2011

Proteomic Analysis of the Blood of *Alligator mississippiensis*

Lancia Nadinia Fallen Darville-Bowleg
Louisiana State University and Agricultural and Mechanical College, ldarvi1@lsu.edu

PROTEOMIC ANALYSIS OF THE BLOOD OF *ALLIGATOR MISSISSIPPIENSIS*

A Dissertation

Submitted to the Graduate Faculty of the
Louisianan State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Chemistry

by
Lancia Nadinia Fallen Darville-Bowleg
B.S., University of Nebraska at Kearney, 2003
M.S., University of Nebraska-Lincoln, 2005
December 2011

DEDICATION

This dissertation is dedicated to my grandmothers Fairleen Penn Lightbourne-Smith and Victoria Darville, who I wished could have been here to share this special moment that they had so proudly anticipated. I love you and you both will always be an inspiration.

ACKNOWLEDGEMENTS

First and foremost, I thank GOD the Almighty, who has given me the strength to persevere and see another goal accomplished. This dissertation would not have been possible without the love, support and encouragement of many people.

To my dearest husband Mario Bowleg, I am so blessed to have you in my life. Your love, patience and support throughout this process were invaluable and will always be treasured. I cherish the moments we share and the joy that you have brought to my life, especially in my darkest days. To my parents, especially my mother Christine Darville, who always stressed the importance of hard work, perseverance, and faith in GOD, I can't express my gratitude enough. Thank you for always believing in me and for your encouraging words and prayers when I needed them most. To my sister, Lanecia Darville and my brother, Lance Darville Jr. thank you for your constant love and support, you both are my biggest motivators.

To my research advisor, Dr. Kermit K. Murray, I thank you for your counsel, patience and belief in my abilities throughout my studies at Louisiana State University. I am forever grateful for the opportunity to have worked in your research group and for you allowing me to "spread my wings" and adventure into research that required outside collaboration. Your leadership has helped to shape me into the scientist I am today. I am greatly appreciative and I owe you my sincerest gratitude.

To my committee members, Dr. Kermit K. Murray, Dr. Doug Gilman, Dr. Bin Chen, Dr. Graca Vicente and Dr. Jeffrey Gillespie, I appreciate your valuable discussions, guidance and encouraging words. Special thanks to Dr. Mark Merchant, my collaborator. I appreciate the opportunity to have worked with you and for all of your guidance and mentoring during this process. To Dr. Azeem Hasan, thank you for allowing me to use the mass spectrometry facility

freely and for your gifts of wisdom that you willingly shared with me. I am forever in debt to your willingness to collaborate and share valuable scientific discussions with me.

To my colleagues in the Murray Research group, past and present, I greatly appreciate your friendship and I will always remember and cherish the great times we spent with each other. Finally, I want to thank all of my family and friends both near and far for all of their love and support and for making me feel at home whiles in Louisiana.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Name
2D-GE	Two-Dimensional Gel Electrophoresis
2D-PAGE	Two-Dimensional Polyacrylamide Gel Electrophoresis
ACN	Acetonitrile
AFM	Atomic Force Microscopy
AMP	Antimicrobial Peptides
ATCC	American Type Culture Collection
BLAST	Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
CCA	α-Cyano-4-Hydroxycinnamic Acid
CCD	Charge Coupled Device
CHAPS	3-[(3-cholamidopropyl) dimethylammonio]-1 propanesulfonate
CID	Collision Induced Dissociation
CNBr	Cyanogen Bromide
CRD	Carbohydrate-Recognition Domain
DHB	2,5-Dihydroxybenzoic Acid
DTT	Dithiothreitol
ECD	Electron-Capture Dissociation
EDTA	Ethylenediaminetetraacetate
Endo F	Endo-β-N-Acetylglucosaminidases F
Endo H	Endo-β-N-Acetylglucosaminidases H
ESI	Electrospray Ionization
ESI-Q-TOF	Electrospray Ionization-Quadrupole-Time-of-Flight

ESI-TOF	Electrospray Ionization Time-of-Flight
EST	Expressed Sequence Tag
ETD	Electron-Transfer Dissociation
FA	Formic Acid
Gal-GalNAc	Galactose-N-Acetylgalactosamine
GlcNAc	N-Acetylglucosamine
HPLC	High Performance Liquid Chromatography
HSP	High-Scoring Segment Pairs
HIV-1	Human Immunodeficiency Virus Type 1
IA	Iodacetaminde
ID	Inner Diameter
IDA	Information Dependent Acquisition
IEF	Isoelectric Focusing
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IM	Ion Mobility
IM-MS	Ion Mobility Mass Spectrometry
IMS	Ion Mobility Spectrometry
LC	Liquid Chromatography
LC-IMS	Liquid Chromatography-Ion Mobility Spectrometry
LC MS/MS	Liquid Chromatography Tandem Mass Spectrometry
LPS	Lipopolysaccharide
Man	Mannose
MALDI	Matrix Assisted Laser Desorption Ionization

MALDI-TOF-TOF	Matrix-Assisted Laser Desorption Ionization Tandem Time-of-Flight
MBL	Mannose-Binding Lectin
MCP	Microchannel Plate
MIC	Minimal Inhibitory Concentration
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MSDB	Mass Spectrometry Protein Sequence Data Base
NCBI	National Center for Biotechnology Information
NMR	Nuclear Magnetic Resonance
OMSSA	Open Mass Spectrometry Search Algorithm
PAGE	Polyacrylamide Gel Electrophoresis
PAS	Periodic Schiff
PDA	Photodiode Array
pI	Isoelectric Point
PMF	Peptide Mass Fingerprinting
PNGaseF	Peptide-N-(N-acetyl-β-glucosaminyl) asparagine amidase
Q/IMS/oa-TOF	Quadrupole/Ion Mobility Separator/Orthogonal Acceleration Time-of-Flight
QTOF	Quadrupole-Time-of-Flight
RPLC	Reversed-Phase Liquid Chromatography
RP-HPLC	Reversed-Phase High-Performance Liquid Chromatography
SA	Sinapininc Acid or 3,5-dimethoxy-4-hydroxycinnamic acid
SCX	Strong Cation-Exchange Chromatography
sDHB	2-hydroxy-5-methoxybenzoic acid

SDS	Sodium Dodecyl Sulfate
SEC	Size-Exclusion Chromatography
TC	T-cell
TCA	Trichloroacetic Acid
TFA	Trifluoroacetic Acid
TFMS	Trifluoromethanesulfonic Acid
TH	T Helper Cell
TOF/TOF	Tandem Time-of-Flight
TSB	Trypticase Soy Broth
TWIMS	Travelling Wave Ion Mobility Spectrometry
UPLC	Ultra Performance Liquid Chromatography

ABSTRACT

My research focus was to investigate alligator blood using mass spectrometry-based proteomics methods to understand their innate immune systems. The first goal was to sequence peptides and proteins from the blood serum and leukocytes using tandem mass spectrometry and *de novo* sequencing. The second goal was to determine the function of these biological molecules and their relationship to the immune system.

One- and two-dimensional gel electrophoresis was used to separate proteins from alligator leukocytes, which were enzymatically digested. The peptides were measured using reversed phase nano-high performance liquid chromatography coupled with tandem mass spectrometry (nano-HPLC-MS/MS) followed with *de novo* sequencing. The results, as described in Chapter 3 show that alligator leukocytes contain proteins that are similar to proteins found in other vertebrates such as mammals and reptiles that are related to immune responses. Isolation of small molecule interferences and peptides exhibiting antimicrobial activity from alligator leukocyte extracts are described in Chapters 4 and 5. Reversed-phase HPLC was used to separate the leukocyte mixture and antimicrobial activity tests were used to determine the active fractions. Interferants, EDTA and spermine were present and showed activity in early fractions. Two major peptides measured at 4.7 and 4.9 kDa in an active fraction were further separated on the basis of their charge, size and shape using ion mobility-mass spectrometry (IM-MS). Due to the limited fragmentation of the peptides using IM-MS, the peptides were isolated and fragmented using MALDI TOF/TOF MS for *de novo* sequencing.

Lectins are a class of carbohydrate selective proteins that are part of the complement immune system. Chapter 6 presents results for a lectin isolated from alligators that have mannan and mannose binding activity. In this study, the monomeric lectin was isolated and enzymatically digested using five different proteases to create small and large peptides which were analyzed by

LC-MS/MS. The peptides were determined via *de novo* sequencing and overlapped to generate the lectin sequence. Lectins may have varying degrees of glycosylation, therefore deglycosylation procedures suitable for mass spectrometry analysis are described in Chapter 7. Conclusions and future directions for the work in this dissertation will be summarized in Chapter 8.

CHAPTER 1. INTRODUCTION

Alligators exhibit strong immune responses when exposed to microorganisms, and are highly resistant to microbial infections.[1-3] Therefore, characterization of biological components that make up the immune system of alligators is of interest because these compounds may have therapeutic applications. Antibiotic resistance has become a global public health issue and is increasing at a rate that exceeds the pace of the development of new drugs. Hence, isolation of antimicrobial peptides from alligators may lead to a new class of antibiotics.

Another class of molecules that are part of the immune system is the set of proteins known as lectins, which show promise for their application in cancer therapy, pharmacology and immunology.[5] The study of lectins isolated from *Alligator mississippiensis*, can provide insight on its structure-function relationship within the crocodile's immune system.

The goal of this research was to use mass spectrometry-based proteomics to isolate and characterize biological molecules in alligator blood and identify specific components that may have medicinal use.

1.1 Biochemistry of the Immune System

The vertebrate immune system is composed of a network of circulating cells and molecules[7] that recognize and respond to the invasion of foreign pathogens such as viruses, bacteria, fungi and protists.[8] These molecules include antimicrobial molecules such as lysozymes and defensins, molecules involved in specific recognition of foreign antigens such as immunoglobulins and T-cell receptors, molecules that carry intracellular signals such as cytokines, and molecules that receive and transduce these signals such as cytokine receptors.[8] Vertebrates have both innate and adaptive immune systems that work together.[7] The innate immune system responds rapidly and is the first line of defense against pathogens and does not

require previous exposure to exhibit a full immunological response.[7] In contrast to the innate immune system, the adaptive immune system requires previous exposure to an antigen to exhibit a full immunological response and can take several days to weeks to become completely active.[7]

Most studies on vertebrate immune systems have been performed on mammals[8] and the evolutionary development of the immune systems of major group of vertebrates, including mammals, birds, amphibians and fish is well known compared to other groups of organisms.[8] In contrast, little is known about the reptilian immune systems.[8] A better understanding of reptilian immunology will provide important information on the evolution of both innate and adaptive immune mechanisms. Crocodilians are of interest because they are the only living reptilian archosaurs, which include the dinosaurs, pterosaurs, and crocodilians.[9]

Crocodilians are vertebrates with a complex immune system. The components that make up the reptilian immune system are antimicrobial peptides, macrophages, heterophils, neutrophils, basophils, eosinophils, phagocytic B cells, and proteins of the complement system. Like other vertebrates, crocodilians also have an innate and adaptive immune system.[7, 10] Crocodilians thrive in microbe containing environments but exhibit a strong resistance to infections.

Alligator serum has been shown to have antibacterial,[3] antiviral,[2] and antiamoebacidal properties.[11] Antibacterial activity has also been observed in crocodile serum, particularly *Crocodylus siamensis*.[12, 13] In addition, the leukocytes of *Alligator mississippiensis* have also shown to produce a broad antimicrobial spectrum.[1] The alligator complement system, which is part of the innate immune system, has also been shown to be effective against gram-positive bacteria.[3, 14] When alligator serum was compared to human serum, the alligator serum was effective against different strains of Gram-positive bacteria, unlike human serum which had no antibacterial activity. It was proposed that the complement is also responsible for antiviral

activity from the alligator serum. Human T-cells were infected with human immunodeficiency virus type 1 (HIV-1) and when incubated with alligator serum potent antiviral activity was observed.[2] These studies suggest that crocodilians have a strong innate immune system.

