for these two genes and that they could be used to further characterize these two families of genes. The previously mentioned genes likely common to both cDNA libraries are large genes and the flea sequences are only parts of the total so that determining the entire sequence would require a number of additional sequencing steps.

Overall this research has shown that the flea, like many other insects, is capable of mounting a complex innate immune response when challenged by bacteria. They are capable of secreting antibacterial components into their hemolymph that have a bactericidal effect against both Gram-negative and Gram-positive bacteria. The result of gel analysis on the flea hemolymph indicates that bacterial inoculation of the flea can induce the production of several as yet unidentified proteins. The production of a subtracted cDNA library has identified a variety of genes that are putatively linked to the innate immune response in the flea by sequence similarity to other insect immune response genes. The complete sequences have been isolated for two putative flea antibacterials, a defensin and a cecropin. In addition the complete sequence was found for a putative flea peptidoglycan binding protein. PgRP has been implicated in immune recognition in other insects and mammals and has been shown to preferentially bind to Gram-positive.
bacteria. A partial sequence has been identified for a putative flea trypsin-like serine protease that had previously been linked to the immune response in mosquitos and Drosophila spp. Also, there is evidence that X. cheopis produces several other antibacterial peptides and an antibacterial protein, as well as a group of proteases and protease inhibitors, that may play a role in immune function in the flea.

When looked at as a whole it is likely that the flea innate response is capable of recognizing conserved bacterial patterns, inducing specific antibacterial responses, and activating antibacterials or protein cascades designed to eliminate microbial pathogens. Although a recognition protein (PgRP) was identified for Gram positive bacteria it is disappointing that a similar protein for Gram negative bacterial identification was not found. The results from plate bioassays indicated that inoculation of fleas with E. coli could induce a response that produced an antibacterial effective against both Gram negative and Gram positive bacteria. Inoculation of fleas with M. luteus produced an antibacterial that was only effective against Gram positive bacteria. Since the bacterially-induced cDNA library was produced using only E. coli it seems more likely that a response would have been produced against the LPS moiety of the bacteria rather than peptidoglycan. A Gram negative binding protein, or an LPS binding protein, as had been
shown in other insects could have mediated this response. However, the inability for purified LPS to produce any antibacterial response in the flea may indicate the lack of a flea LPS recognition molecule. Once the immune response has been activated there is evidence that the flea can produce antibacterials against different types of bacteria. Defensins are more effective against Gram-positive bacteria while cecropins are more effective against Gram-negative bacteria. The coleopterincin, attacin, and dipterincin B sequences, if proven to be valid, are all effective against bacteria. There were no sequences found that aligned to antifungal peptides. Given results from previous research using other arthropod species and the environment fleas live in it is likely that the flea mounts an antifungal response and that it will require a fungal challenge to induce it. Finally, the identification of a putative serine protease in the flea indicates that like in other insects the flea has mechanisms for control of immune activation and for induction of protein cascades similar to the prophenyloxidase cascade.

Future directions for this research include the sequencing of the complete genes for the flea 14D2-like serine protease and of the other partial antibacterial gene sequences. The X. cheopis subtracted cDNA library has given us a look at the innate antibacterial immune response and although there have been three complete genes identified, additional research could
confirm the presence of four more. Complete genes for proteases and inhibitors that could control some immune functions could also be identified from the flea. Additionally, a subtracted library approach using induction of the fleas by only a Gram-positive bacteria or a fungus, could yield different types of antibacterials or antifungals that could further elucidate the immune response in the flea.
References


of the antibacterial peptide cecropin A: a nuclear magnetic resonance and

Hultmark, D., A. Engstrom, K. Andersson, H. Steiner, H. Bennich and H. G.
Boman. 1983. Insect immunity. Attacins, a family of antibacterial proteins

Insect immunity: isolation and structure of cecropin D and four minor

immunity. Purification and properties of three inducible bactericidal proteins
from hemolymph of immunized pupae of Hyalophora cecropia. Eur J
Biochem 106:7-16.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Interactions of bacterial cationic peptide antibiotics with outer and
cytoplasmic membranes of *Pseudomonas aeruginosa*. Antimicrob Agents
Chemother *44*:3317-21.