1.1.1 Innate Immune System

The innate immune system comprises different molecules and cells including lysozymes, proteins of the complement system, non-specific leukocytes, and antimicrobial peptides.[7] Lysozymes are enzymes that cause bacteria cells to be lysed by the hydrolysis of their cell wall.[15] Lysozymes have been isolated from several reptilian organisms such as, lizards,[16] turtles,[16] crocodiles,[17] and alligators[6]. The complement system consists of various proteins found in plasma that kill bacteria via lysis or opsonization.[7] Lysis entails the complement proteins rupturing the bacterial membrane therefore killing the invading bacteria.[18] Opsonization is the process by which opsonin proteins found in blood serum, bind to the bacterial membrane allowing the bacteria to be recognized by macrophages. The macrophages then engulf the bacteria through phagocytosis.[7]

There are three different pathways to the complement immune system: classical, alternative, and lectin.[19] The classical pathway is activated by the immunoglobulins, immunoglobulin G (IgG) and immunoglobulin M (IgM) to activate an immune response.[7, 19] The alternative pathway does not require antibodies but is activated by molecules such as viruses or lipopolysaccharide (LPS) that are found on the surface of bacteria.[18] Finally, the lectin pathway is activated by mannose sugars of proteins that are on the surface of bacteria.[19] The complement immune system has been characterized in the American alligator[14] and is believed to be responsible for antiviral activity exhibited by alligator serum.[2]

There are non-specific leukocytes in reptiles including eosinophils, herterophils, basophils, monocytes, and macrophages.[7] Limited information is known about the function of

eosinophils in reptiles; however, in mammals they play a key role in the defense against parasitic infections.[19] Heterophils are involved in the inflammatory response in reptiles and are also responsible for suppressing microbial invasion.[20] Basophils contain immunoglobulins on their surface and when triggered by an antigen releases histamine.[21] Monocytes and macrophages are phagocytic cells that are responsible for processing and releasing antigens as well as releasing cytokines.[19] Cytokines are regulatory proteins that are released by the cells of the immune system to generate an immune response.[22]

An important component of the innate immune system is antimicrobial peptides and proteins. Antimicrobial peptides are found in the host defense system and exhibit antimicrobial activity.[23] Antimicrobial molecules are typically amphipathic and cationic small peptides less than10 kDa in mass.[24] However, they can also be anionic peptides[25] or proteins.[26] Antimicrobial peptides can be linear and α-helical or cysteine containing and β-sheet. The linear peptides typically consist of 12–25 residues and no cysteines.[24] The β-sheet peptides have several antiparallel β-strands and are stabilized with up to six disulfide bonds.[24] There are also antimicrobial peptides that are rich in specific residues, such as tryptophan,[27] proline, and/or arginine[28, 29] and histidine.[30]

There are two major antimicrobial peptide families found in vertebrates: defensins and cathelicidins. Defensins are antimicrobial peptides that are rich in arginine residues and have a characteristic β-sheet fold and six disulfide linked cysteines.[23] They are cationic peptides that bind to microbes via electrostatic interactions.[31] Defensins have been found in mammals and birds[32] and recently the first reptilian defensin was discovered in the European pond turtle *Emys orbicularis*.[33] Defensins are 38-42 residues and found in cells and tissues involved in the host defense system. In many animals the highest concentration of defensins are found in the

granules, where the leukocytes are stored.[23] Defensins exhibit antibacterial, antifungal, and antiviral activity.[23]

Another family of antimicrobial peptides is the cathelicidins. Cathelicidins are linear molecules that range in size from 12 to 80 amino acids and, unlike the β-defensins, they lack disulfide bridges.[34] Cathelicidins have antifungal[35] and antibacterial activity.[36] They have been isolated from fish,[37] birds,[38] mammals,[39] and reptiles.[40] Cathelicidins are produced in the myeloid cells in the bone marrow and stored in the neutrophil granules.[41] Cathelicidins have also been found in monocytes, epithelial cells of the skin, respiratory tract, urogenital tract, as well as T and B lymphocytes.[41]

The membranes of microbes and multicellular animals differ in that microbes, particularly bacteria, have an outer membrane surface that is composed of lipids with negatively charged phospholipid head groups. The outer membranes of plants and animals are composed of lipids that do not have a net negative charge (Figure 1-1); the negatively charged head groups are positioned towards the cytoplasm.[42, 43]

The primary model that explains the mechanism of antimicrobial peptides is the Shai-Matsuzaki-Huang model.[43, 44] The Shai-Matsuzaki-Huang Model postulates that the peptide interacts with the pathogen membrane, displacing the lipids and disrupting the membrane structure. In some cases the peptide may also penetrate the cell.[44] The membrane of multicellular animal cells contains cholesterol, which reduces the activity of the antimicrobial peptide via peptide interaction with the cholesterol or stabilization of the lipid bilayer. In addition to this proposed mechanism, studies suggest that there are other mechanisms involved. These mechanisms include 1) depolarization of the bacterial membrane leading to death,[45] 2) creation of holes in the cell causing cellular leakage,[46] 3) activation of processes that cause cell death (e.g. hydrolases which destroys the cell wall),[47] 4) disruption of the membrane function by

rearrangement of the lipids on the outer cell membrane,[43] and 5) destruction of the internal cellular components after internalization of the antimicrobial peptide.[48]

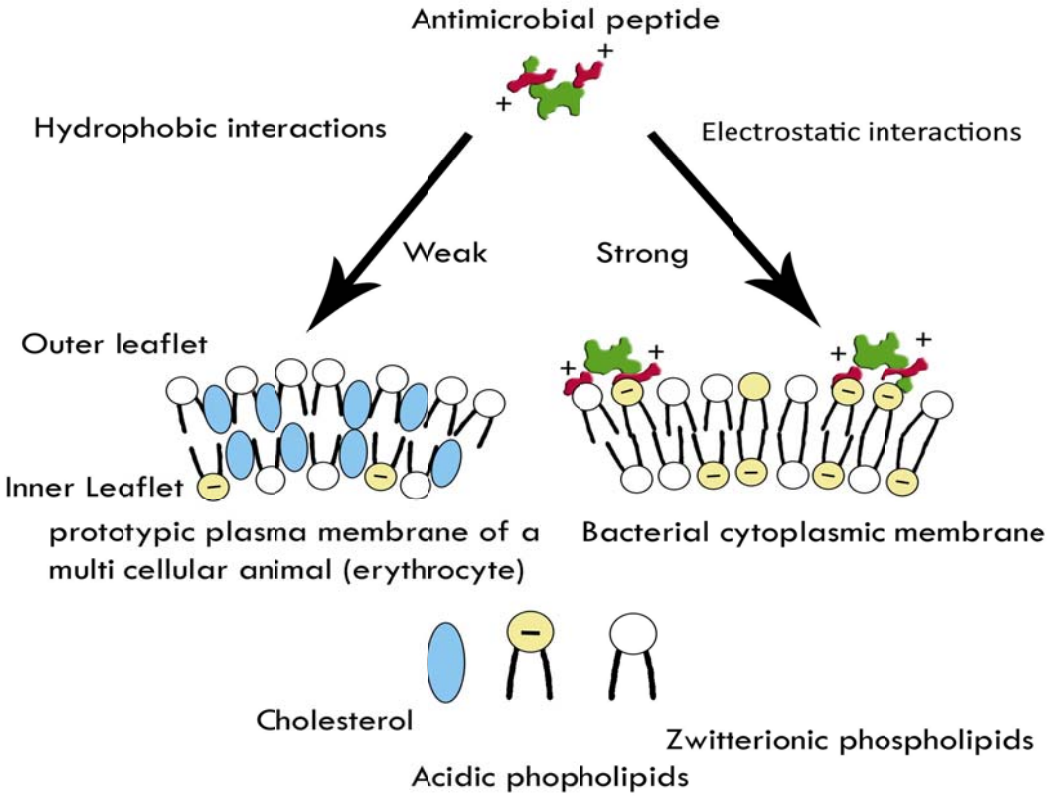


Figure 1-1. Comparison of antimicrobial peptide interaction with animal and bacterial membranes.

Some antimicrobial peptides have been tested for pharmaceutical use, primarily as a topical treatment. The first antimicrobial peptide to undergo clinical trial was the magainin peptide isolated from frog skin.[49] The magainin peptide was developed as a topical treatment and progressed to Phase III clinical trials to treat diabetic patients suffering from foot ulcers.[49, 50] The study showed the magainin peptide to be equally as effective when compared to the oral antibiotic ofloxacin.[51] A variety of antimicrobial peptides are currently being developed as potential antibiotics as indicated in Table 1-1.

The broad antimicrobial spectrum of the antimicrobial peptides allows them to be used for diverse applications. However, a major drawback of antimicrobial peptides for clinical use is the level of toxicity. The level at which the antimicrobial peptides are effective *in vivo* are usually toxic.[52] Other factors are stability and immunogenicity.[49]

Table 1-1. Antimicrobial peptides available for therapeutic use.[49, 53]

Peptide	Pharmaceutical Name	Origin	Mode of Application	Application	Stage
Magainin 2	Pexiganan	African clawed frog skin	Topical	Foot ulcers	Completed Phase III. (Not approved by FDA)
Indolicidin	Omiganan (MBI-226)	Synthetic analog of indolicidin	Topical	Catheter infection	Phase III
Indolicidin	MBI-594AN	Cow erythrocytes	Topical	Acne	Phase III
Protegrin	Iseganan (IB-367)	Pig leukocytes	Oral	Mucositis	Phase III
Histatin	P113P113D	Human	Oral	Gingivitis	Phase II
Heliomycin	Heliomycin	Tobacco budworm	Systemic	Antifungal	Preclinical
Lactoferricin	Lactoferricin	Human	Systemic	Antibacterial	Preclinical
Bactericidal permeability increasing protein histatin	XMP.629	Human	Systemic	Meningococcal meningitis	Phase III

Another group of molecules that play an important role in vertebrate innate immune systems is lectins. Lectins have been studied for more than 100 years in plants and have recently been recognized for their importance in animals.[54] Lectins are proteins that can recognize carbohydrates that are endogenous to the animal or those on microbes.[55] They have been

classified based on their carbohydrate ligands, biological processes, subcellular location, and dependence on divalent cations.[55] They have a wide range of structures, but carbohydrate-binding activity is limited to a particular region of the protein known as the carbohydrate-recognition domain (CRD).[56]

Animal lectins are grouped into two structural families: C-type and S-type.[56] C-type lectins are found in serum, extracellular matrix, and plasma membranes. They require calcium for binding and bind to a variety of sugars. S-type lectins are found intracellularly and require sulfydryl for binding to β-galactosides.[57]

Animal lectins possess a wide variety of functions such as cellular growth regulation, mediation of endocytosis, intracellular routing of glycoconjugates, and urate transport.[58] However, the major function of lectins is to behave as recognition molecules within the immune system.[58] These roles within the immune system include: direct defense, cell recognition and trafficking, immune regulation, and prevention of autoimmunity.[58] Different applications of lectins have been demonstrated such as immobilization onto columns for affinity separation of glycoproteins, glycopeptides and oligosaccharides,[59] selective agglutination,[60] and blood typing.[61] Lectins show promise for their application in agriculture, cancer therapy, pharmacology, and immunology.[5]

1.1.2 Adaptive Immune System

The adaptive immune system becomes activated following the innate immune response. There are two adaptive immunity responses: cell-mediated and humoral adaptive immunity.[7] Cell-mediated immunity involves T-cells that are responsible for regulating antibody production. T-cells can differentiate into two types of cells: cytotoxic T-cell (TC) or T helper cell (TH). TC can rapidly kill bacterial or viral infected cells through apotosis. TC regulates other immune cells. TH helps control other immune cells. T-cells also release cytokines that affect the humoral

response (the immune responses mediated by antibodies).[7] Humoral adaptive immunity relies on B-cells that recognize antigens even before being processed.[19]

1.1.3 Study of the Immune System

To study the immune system of vertebrates, serological assays are typically used, in which *in vitro* reactions between antigen and serum antibodies are studied.[62] There are many different types of serological assays including agglutination, precipitation, immunoassays, immunofluorescence, fluorescence-activated cell sorting analysis, and lymphocyte function.[62] These assays help elucidate the immune system's functional and regulatory properties through lymphocyte function measurements and the responses of B- and T-cells as well as antibodies.

Another approach to studying the immune system is hematology, which is the study of blood, including white blood cells, red blood cells, hemoglobin, and platelets.[62] Hematological tests allow for observation of live blood (whole blood that is unaltered and unstained) with microscopy. Using this test, microbial activity in the blood and its potential effects can be measured. In addition, the white blood cells can be quantified, which can provide insight on how the immune system is functioning. There is also the standard blood microscopy technique in which the blood is stained and fixed; however, staining kills the blood cells.

Another technique for studying the immune system is genomics and proteomics, which is discussed in Section 1.2. Genomics and proteomics provide a more comprehensive view of the immune system and its function as compared to serological assays and hematology tests.

1.2 Chemical Analysis Methods for Proteomics

Proteomics is the approach used to understand complex biological systems by analyzing protein expression, function, modifications, and interactions.[63] Proteomics is related to genomics, which is the study of the genetic make-up of an organism.[64] An important step towards understanding an organism's biology is to determine its genome sequence. However, the genome

sequence is not enough to provide information on complex cellular processes; the complement of proteins associated with a particular genome is essential to this understanding.[65]

Proteomics is complimentary to genomics and provides an additional component to the understanding of biological systems. However, there are significant challenges in proteomics such as limited sample quantity, sample degradation, broad dynamic range (>10^6 fold for protein abundance), post-translational modifications, and disease changes.[66] Protein concentrations typically exceed the dynamic range of a single analytical instrument or method necessitating the use of one or more dimensions of separation.[63]

Proteomics can be divided into three major branches: structural proteomics,[67] expression proteomics,[68] and functional proteomics.[68] Structural proteomics involves determining the structures of proteins, such as its three-dimensional shape (secondary and tertiary structure) and its amino acid sequence (primary structure).[67] A commonly used approach for protein sequencing is Edman degradation. Edman degradation was developed by Pehr Edman in 1950 and is one of the oldest and most developed techniques for protein sequencing.[69] However, Edman sequencing chemistry has largely been replaced by mass spectrometry for protein sequencing and identification.[70, 71] Protein secondary and tertiary structures can be characterized by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy.[67]

Expression proteomics involves the quantitative and qualitative analysis of proteins under different conditions.[68] This approach allows disease-specific proteins to be identified by comparing the entire proteome between two samples. Proteins that are over-expressed or under-expressed can be identified and characterized. The techniques commonly used for expression proteomics are two-dimensional gel electrophoresis,[72] multi-dimensional chromatography with mass spectrometry,[73] and micro-array techniques.[74] Two-dimensional electrophoresis suffers from limitations such as the large dynamic range of protein expression in biological systems and

difficulty in analyzing proteins that are post-translationally modified.[68] Limitations of microarray technology include the sensitivity of the arrays for detection of low abundance genes[75] as well as its inability to measure post-translational modifications.[76] Unlike 2-D electrophoresis and micro-array techniques, multi-dimensional chromatography with mass spectrometry can measure post-translational modifications in expression proteomics.[77]

Functional proteomics is an approach to analyze and understand the properties of macromolecular networks involved in cells.[68] Proteins and their specific roles in metabolic activities can be identified.

1.2.1 Sample Purification

Blood plasma and serum are commonly used biological fluids for proteomic analysis because many cells release a portion of their content into the plasma when damaged or upon cell death. There are approximately 10,000 proteins present in human serum[78] and many proteins of interest are present in low abundance. Plasma comprises ~97% high abundance proteins including albumin (57–71%) and immunoglobulins (8–26%).[79] Hence plasma samples are difficult to analyze directly and purification is required. In addition, proteomic samples usually contain a large amount of contaminants such as lipids, nucleic acids, and surfactants.[80] The dynamic range of expressed proteins is greater than six orders of magnitude[81] and protein mixtures can be quite complex and contain proteins with different solubility, hydrophobicity and hydrophilicity, pI, and molecular masses.[82] Therefore, sample purification is necessary before mass spectrometry analysis.

Separation methods such as affinity-based techniques, chromatography and centrifugation have been employed.[83] It is important that the proteins of interest are well resolved with limited sample purification steps to avoid sample loss. Affinity-based techniques have been established for removal of albumin and IgG using immobilized antibodies that are selective

against albumin[84] and protein A or protein G that selectively capture IgG on the columns.[85] The removal of salts and other contaminants from protein samples has been accomplished using precipitation with centrifugation.[80] Commonly used precipitation methods include acetone, trichloroacetic acid (TCA), ammonium sulfate, and chloroform/methanol precipitation.

1.2.2 Digestion

Protein digestion can be performed using three common approaches: in-gel,[86, 87] in-solution,[88] or solid phase.[89, 90] In an in-gel digestion, the proteins are separated on a 1- or 2-D gel and the gel bands are excised for chemical or proteolytic digestion.[87] A major advantage of in-gel digestion is that it removes detergents and salts that can be interfants in the mass spectrometer.[87] However, a limitation to this method is the loss of peptides during in-gel digestion through binding to the polyacrylamide.[87] Another approach that has been developed is in-solution digestion which entails digesting proteins directly in buffers or solvents such as ammonium bicarbonate or acetonitrile.[91] This approach is advantageous in that low abundance molecules that may otherwise be lost in the gel can be detected. A drawback for this method includes longer incubation times due to lower enzymatic concentrations.[92] Solid phase is another digestion approach which includes immobilization of an endoprotease on a solid support; examples include monolithic columns for trypsin digestion[93] and microfluidic devises with integrated trypsin digestion.[90] Solid phase digestion offers advantages of increased digestion rates, reduced interferences from trypsin autolysis products, low sample consumption, and fast response.[94, 95] A limitation to this method is the use of organic solvents which improve digestion efficiency but can damage the immobilized enzyme.[96]

Digestion efficiency can be improved by cysteine reduction before digestion. The disulfide bonds are reduced with a reagent such as dithiothreitol (DTT) and an alkylation reaction is performed with iodoacetamide to prevent new disulfide bridges from forming.[97]

Many of the endoproteases and chemicals cleave proteins at specific amino acids generating peptide fragments of varying lengths. Peptide fragments between 6–20 amino acids are best for MS analysis and protein database searching.[98] Endoproteases and chemicals used for protein analysis are indicated in Table 1-2. The most commonly used endoprotease for proteomic analysis is trypsin. Trypsin cleaves at lysine and arginine residues, unless followed by a proline residue in the C-terminus direction.[99] Trypsin has good activity in both in-gel and in-solution digests.

Table 1-2. Chemicals and proteases used for enzymatic and chemical cleavage. Cleavage with the endoproteases only occurs if the residue after the cleavage site is not proline, except for Asp-N.[100]

Endoproteases	**Cleavage Specificity**
Trypsin	K, R
Glu-C	E, D
Lys-C	K
Asp-N	D
Arg-C	R
Chymotrypsin	W, Y, F, L, M
Chemical Agents	**Cleavage Specificity**
70% Formic acid	D
Cyanogen bromide	M
2-nitro-5-thiocyanobenzoate, pH 9[101]	C
Hydroxylamine, pH 9[102]	N, G
Iodobenzoic acid	W

Many proteins contain a significant number of lysine and arginine residues that are spaced sufficiently in the sequence so that trypsin produces fragments that are a suitable length for MS analysis. Another complementary enzyme used is Glu-C which cleaves at the carboxyl

side of glutamate residues.[103, 104] In the presence of selected buffers such as sodium phosphate, it can cleave at both the glutamate and aspartate residues. Other proteases listed in Table 2 with cleavage specificities are useful for producing peptides of varying lengths depending on how many cleavages occur. This is useful for obtaining additional sequence information, especially if the protein is unknown.

There are also non-specific endoproteases such as pepsin as well as endoproteases with broad specificities such as chymotrypsin that are useful for producing multiple overlapping peptides that can increase sequence coverage.[98] Proteins can also be cleaved with cyanogen bromide, formic acid, and hydroxylamine. Cyanogen bromide is the most commonly used for protein cleavage; it cleaves specifically at methionine residues.[105]

1.2.3 Separations

Two separation methods used in proteomics are gel electrophoresis and liquid chromatography (LC). Gel electrophoresis is a sensitive method for separating and identifying proteins in a gel matrix such as agarose or polyacrylamide.[80] Agarose is typically used to separate larger macromolecules such as nucleic acids and polyacrylamide is typically used to separate proteins. Polyacrylamide gel electrophoresis can be used to determine the size, isoelectric point and purity of proteins.[80] The gel pores are made by crosslinking of the polyacrylamide with bis-acrylamide to form a network of pores that allows the molecules to move through the gel matrix like a sieve. The gel pore size is determined by the acrylamide monomer concentration.[106]

Gel electrophoresis separates molecules based on the differences in migration velocity of ions in the gel under the influence of an electric field. The migration velocity is the product of the electrophoretic mobility and the applied electric field. The electrophoretic mobility is proportional to the ion charge and inversely proportional to the frictional forces. The frictional

forces depend on the analyte size and the viscosity of the solvent. Smaller analytes have a greater mobility and migrate farther down the medium in a given time. Polyacrylamide gel electrophoresis (PAGE) is used to separate proteins and peptides based on their size. Sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) is the most commonly used gel based technique for separating proteins. SDS is used to denature the proteins and gives the protein an overall net negative charge.[107]

Two-dimensional gel electrophoresis (2D-GE) is used to separate proteins based on their charge and mass.[107, 108] The first dimension separates proteins based on their net charge using isoelectric focusing.[108] Protein separation is performed in a pH gradient; the proteins migrate to their isoelectric point (pI), which is the pH where the protein has a net charge of zero. A protein's pI is determined by the type and number of acidic and basic residues it contains.[107, 108] The second dimension in 2D-GE separates proteins based on their mass and is usually performed in a SDS gel (SDS-PAGE). 2D-GE has difficulty resolving large proteins, or those with extreme pI or hydrophobicity, and suffers from lack of reproducibility.[109]

Liquid chromatography (LC) is a technique used to separate components on a stationary phase using a liquid mobile phase. Reversed-phase high-performance liquid chromatography (RP-HPLC) separates proteins and peptides by hydrophobicity.[110] It is one of the most powerful and commonly used liquid chromatography techniques.[111] Commonly used hydrocarbon ligands for reversed-phase resins include C_4 and C_{18}.[110] C_4 is commonly used for polar proteins and C_{18} is most used for peptides.

Ultra performance liquid chromatography (UPLC) uses smaller particles as well as high speed and peak capacity (the number of peaks that can be resolved per unit time).[112] Compared to conventional HPLC columns which are packed with 3.5 to 5 μm particles, UPLC columns are

packed with 1.7 μm particles.[113] Smaller particles shorten the analyte's diffusion path which improves separation efficiency, speed, and resolution.[114, 115]

Ion exchange is another form of chromatography which involves the separation of proteins and peptides based on their charge.[110] A cationic or anionic resin is used and proteins or peptides of opposite charge are retained due to charge attraction. Hydrophilic-interaction chromatography separates proteins based on their hydrophilic properties; hence the stationary phase is polar.[116] Another separation technique is affinity chromatography, which separates proteins and peptides based on their specific ligand-binding affinity.[110] There are two fractions collected from affinity separation, the unbound and the bound proteins and peptides. The analyses and detection of low abundant proteins, primarily in plasma, can be difficult due to the presence of high abundant proteins such as albumin, immunoglobulins, and transferrin.[117] Therefore, affinity-based approaches can be used to remove high abundant proteins or low abundant proteins can be enriched.[109]

Due to the complexity of the protein samples, one-dimensional separation techniques are usually insufficient and multi-dimensional separations are employed. In multi-dimensional separation, two or more separation techniques are coupled together to improve the resolving power. These separation techniques are orthogonal to one another.

1.2.4 Mass Spectrometry

Mass spectrometry (MS) is an analytical technique used for measuring the mass and chemical structure of molecules and is widely used for proteome analysis. A variety of ionization techniques can be used for mass spectrometry but the most commonly used techniques for the analysis of biomolecules are electrospray ionization (ESI)[118] and matrix assisted laser desorption ionization (MALDI).[119-121] MALDI uses a matrix that absorbs laser energy and aids in ionization of the analyte. The ions generated are typically singly charged. ESI can also be used to analyze

biomolecules. Unlike MALDI, its ions are produced from solution. After the ions are formed they are transferred into a mass analyzer by an electric field where they are separated according to their mass-to-charge ratio. Two stages of mass separation can be coupled (either in space or in time) to obtain additional information of the sample being analyzed which is known as tandem mass spectrometry (MS/MS).[122] Tandem mass spectrometry is used to determine peptide sequences from proteins.[83] A peptide is separated from a mixture of peptides in the first stage of mass spectrometry and dissociated by collision with an inert gas. The generated fragments are separated in the second stage of mass spectrometry.[83, 123]

There are three MS proteome analysis approaches: 1) bottom-up proteomics, 2) shotgun proteomics, and 3) top-down proteomics. In the bottom-up approach the protein mixture is separated usually by 1 or 2-dimensional electrophoresis and the individual protein bands or spots are cut and digested with an enzyme such as trypsin resulting in peptides. The peptides are analyzed by mass spectrometry using peptide mass fingerprinting or tandem mass spectrometry (MS/MS) to create sequence tags for database searching.[124-126] Some of the major advantages of using the bottom-up approach are the ability to obtain high-resolution separations and a comprehensive coverage of proteins. It's the most widely used technique in proteomics,[127] hence several bioinformatics tools are available. In addition, proteins can be separated in a complex mixture before digestion so there is a greater chance of identification. The drawback of this approach is the limited dynamic range[128] and difficulty separating membrane proteins.[129, 130]

In shotgun proteomics, a mixture of intact proteins is enzymatically digested and separated using strong cation-exchange chromatography (SCX) followed by reversed-phase liquid chromatography (RPLC).[109, 131] The separated peptides are subjected to tandem mass spectrometry and database searching.[131] A major advantage of this technique is that thousands of proteins can be identified in a single analysis and the technique is better suited to membrane

proteins. However, limitations of this technique includes the need for complex mixtures to be purified prior to separation,[132] limited dynamic range,[63] and bioinformatics challenges to identify peptide and protein sequences from a large number of acquired spectra.[63]

In the top-down approach, intact proteins are separated by gel electrophoresis or HPLC before being introduced into the mass spectrometer.[124, 126] The mass of the protein is measured and tandem mass spectrometry is used to generate sequence tags (a short sub-sequence of a peptide sequence) for database searching. Alternately, *de novo* sequencing (an approach to determining a peptide sequence without prior knowledge of the sequence) can be performed.[126] Top-down sequencing can be used to locate and characterize post-translational modifications, determine the complete protein sequence, and minimize time-consuming preparation steps such as separation and digestion procedures. Conversely, spectra generated by multiply charged proteins can be very complex and bioinformatics tools still needs more developing for protein identification. In addition, top-down sequencing does not work well with intact proteins greater than 50 kDa.

Separation can also be performed in the gas phase using ion mobility mass spectrometry (IM-MS).[133-135] Ion mobility is a gas-phase technique that separates ions based on their ability to migrate through a gas in the presence of an electric field.[133, 134] The mobility is dependent on the collision cross section of the ion. An ion mobility spectrometer consists of a gas filled cell where ions travel under the influence of an electric field.[133] Ions with larger diameters undergo more collisions with the buffer gas, hence their passage through the drift cell is slower, whereas smaller molecules undergo fewer collisions and pass through the drift cell more rapidly.[133] When coupled with MS both the mass to charge and size to charge ratio of the ions can be determined. Both MALDI and ESI can be used as an ionization source for IM-MS.[136] However, many IM studies reported in literature of peptides and proteins describe the use of an ESI source coupled

with IM.[137-139] IM can also be coupled with tandem MS for additional peptide and protein information.[140]

IM-MS separations can obtain comparable resolutions to HPLC and CE.[134] LC-MS, a more commonly used proteomics technique is advantageous in that it increases the dynamic range,[141] however, coupling LC to MS limits optimization of LC and LC can take minutes to hours for separation which limits sample throughput.[142] IM offers two major advantages over LC techniques; it reduces separation time and the post-ionization separation provides information on the product.[134] Another major advantage of IM-MS is that it adds two separation dimensions for different chemical classes,[143, 144] for example peptides, DNA, oligonucleotide and lipids or conformational classes such as α-helix and random coil.[145] However, IM-MS has its shortcomings for proteomic applications, specifically poor sensitivity and limited peak capacity.[134] Despite the limitations of IM-MS for proteomics it has still been proven to be a useful method for proteomics,[134] metabolomics,[146] and glycomics.[147]

1.2.5 Protein Bioinformatics

Bioinformatics is the approach used to analyze large numbers of genes and proteins[148] and is important for the elaboration of mass spectrometry data due to the large amount of data produced.[149] It is used in proteomics to provide functional analysis and mining of data sets.[148] Peptide and protein data can be interpreted via peptide mass fingerprinting, database searching, or *de novo* sequencing.

Peptide mass fingerprinting (PMF) is a high-throughput protein identification method.[150] In this method the protein is purified and cleaved using an enzymatic or chemical approach to generate peptides. The peptides are analyzed via ESI or MALDI mass spectrometry and a peptide mass fingerprint (the masses of the intact peptides in the sample) is obtained. The mass fingerprint is compared to theoretical cleavages of protein sequences in databases and the protein

matches are scored based on the best match.[150, 151] Several programs have been developed for peptide mass fingerprinting including MassSearch, Mowse, MS-FIT, PepMAPPER, PepSea, PeptideSearch, ProFound and PeptIdent.[152] Peptide mass fingerprinting will only yield hits for proteins that are in a sequence database.

An alternate approach to database searching is mass spectral matching. This entails matching the experimental spectrum to a library of previously obtained MS/MS data.[153] This method is a fast and precise means to identifying peptides whose proteome has been previously identified. Its major limitation is its inability to be used for identifying or discovering new peptides.

Another approach for database searching compares experimental spectra to theoretical spectra to identify peptides in the protein database. Theoretical tandem mass spectra are produced from fragmentation patterns that are known for a specific series of amino acids. Search engines that are used for database searching includes Mascot,[154], SEQUEST,[155] X!TANDEM, [156] Open mass spectrometry search algorithm (OMSSA),[157] SONAR,[158] ProbID,[159] PeptideProphet[160] and OLAV-PMF.[161] Two commonly used search engines include Mascot and SEQUEST. SEQUEST uses a cross-correlation score to match hypothetical spectra to experimental spectra[155] whereas Mascot uses a probability score that indicates the probability of whether or not a spectral match was random.[154] When some commonly used search engines were compared, including SEQUEST and Mascot, Mascot proved to be able to better discriminate between a correct and incorrect hit as compared to SEQUEST.[162] An overall evaluation showed that Mascot outperformed the other algorithms used in the study, which included PeptideProphet, Spectrum Mill, SONAR and X!TANDEM.

Mascot is based on probability scoring and the lowest probability is the best match. The match significance criteria depend on the size of the database. The score is reported as -10log (P), where P is the probability. Hence, the best match has the highest score.

Database searching offers several advantages including high-throughput, robustness and well annotated proteins (detailed information on each protein). Despite these advantages there are also some disadvantages that are associated with database searching including false positive identification due to selection of background peaks, unidentified peptides due to post translational modifications, scoring a longer peptide that may be from a lower quality MS/MS spectrum (low signal-to-noise ratio) with a higher score than a shorter peptide from a higher quality MS/MS spectrum (high signal-to-noise ratio), and, most importantly, it is impossible to identify a peptide that is not part of a protein in the database.[163]

De Novo sequencing is an approach to identifying peptides without database searching, for example for a species whose genome has not been previously sequenced. It is also used to identify post-translational modifications. The *de novo* approach determines peptide sequences using information such as the type of fragmentation method, for example collision induced dissociation (CID),[164] electron-transfer dissociation (ETD),[165] or electron-capture dissociation (ECD),[164, 166] the type of enzyme used, as well as any chemical modifications. Some commonly used *de novo* sequencing programs are PEAKS[167] Mascot Distiller[97], Lutefisk,[168] PepNovo,[169] and SHERENGA.[170] Tandem mass spectrometry data can also be searched against expressed sequence tag (EST) databases to identify peptides and proteins for organisms without complete genomes. ESTs are nucleotide sequences (200 to 500 nucleotides long) that are generated by sequencing either one or both ends of an expressed gene originating from specific tissues.[171] These nucleotide sequences are translated into protein sequences for protein identification from tandem mass spectra.

Under CID, peptides fragment along the peptide backbone and fragment ions generated from the N-terminus of the peptide are labeled a, b, and c, whereas fragments generated from the C-terminus of the peptide are labeled x, y, and z.[172] *De novo* spectra generated from low energy CID gives only partial peptide ion coverage because of its backbone cleavage specificity; in low energy CID spectra c, x, z and a-type fragment ions are not observed.[172] Hence, it is usually beneficial to collect peptide spectra from other fragmentation methods such as ETD or ECD. Some of the limitations associated with CID include overlapping fragment ion peaks (which can cause incorrect peak assignment), low signal for some of the ions in the CID spectra, difficulty identifying post-translational modifications, and the inability to differentiate between the amino acids leucine and isoleucine.[163] The ETD and ECD techniques can be used which can differentiate between leucine and isoleucine as well as identify post-translational modifications. ECD also produces less specific backbone cleavage as compared to CID; therefore, more extensive sequence information can be obtained on proteins.[164] However, a limitation of ETD and ECD is their inability to produce good quality data with shorter peptides, such as those generated from tryptic digests.[163]

Basic local alignment search tool (BLAST) is a search algorithm that is used to compare sequence similarities between experimentally determined nucleotide or protein sequences with nucleotide or protein databases.[173] This approach is useful for the identification of proteins from organisms that have unsequenced genomes.[174, 175] A BLAST alignment pairs each amino acid in the queried sequence to those in another sequence from a protein database. BLAST begins a search by indexing short character strings (amino acid sequences) within the peptide query by their starting position in the query. The "word size" (length of the amino acid sequence) for a protein-to-protein sequence comparison is typically three. The BLAST software then searches the database to look for matches between the indexed "words" from the queried peptide to

character strings within the sequence in the database. Whenever a word match is found, BLAST then extends the sequence (using the database sequence) in the forward and backward direction to create an alignment. The BLAST score value increases as long as the alignment matches and will begin to decrease once it encounters mismatches.[176, 177]

The BLAST results are quantified by comparing them to the expect value (E-value). The E-value threshold represents the number of times a good match is expected to occur by chance and is proportional to the size of the database. BLAST determined E-values that are greater than the threshold E-value are considered significant. The higher the similarity between the queried sequence and the sequence in the database the lower the E-value is. This can be seen in Equation 1-1,

$$E = K * m * n * e^{-\lambda S} \qquad \textbf{EQUATION 1-1}$$

where K is a constant (scaling factor), m is the length of the query sequence, n is the length of the database sequence, λ is the decay constant from the extreme value distribution (scales for the specific scoring matrix used) and S is the similarity score.[176]

1.3 Research Objectives

The objective of this research is to characterize the biological molecules isolated from the *Alligator mississippiensis* immune system and identify antimicrobial molecules. To achieve this goal, the specific aims of the project were (1) to isolate and characterize the proteins of the leukocytes, (2) to purify and identify antimicrobial peptides from whole blood, and (3) to isolate and determine the primary sequence of lectin from the serum. The first set of experiments, described in Chapter 3, used gel electrophoresis and multidimensional gel electrophoresis to separate a mixture of isolated leukocyte proteins for mass spectrometry detection and sequence identification. A second set of experiments, described in Chapters 4 and 5, used a combination of

separation techniques for purification of a novel antimicrobial peptide. Although antimicrobial peptides have been identified and characterized in numerous organisms, antimicrobial peptides have not yet been identified in alligators. For the third set of experiments described in Chapter 6, a new lectin from *Alligator mississisippiensis* was identified using *de novo* sequencing with different enzymatic digests and characterized. Chapter 7 describes deglycosylation approaches and compares sample clean-up procedures that can be applied to glycoproteins for mass spectrometry analysis.

CHAPTER 2. EXPERIMENTAL*

This chapter contains a description of the instruments and methods used for analyzing peptides and proteins isolated from the blood of alligators. A summary of the instruments and techniques used for this research, including gel-based protein separations, liquid chromatography, ion mobility, mass spectrometry, and proteomic bioinformatics is described. A detailed description of the experimental parameters and procedures used for the separation techniques as well as the electrospray ionization time-of-flight (ESI-TOF), matrix-assisted laser desorption ionization tandem time-of-flight (MALDI-TOF/TOF) and electrospray ionization-quadrupole-time-of-flight (ESI-QTOF) mass spectrometers is given.

2.1 Blood and Leukocyte Collection

Adult alligators were captured at night by boat at J. D. Murphree State Wildlife Refuge in Port Arthur, Texas, with the use of a spotlight and a cable harness and housed at the Rockefeller State Wildlife Refuge in Grand Chenier, Louisiana. The animals were kept in fiberglass-lined concrete tanks approximately 5 m long and 2 m wide, with 50% dry bottom and 50% water of approximate 0.2 m depth. The temperature was maintained at 31 ºC. Several alligators up to 1.5 m in length were housed in a single tank. The alligators were fed formulated dry pellets four times per week and the cages were cleaned five times per week.

Blood samples were collected from the supravertebral branch of the internal jugular vein using a heparinized 38 mm long 18 gauge needle and a 60 mL syringe[178, 179] and transferred to 250 mL plasma bottles containing 25 mL of 500 mM EDTA. This method of blood collection has been approved by the McNeese State University Animal Care and Use Committee and is routinely used to collect blood from wild and captive adult alligators (See Appendix D).

*Portions reprinted by permission from the Elsevier. The work reported in this chapter has been published in *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*.[6]

Leukocytes were isolated from the whole blood by differential sedimentation. Whole blood samples were mixed with one volume of 5% dextran and the erythrocytes were allowed to settle for approximately two hours. The leukocytes were separated manually from the top phase using a transfer pipette.[1]

The leukocytes were centrifuged for 15 min. at 1500 *g* and the leukocyte pellet was resuspended in approximately one volume of 10% acetic acid and stored at -80ºC until needed. Prior to analysis, the leukocyte pellets were homogenized in a Dounce homogenizer and then centrifuged at 20,000 *g* for 30 min. The clear supernatant was transferred to 1 kDa concentrator tubes (Centriprep, Millipore Corp., Billerica, MA) and centrifuged at 5000 *g* to concentrate the sample and remove the acetic acid. The concentrated extract and precipitate were resuspended in 10 mL of 0.1% acetic acid. The resuspended extract was centrifuged at 20,000 *g* and the clear supernatant was used in the gel electrophoresis and HPLC analysis.

2.2 Liquid Chromatography

In this work, reversed-phase liquid chromatography was used. Both capillary and nano capillary high performance liquid chromatography (HPLC) as well as nano ultra performance liquid chromatography (UPLC) were used.

2.2.1 Nano HPLC

Nano capillary HPLC (nanoLC) uses columns that have an inner diameter (ID) between 50 and 100 µm and flow rates below approximately 120 nL/min.[180] In this work, nanoLC was coupled to the mass spectrometer described in Section 2.6.4. Typical columns used for nanoLC have lengths of 50-150 mm with 3-5 µm particle diameters. For improved chromatographic separation, acids, including trifluoroacetic acid (TFA) and formic acid (FA), are added to the solvents at 0.1%. TFA is best for offline LC because it provides higher peak capacity with narrower peak width. However, if the LC is coupled online with the mass spectrometer, FA is

preferred because the high surface tension of the TFA solution makes spray formation difficult and trifluoroacetate can form strong ion pair complexes with peptide cationic sites causing signal suppression.[181]

The LC separations were performed on a LC Packings/Dionex capillary and nano HPLC system. The HPLC system is equipped with a low pressure mixing system, micropump, flow splitting unit that allows flow rates from 50 nL/min to 200 µL/min, a sample injector, and UV detector. A Vydac C_{18} 75 µm ID × 150 mm, 5.0 µm column was used. The solvents were water in 0.1% FA and acetonitrile in 0.1% FA. The column flow rate was 200 nL/min and a linear gradient was used for separation.

2.2.2 Capillary HPLC

Capillary HPLC is separation on a column that has an ID between 100 and 300 µm and flow rate that is several µL/min. A capillary HPLC (Agilent 1200 series) was used both online with the Agilent mass spectrometer (see Section 2.6.3 below) and offline with a photodiode array (PDA) detector over a wavelength range between 190 and 400 nm. The capillary HPLC is equipped with a binary and gradient pump, diode array detector, autosampler and thermostat. In this work, a C_{18} 2.1 mm ID × 50 mm, 5.0 µm column (Zorbax) and a C_{18} 4.6 mm ID × 250 mm, 5.0 µm column (Vydac) were used. The solvents were water in 0.1% FA and acetonitrile in 0.1% FA. The column flow rate was 300 µL/min and a linear gradient was used for separation.

2.2.3 Nano Ultra Performance Liquid Chromatography

Nano ultra performance liquid chromatography (nano UPLC) uses columns that typically have an ID between 75 and 1000 µm and flow rates between 200 nL/min and 100 µL/min. A UPLC system (Waters nanoAcquity) was used online with the Waters Synapt instrument described in Section 2.6.5. The system is equipped with a binary solvent manager, sample manager, a column heating compartment and a UV detector. In this study, an C_{18} column

(Acquity) packed with 1.7 µm particles, 5 cm long × 2.1 mm ID was used. The samples were diluted in 0.1% FA and eluted using a gradient with 3–85% 0.1% FA in acetonitrile (v/v) for 32 minutes. The column flow rate was 500 nL/min and 214 nm UV absorption was recorded.

2.3 Ion Mobility

Another separation technique used in this work was ion mobility (IM). The ion mobility cell was built into a Q-TOF mass spectrometer (Waters Corporation), details of this mass spectrometer are found in Section 2.6.2. The peptide samples were diluted in 0.1% FA and injected on the UPLC system at a flow rate of 500 nL/min. The peptides were separated using a gradient of 3–40% B over 30 minutes. Solvent A was 95% water and 5% acetonitrile containing 0.1% FA. Solvent B was 80% acetonitrile and 20% water containing 0.1% FA. The peptides were ionized using a nano-electrospray source and analyzed via a quadrupole mass analyzer before being introduced to travelling wave ion mobility cell. In travelling wave ion mobility spectrometry (TWIMS) pulsed voltage waves guide the ions through the drift cell via stacked ring electrodes.[182] As the waves move, the ions are pushed for a short distance before moving over the wave. Smaller ions undergo fewer collisions with the buffer gas so travel faster in the traveling wave's gradient, whereas larger ions undergo more collisions with the buffer gas slowing them down therefore requiring a longer time to be moved by the traveling waves.[140, 182] A pressure of 0.5 bar of nitrogen was maintained in the IM cell. Each set of experiments was run in triplicate.

2.4 Gel Electrophoresis

Both one- and two-dimensional gel electrophoresis was used to separate protein mixtures prior to digestion and mass spectrometry analysis. One-dimensional gel electrophoresis was performed using a small format cell (BioRad mini-PROTEAN). The main components of this gel apparatus are a gel cassette assembly, electrode assembly, clamping frame, mini tank and a lid

(see Figure 2-1). The electrode assembly can hold gels 8 cm (W) × 7.3 cm (H). In this work, precast polyacrylamide gels were used (BioRad Laboratories).

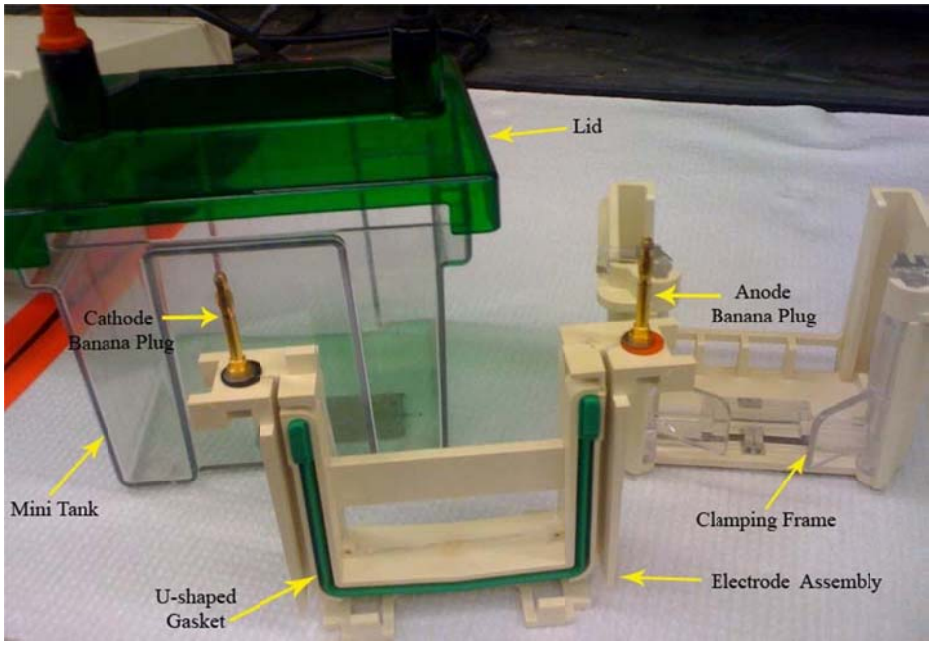


Figure 2-1. Assembly components of the mini-PROTEAN 2D gel electrophoresis cell.

Two-dimensional electrophoresis was performed using either the small format or a standard format electrophoresis system (Protean II XL, Bio-Rad) (see Figure 2-2). For isoelectric focusing, sample was dissolved in sample buffer containing 8 M urea, 2% 3-[(3-cholamidopropyl) dimethylammonio]-1 propanesulfonate (CHAPS), 50 mM dithiothreitol (DTT) and 1% carrier ampholytes. Isoelectric focusing was carried out in an isoelectric focusing cell (Protean, Bio-Rad) at 250 V for 20 min and 14,000 V-h (rapid ramping). SDS-PAGE electrophoresis was performed in slab gels (18.3×19.3 cm) with 1.00 mm thickness. The electrophoresis was conducted at 200 V for 40 min.

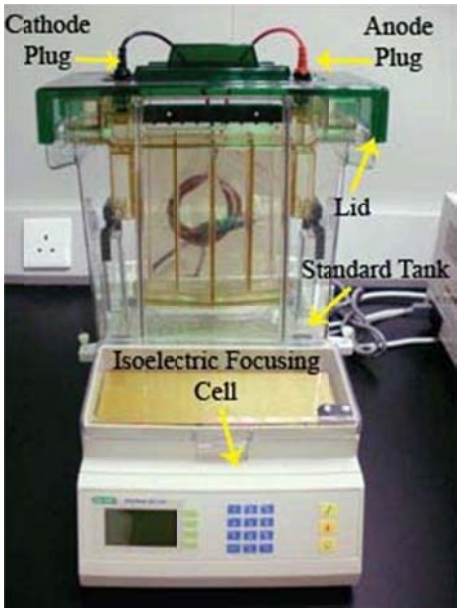


Figure 2-2. Protean II XL large format gel electrophoresis system and isoelectric focusing cell.

Gels can be prepared as a single continuous percentage or as a gradient %T through the gel; examples of commonly used gradient gels include 4–15% and 4–20%. Gradient gels have the advantage that proteins with a wide molecular weight range can be separated and the protein bands are narrower. In this work, 4–20% gels were used unless otherwise indicated.

Another component to the gel electrophoresis is the buffer system. There are two types of buffer systems: continuous and discontinuous.[183] A continuous buffer system has the same buffer at a constant pH and concentration. The gel used is typically made of one continuous %T. A discontinuous buffer system has a different buffer in the gel solution and the solution for the electrode assembly, and a stacking gel is also added to the resolving gel. The stacking gel has a lower acrylamide concentration and pH, which gives the gel larger pores allowing the proteins in the sample to be concentrated before entering the resolving gel. The resolving gel has a higher acrylamide concentration and higher pH which gives the gel narrow pores. This gives better protein resolution and separation.

The sample buffer system used in this work was a discontinuous buffer system that incorporated sodium dodecyl sulphate (SDS) and a reducing agent, β-mercaptoethanol, in the buffer. SDS is an anionic detergent that denatures and binds to proteins. The polypeptides become rod-like in shape with a negative charge.[184] Because all the proteins have the same charge, they separate according to their molecular mass. The β-mercaptoethanol is used to reduce the disulfide bridges in the proteins so that they can assume the random-coil configuration required for optimum separation.[106]

To resolve proteins with similar molecular weights, two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) was performed. 2D-PAGE is a powerful proteomics separation tool as it is capable of separating thousands of proteins in a typical cell.[185] The first dimension, isoelectric focusing (IEF) was performed in an isoelectric focusing cell (Bio-Rad Protean IEF). IEF is usually performed in an immobilized pH gradient strip.[186] The pH gradients used in this work were 3–10 and 3–6. Prior to the second dimension, the proteins separated by IEF underwent an equilibration step. The proteins were reduced and alkylated using DTT and iodoacetamide and SDS was added to give the proteins a negative charge. This allowed the separation based on size in the second dimension. The second dimension was performed using the Bio-Rad Protean II 2-D electrophoresis cell.

The gels were stained to visualize the proteins. Several stains are commercially available: Coomassie blue R-250, Bio-safe Coomassie, Sypro ruby fluorescent stain, and silver stain.[106, 185] The choice of stain used depends on several factors such as sensitivity, ease of use, and type of imaging equipment available. Coomassie R-250 stain (wool dye with a red hue) is the least sensitive: it requires 40 ng of protein for detection.[185] Bio-safe Coomasie stain is prepared with Coomassie G-250 (wool dye with a green hue) and can be used to detect as little as 10 ng of protein.[185] The most sensitive stains are Sypro ruby fluorescent and silver stains which both can

be used to detect as little as 1 ng of protein.[185] In this work, Bio-safe Coomasie or silver stains were used. Both stains are mass spectrometry compatible, meaning that the stains have high sensitivity for protein detection while minimizing crosslinking of the protein to the gel, which can inhibit recovery for mass spectrometry.[187]

There are also specific stains that can be used for detecting modified proteins such as glycoproteins. The most commonly used methods for detecting glycoproteins in polyacrylamide gels involves detection by periodic Schiff (PAS) staining using chromogenic detection, for example acid fuchsin dye.[188] Another approach utilizes Pro-Q Emerald 300 fluorescent dye, that links to glycoproteins using periodic Schiff conjugation.[189] During the staining process the glycols in the glycoproteins are oxidized to aldehydes with periodic acid. The dye then reacts with the aldehydes on the glycoproteins and a fluorescent conjugate is generated.[189] This stain can be used to detect as little as 0.5 ng of glycoprotein and is compatible with mass spectrometry analysis. This glycoprotein stain was used in the initial characterization of lectin proteins studied in this work.

Gel images were captured using a camera (Gel Doc XR System with Quantity One 1-D analysis software, BioRad) which is capable of imaging Coomassie stained, silver stained, and fluorescent gels.

2.5 Edman Sequencing

Edman sequencing involves labeling the N-terminal residue of a protein or peptide with phenyl isothiocyanate. The peptide bond adjacent to the modified residue is cleaved via mild acid hydrolysis, leaving the rest of the protein intact. The terminal amino acid residue that has been modified is identified by liquid chromatography. The process is repeated successively for the residues in the peptide sequence until a complete sequence is determined.[69] Protein bands were excised from SDS polyacrylamide gels and submitted to the University of Texas Medical

Branch, Protein Chemistry Core laboratory for N-terminal sequencing using a protein sequencer (LC 494, Applied Biosystems, Foster City, California).

2.6 Proteolysis of Isolated Proteins

Proteolysis generates peptides from proteins using either enzymatic or chemical cleavage. Proteolysis experiments were performed for peptide sequencing. Digestions were performed both in-gel and in-solution. Prior to proteolytic digestions, reduction and alkylation was performed to reduce the disulfide bonds for more efficient digestion. For reduction and alkylation a reaction buffer containing 25 mM NH_4HCO_3 was mixed with 10 mM dithiothreitol (DTT). To a vial containing approximately 100 µg of lyophilized protein in 25 mM NH_4HCO_3, 10 µL of DTT solution was added and the sample mixture was heated for one hour at 57ºC. The vial was removed from the heat block and allowed to cool to room temperature and 40 µL of 55 mM of iodacetaminde (IA) solution was added and incubated at room temperature in the dark for 30 minutes.

Several enzymatic and chemical proteolytic agents were used: trypsin, α-chymotrypsin, Glu-C, Asp-N, Lys-C, cyanogen bromide (CNBr), and formic acid. The enzymatic digestions were performed in a heat block at 37ºC, the CNBr digestions were performed at room temperature in a fume hood, and the FA digestions were performed at 37ºC.

For the enzymatic digestions, a 0.2 µg/uL solution of trypsin was prepared in 1 mM HCl and stored for a maximum of four weeks at -20ºC. The reaction solution contained 1 µl of trypsin (0.2 µg), 9 µL of 40 mM NH_4HCO_3 and 9% acetonitrile by volume. For enzymatic digestions using Lys-C, Asp-N and Glu-C, the preparations were as follows: 0.1 µg/µL of Lys-C in a 1:20 ratio (v/v), 0.04 µg/µL of Asp-N in a 1:50 ratio and 0.1 µg/µL of Glu-C in a 1:20 ratio. For CNBr digestions a 1 M solution was prepared in acetonitrile (ACN) and stored at -20ºC. The CNBr solution contained 5 µL of 5% TFA/50% FA and 5 µL of water (v/v).

2.7 Mass Spectrometry

In this work, four commercial mass spectrometers were used: an ESI quadrupole-time-of-flight (QTOF), a MALDI tandem time-of-flight (TOF/TOF), an ESI time-of-flight, and a quadrupole/ion mobility separator/orthogonal acceleration time-of-flight (Q/IMS/oa-TOF) mass spectrometer

2.7.1 MALDI TOF/TOF

The MALDI TOF/TOF mass spectrometer used in this work was a Bruker UltrafleXtreme. A diagram of this instrument is shown in Figure 2-3. The Bruker TOF/TOF mass spectrometer is equipped with a Nd:YAG laser (wavelength, 355 nm; repetition rate 1 kHz) with beam homogenization (Smartbeam). It is also equipped with delayed extraction, a collision cell, linear and reflectron flight tubes.[190]

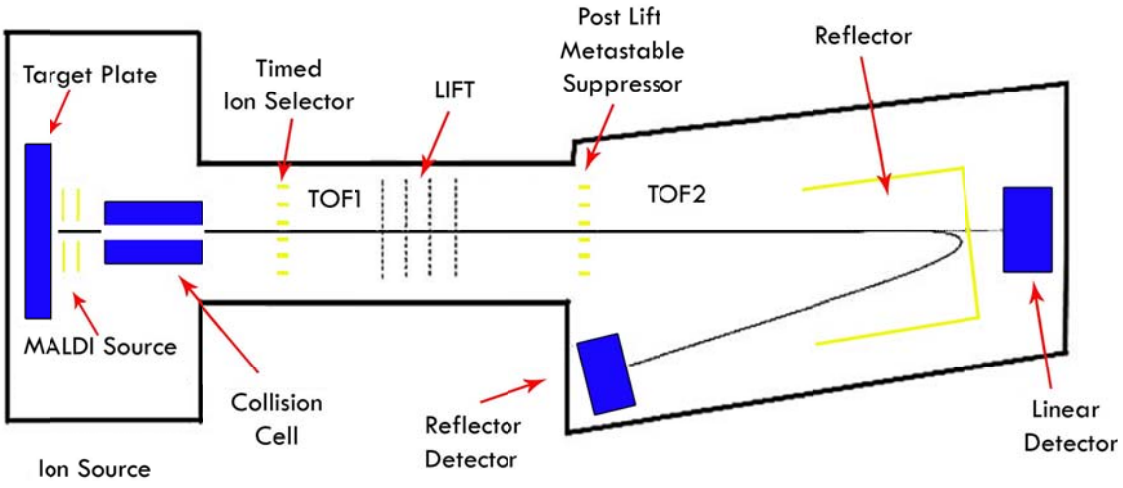


Figure 2-3. TOF/TOF mass spectrometer schematic.

This instrument can be operated in both linear and reflector mode and it can be operated in MS and MS/MS mode. The instrument can achieve a mass resolution ≥ 40,000 from 500 to 4000 Da and for intact proteins, with a mass accuracy < 1 ppm.[191]

The targets provided with the instrument have an integrated target transponder, which is read by the instrument allowing it to recognize the type of target plate being used for analysis. This instrument can use several different types of target plates however, for purposes of this study a 384 spot stainless steel plate was used.

The proteins and peptides were mixed on the 384 spot stainless steel plate using a 1:1 ratio of sample and saturated matrix solution. The matrices were 2-hydroxy-5-methoxybenzoic acid (sDHB), 2,5-dihydroxybenzoic acid (DHB), and 2,4-dimethoxy-3-hydroxycinnamic acid (sinapininc acid, SA) for proteins, and α-cyano-4-hydroxycinnamic acid (CHCA) for peptides. Structures and applications of the matrices used are found in Table 2-1. The protein samples were analyzed in linear positive ion mode and 500 shots averages were collected using a mass range from 30–100 kDa. For peptides, the spectra were obtained in reflector positive ion mode, using a mass range from 1–5 kDa. Masses were determined using FlexAnalysis software (Version 2.0, Bruker Daltonics).

Table 2-1. Commonly used matrices in MALDI for biomolecules.[192]

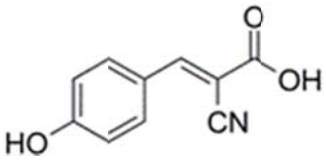
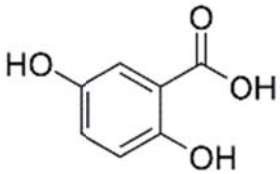
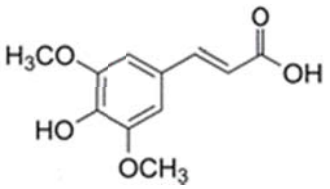
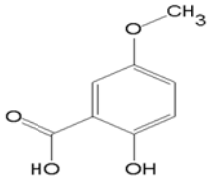
Matrix	Structure	Mass (Da)	Laser λ (nm)	Applications
α-cyano-4-hydroxycinnamic acid (CHCA)		189	337/355	Peptide/proteins mass<10 kDa
2,5-dihydroxybenzoic acid (DHB)		154	337/355	Carbohydrates & Proteins mass<5 kDa

Table 2-1. cont'd.

trans-3,5-dimethoxy-4-hydroxycinnamic acid (SA)		224	266/337/355	Peptide/proteins mass>10 kDa
2-hydroxy-5-methoxybenzoic acid (sDHB)		168	337/355	Carbohydrates & Proteins mass<5 kDa

2.7.2 LC-TOF Mass Spectrometer

An ESI-TOF MS (Agilent 6210) was used with a liquid chromatography HPLC system. The HPLC system used with this instrument was described in Section 2.2.2. This instrument consists of an electrospray source, a nebulizer and nebulizer gas inlet, capillary, skimmer, octopole ion guides, ion pulser for the orthogonal acceleration TOF, reflectron mirror, and a microchannel plate (MCP) detector. The capillary separates the spray chamber from the ion optics and has an applied potential from -4000 to 150 V. The ions are focused by lenses and enter the TOF mass analyzer. The ions pass into the TOF ion pulser and a high voltage pulse is applied to accelerate the ions into the flight tube. The ions are reflected by a reflectron into the detector. This instrument was used in this study to obtain intact protein mass measurements and has a mass accuracy of < 2 ppm and a mass resolution >13,000 measured at *m/z* 2722. The limit of detection for this instrument using ESI is in the range of 1–10 ng/mL. Mass spectra were acquired from 600 to 3000 *m/z* using ESI at 4200 V. The spectra were extracted and deconvoluted using Analyst software (Version 1.5, Agilent).

2.7.3 ESI-Q-TOF Mass Spectrometer

The mass spectrometer used in this work for peptide sequencing was a hybrid quadrupole time of flight tandem mass spectrometer (Applied Biosystems QSTAR XL). A diagram of the instrument is shown in Figure 2-4. The mass spectrometer consists of an ESI source, three

quadrupoles (focusing quadrupole, q_0; mass selection quadrupole, Q_1; and collision quadrupole, q_2), an orthogonal TOF, reflectron ion mirror, and a MCP detector. This instrument was operated in MS/MS mode for these studies. The instrument has an orthogonal pulser that pulses the ions out at repetition rate of several kHz using an applied electric field of 10 kV.

To perform MS/MS experiments the instrument software (Analyst) was used to 1) select the three most intense peaks for fragmentation, 2) select three different collision energies (25, 32 and 38 eV), and 3) a repetition rate ranging between 3 to 20 kHz for pulsing the ions into the TOF MS.

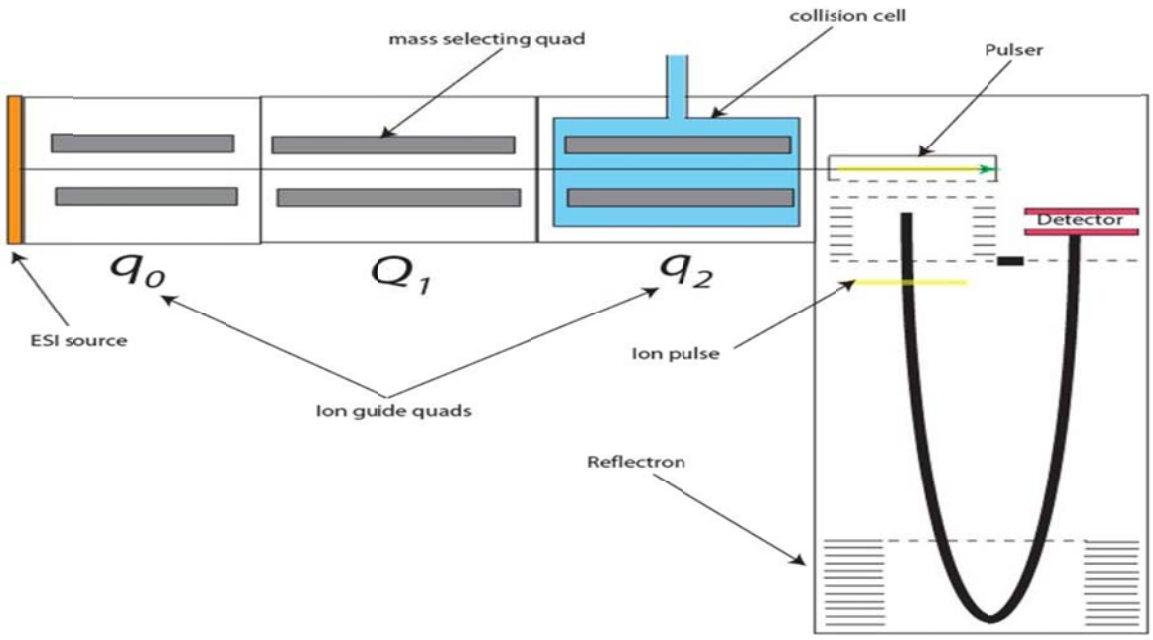


Figure 2-4. Q-TOF mass spectrometer schematic.

2.7.4 ESI-Q/IMS/TOF Mass Spectrometer

A hybrid quadrupole time-of-flight instrument with a travelling wave ion mobility cell was used in this work (Waters Synapt; Figure 2-5). This instrument has an ESI source, ion guide, quadrupole, ion mobility section, and an orthogonal time-of-flight. The ion mobility section is after the mass selection quadrupole and consists of a linear quadrupole ion trap which is used to

accumulate ions before the mobility separation; the accumulated ions are then released into the ion mobility cell.[140] The traveling wave transfer ion guide is used to transport the mobility-separated ions into the oa-TOF for mass analysis. A nano UPLC system is interfaced to a nanospray ionization source of the mass spectrometer. This instrument has a resolution of 15,000 and a mass accuracy < 10 ppm.

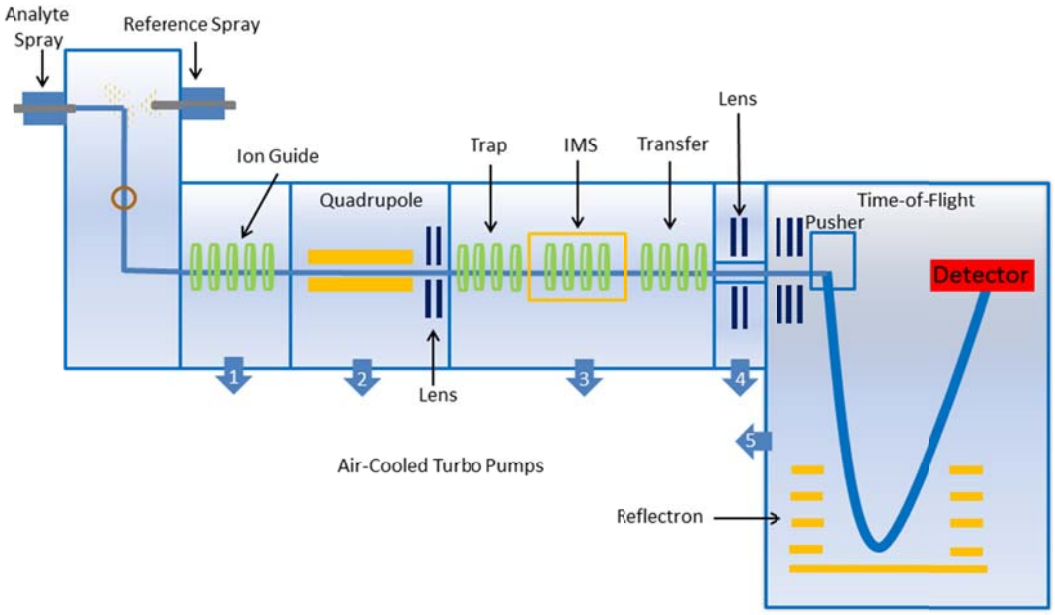


Figure 2-5. Schematic of the hybrid quadrupole/ion mobility separator/orthogonal acceleration time-of-flight.

In this study the instrument was operated in positive ion electrospray mode. The quadrupole was operated in RF mode in order to record full mobility spectra and the data was acquired using MassLynx software version 4.1. The peptide sample analysis on this instrument was performed at Waters Corporation, Milford, MA.

2.7.5 Protein Bioinformatics

Two approaches were used to sequence proteins: database searching and *de novo* sequencing. In the first method the tandem mass spectra from digested proteins were used to determine peptide sequences that were compared to protein sequences in databases. In this work,

the Mascot search engine was used for protein identification. The database search engine calculates peptide fragment masses *in silico* using the enzyme specificity for each protein sequence in the database. These calculated fragment masses are compared with the measured fragment masses and a score is calculated for comparison.[152, 154] Search engines can also extract partial sequence information from the spectrum based on the ion series. The nomenclature for fragment ions was proposed by Roepstorff and Fohlman[193] and later modified by Biemann[194] to designate peptide fragment ions. When the charge is retained on the N-terminus the ions are represented by the symbols a, b, and c and when the charge is retained on the C-terminus the ions are represented by the symbols x, y and z (Figure 2-6). The most common ions produced are b and y-ions.[195]

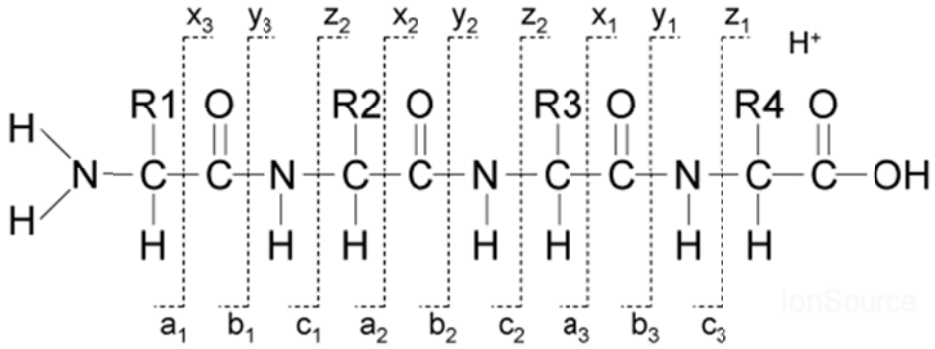


Figure 2-6. Common ions produced from peptide fragmentation.

To obtain sequence information from the tandem mass spectrum the molecular ion of the peptide is identified, which is the precursor ion in CID spectra followed by identifying the sequence ion series. The mass difference between the fragment ion peaks is determined and the masses of the 20 common amino acids are considered.

Mascot is based on the Mowse algorithm[196] which takes into account how many peptides of a given mass are in the database and also takes the protein size into consideration. Mascot also

calculates the probability that the observed match between the experimental data and a protein sequence is random.[152]

The second approach, *de novo* sequencing with database searching, is necessary for organisms that do not have a large number of proteins in the sequence database. One source for protein database information is the genome sequence of an organism. Protein databases can be constructed using the translated protein sequence. Three *de novo* sequencing software tools were used in this work: Mascot Distiller,[197] PEAKS,[167] and BioTools. Mascot Distiller (Version 2.4, Matrix Science, Boston, MA) generates peak lists for *de novo* sequencing. The sequencing software fits each peak in the mass spectrum to a hypothetical amino acid combination of the same mass. A list of amino acid combinations corresponding to each peak is generated from the best fit data.[197]

The second software used, PEAKS (Version 5.2, Bioinformatics Solutions, Waterloo, ON), applies four processes: preprocessing, candidate computation, refined scoring, and global and positional confidence scoring. First, the raw MS/MS data is smoothed and deconvoluted from doubly and triply charged ions to singly charged ions, and peak centering (taking the centroid of the first peak of the isotope distribution and removal of the other isotopic peaks) is performed. Second, sequences of all possible combinations of amino acids for the selected precursor mass are determined. The fragment ion series a, b, c, x and y are used to determine the peptide sequences. Next, the best sequences are scored using a stringent scoring scheme where the ion mass error tolerance is stricter. Last, PEAKS generates a confidence score for each of the top-scoring peptide sequences.[167]

BioTools (Version 2.2, Bruker Daltonics, Billerica, MA) uses MS/MS spectra from MALDI-TOF/TOF to generate sequence tags of various lengths and scores. Proposed sequences are calculated from the masses in the MS/MS spectrum using sequence tags. Sequence tags are

produced using the fragment masses and information added by the user to the software, such as terminal amino acids and enzymes used. The suggested sequence tags are then extended to produce a list of all peptide sequences that could fit the MS/MS data. The sequences are scored based on the number of matched ions. Several candidate sequences may be calculated with similar scores; in this case the proposed sequences needs to be manually fitted to the MS/MS data. However, if the peptide sequence produced complete fragmentation in the MS/MS spectrum one peptide sequence may be the best fit.[198]

The peptide sequences found by *de novo* sequencing are used in protein database searches in order to find proteins with similar sequences. The protein databases used in this work were the National Center for Biotechnology Information (NCBI) database, Mass Spectrometry Protein Sequence Database (MSDB), and Swissprot. NCBI is a non-identical protein and nucleic acid database that has sequences taken from multiple sources including Swissprot, the Protein Data Bank, and translations from annotated coding regions in the GenBank. One of the major advantages of NCBI is that it is frequently updated; therefore, this database was used the most in this work. MSDB is also a non-identical protein database that comprises a number of other protein databases; however, it is not updated as frequently. Swissprot is a non-redundant protein database that provides a high level of annotations such as protein function and post-translational modifications.

Proteins were also determined from peptide sequences using Basic Local Alignment Search Tool (BLAST). BLAST is used for comparing high-scoring segment pairs (HSP) in nucleotide or protein sequences to sequences in nucleotide or protein databases. HSP is a region between a pair of sequences, either nucleotide or amino acid, that share high level of similarity.[174, 199] BLAST can also provide evolutionary relationships for the species under consideration, which shows how two species are related with respect to their evolutionary

descent. For this work, BLAST was used to compare peptide sequences generated from alligator leukocyte and lectin proteins to proteins in the protein database. For BLAST searching, Swiss-Prot and NCBI databases were used to search peptide sequences using all taxonomies. The expect threshold was set to 20,000 and the score matrix used was PAM30.

To look for protein sequence similarities DIALIGN and CLUSTALW software were used. The DIALIGN software uses both local and global alignment features by either aligning all residues in each sequence or aligning only residue segments that are closely related.[200] CLUSTALW allows for the alignment of multiple protein sequences in order to determine similarities and differences.[201]

2.8 X-Ray Crystallography

In this work, small molecules were isolated from alligator leukocyte extracts and submitted for structural characterization using an X-ray diffractometer with charge coupled device (CCD) area-detector (Bruker Kappa Apex II, Madison, WI, USA). The X-ray diffractometer provides accurate and precise measurements of the entire three-dimensional structure of a molecule, including the bond angles and distances as well as crystal structure.

2.9 Deglycosylation

Deglycosylation is the removal of a sugar from a glycogen, especially a glycoprotein. The oligosaccharides can be removed using either an enzyme or a chemical reagent. There are several enzymes available for deglycosylation of N- and O-linked oligosaccharides; these include peptide-N-(N-acetyl-β-glucosaminyl)asparagine amidase (PNGaseF), endo F and endo H (endo-β-N-acetylglucosaminidases F and H) for removal of N-linked oligosaccharides,[202] as well as O-glycanase for removal of only galactose-N-acetylgalactosamine (Gal-GalNAc) O-linked oligosaccharides.[203] A commonly used enzyme is PNGaseF, which was used in this work. PNGase cleaves between the innermost N-acetylglucosamine (GlcNAc) and asparagine residues

of high mannose (Man) content and complex oligosaccharides from N-linked glycoproteins (Figure 2-7).[202, 204] This method is advantageous due to its high specificity and its nondestructive nature that allows the oligosaccharides to be recovered for further analysis.

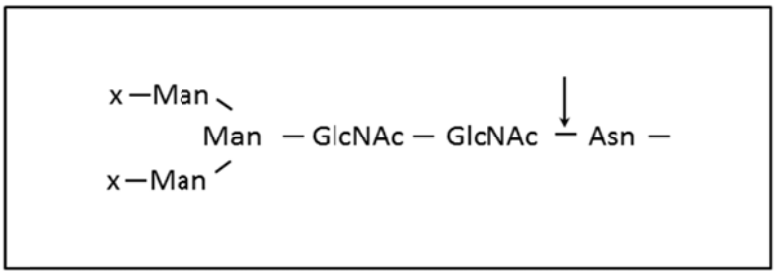


Figure 2-7. Illustration of PNGaseF cleavage location in glycoproteins. (x = H or sugars)

Another deglycosylation procedure used in this work was trifluoromethanesulfonic acid (TFMS) deglycosylation. TFMS is a commonly used chemical reagent that is used to remove oligosaccharides from glycoproteins.[205] Protein amide bonds are stable to TFMS whereas glycosidic bonds are not. The chemical reaction is carried out in anhydrous conditions in TFMS without a solvent in the presence of a scavenger that protects the amino acids. During the solvolysis reaction, cleavage of the glycosidic bonds occurs and the protein remains intact. It is important to ensure that the reaction is performed in the absence of water to prevent cleavage of peptide bonds. TFMS is advantageous when compared to enzymatic deglycosylation because carbohydrates can be removed from a glycoprotein regardless of its structure or its linkage (N- or O-linkage).[205]

2.10 Radial Diffusion Assay

Antimicrobial activity can be monitored using growth inhibition assays. The antimicrobial peptides isolated in this work were tested at various stages using a radial diffusion assay.[206, 207] Bacteria were incubated overnight at 37ºC in trypticase soy broth (TSB) (Sigma

Aldrich) that was previously autoclaved at 120ºC for 20 minutes. The TSB was prepared using 3.0 mg of TSB medium dissolved in 100 mL of nanopure water. A 20 µL volume of bacteria was added to 15 mL of the TSB medium containing 1% (wt/vol) low-electroendomosis-type agarose (Sigma Aldrich). The previously prepared TSB medium containing agarose (agar) was heated in a water bath at 42ºC before use. The agar was poured into an agar plate to form a uniform layer and allowed to solidify. Evenly spaced holes were punched using a sterilized pipette tip connected to a vacuum for suction to create wells. Added to individual holes was 5 µL of control sample (0.1% acetic acid) and polymyxin B, which was used as the standard and isolated peptide samples (antimicrobial peptides). The samples were added in triplicate and the plates were incubated for 3 hours at room temperature followed by an overnight incubation at 37ºC. After incubation, the size of the clear zone around each well was measured. The size of a clear zone around the well indicates the antimicrobial activity of the peptides. The absence of a clear zone indicates no antimicrobial activity.

2.11 Reagents and Chemicals

Ammonium bicarbonate and cyanogen bromide (CNBr) were obtained from Sigma Aldrich (St. Louis, MO). The chemicals trifluoroacetic acid (TFA) and acetic acid were purchased from Fisher Scientific. The MALDI matrix α-cyano-4-hydroxycinnamic acid (CHCA) was obtained from Sigma and 3, 5-dimethoxy-4-hydroxycinnamic acid (sinapic acid), DHB, and sDHB were obtained from Fluka. Trypsin, α-chymotrypsin, endoprotease Asp-N from *Pseudomonas fragi*, endoprotease Glu-C from *Staphylococcus aureus*, and endoprotease Lys-C from *Lysobacter enzymogenes* were obtained from Sigma-Aldrich (St. Louis, MO, USA). Dithiothreitol (DTT), iodacetamide, HPLC grade acetonitrile, formic acid, and glu-fibrinopeptide standards (Sigma-Aldrich) were used without further purification.

Laemmli sample buffer, 10× SDS-tris-glycine buffer, β-mercaptanol, tris-HCl gradient gel, Coomassie blue stain (Bio-Safe), and molecular weight marker standards were purchased from Bio-Rad (Hercules, CA, USA). Dithiothreitol, iodacetamide, and glu-fibrinopeptide were purchased from Sigma-Aldrich (St. Louis, MO, USA). Trifluoromethanesulfonic acid (TFMS), pyridine, and toluene were purchased as a kit from Prozyme (Hayward, CA, USA).

The bacteria strains used in this study were obtained from American Type Culture Collection (ATCC) (Manassas, VA). The following American Type Culture Collection (ATCC) bacterial strains were used: *Escherichia coli* (35218), *Shigella flexneri* (12022), *Enterobacter cloacae* (23355), *Klebsiella oxytoca* (33496), *Staphylococcus aureus* (51153), and *Staphylococcus epidermidis* (29887). Trypticase soy broth was purchased from Voigt Global Distribution (Broth 23400, Lawrence, KS) and agarose Type I was purchased from Sigma.

CHAPTER 3. PROTEOME ANALYSIS OF THE LEUKOCYTES FROM THE AMERICAN ALLIGATOR (*ALLIGATOR MISSISSIPPIENSIS*) USING MASS SPECTROMETRY*

The purpose of the research described in this chapter was to investigate proteins related to the alligator immune system. This chapter describes the use of mass spectrometry in conjunction with gel electrophoresis and liquid chromatography to determine peptide sequences from American alligator (*Alligator mississippiensis*) leukocytes and to identify similar proteins based on homology. Proteins from leukocyte extracts were separated using two-dimensional gel electrophoresis and the major bands were excised, digested, and analyzed with on-line nano LC MS/MS to generate peptide sequences. The sequences generated were used to identify proteins and infer their function based on similarity to previously identified proteins. Similar proteins were identified based on matching two or more peptides from the alligator protein to similar proteins by searching against the NCBI database using Mascot and Basic Local Alignment Search Tool (BLAST). For those proteins with only one peptide matching, the phylum of the organism corresponding to the matching protein was considered.

3.1 Introduction

A large amount of anecdotal evidence exists to suggest that crocodilians are resistant to bacterial, fungal and viral infections[208-210] and have the ability to fight a variety of diseases that may pose human health risks.[211, 212] Several studies have shown that the serum of alligators exhibit innate immunity against fungi, viruses and various bacterial species.[208, 213, 214] It has been reported that alligator leukocyte extracts exhibit broad-spectrum antimicrobial activity against important human pathogens such as *C. albicans*, *S. faecalis* and *E. coli*.[215] It is anticipated that elucidation of the alligator blood proteome will aid in identifying these antibacterial, antifungal

*Reprinted by permission of the Elsevier. The work reported in this chapter has been published in *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*.[6]

and antiviral peptides believed to be responsible for these properties, which in turn could lead to the development of new antibiotics.

Antimicrobial peptides are typically found in granulocytic leukocytes and are part of the innate immune system;[216] these antimicrobial peptides are active against a wide range of microorganisms such as fungi, bacteria, and viruses.[42, 217] Various classes of antimicrobial peptides have been isolated from organisms such as bacteria, plants and animals[218-220] and are characterized by their amino acid sequence and structure.[217] The majority of the antimicrobial peptides assume linear α-helical or disulfide-stabilized β-sheet conformations[24, 217] and are typically rich in lysine and arginine; however, some are also rich in proline, histidine or tryptophan residues.[217] These antimicrobial peptides are typically less than 10 kDa in molecular mass and are cationic and amphipathic.[24, 217] However, antimicrobial properties have been reported for anionic peptides as well as for proteins.[24]

Mass spectrometry (MS) and tandem mass spectrometry (MS/MS) in combination with multi-dimensional separations have become powerful techniques for peptide and protein identification.[221, 222] Mass spectrometry-based approaches have been extensively employed for the identification and characterization of proteins in blood components such as leukocytes,[33, 223, 224] serum,[225] platelets,[226, 227] plasma[228] and red blood cells.[229] For example, online LC MS/MS was used to identify 1444 proteins in human serum.[230] In other studies, the human blood plasma proteome has been characterized using multi-dimensional separation techniques along with MS/MS and database searching to identify over a thousand proteins.[125, 221, 228]

One of the major challenges in the study of the alligator blood proteome is the limited information available regarding the reptilian genome and proteome.[33, 231] For proteomic analysis of species with limited genomic and proteomic data, *de novo* sequencing can be used to determine protein sequences.[175] In *de novo* sequencing, proteins are isolated and enzymatically

digested and the resulting peptides are analyzed by tandem mass spectrometry. The sequence is determined by observing the mass differences between fragment peaks in the tandem mass spectra which are generated by low energy collision-induced dissociation (CID) of the peptide of interest.[232] The amino acid sequence can be determined from the N-terminus using the b-ion series or from the C-terminus using the y-ion series.[233] *De novo* sequencing has been used to identify a number of proteins from species with limited representation in protein databases.[234-237]

In the present study, proteins were isolated from the leukocytes of the American alligator (*Alligator mississippiensis*) and separated by one or two dimensional gel electrophoresis. One of the challenges in this study is limited proteome data for the alligator hence peptide sequences were determined using *de novo* sequencing with database searching and BLAST search to identify similar proteins. Using this approach 43 proteins were identified that represent abundant proteins from the alligator leukocyte.

3.2 Experimental

Adult alligators were captured for blood samples and leukocytes were isolated as described in Section 2.1. One-dimensional polyacrylamide gel electrophoresis (PAGE) was performed using precast 4–20% polyacrylamide gradient gels (8.6 × 6.8 cm) on a small format gel electrophoresis system. Leukocyte extracts were dissolved in sample diluting buffer (Tris-HCl at pH 6.8, containing SDS, glycerol, 2-mercaptoethanol, and bromphenol blue dye) and boiled for 3 min prior to loading on the gel. A 10 µl sample volume was loaded on each individual lane and in a separate lane, 10 µl of a protein molecular weight standard solution was loaded. Electrophoresis was conducted for 1 hr at 100 V using tris–glycine (pH 8.3) as the gel running buffer. The gels were then stained overnight with Coomassie blue to aid in visualizing the proteins bands. The gels were rinsed twice with distilled water for 10 min each time to

remove excess stain. The molecular weights of the proteins were estimated using protein molecular weight standards.

Two-dimensional electrophoresis was performed using either the small format or a standard format electrophoresis system. Additional two-dimensional PAGE was also performed by Kendrick Lab Inc. For isoelectric focusing, leukocyte extract was dissolved in sample buffer containing 8 M urea, 2% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 50 mM dithiothreitol (DTT) and 1% carrier ampholytes. The linear IEF gel strips (7 cm), pH 3–10 were rehydrated with sample overnight. Isoelectric focusing was carried out in an isoelectric focusing cell at 250 V for 20 min and 14,000 V-h (rapid ramping). SDS-PAGE electrophoresis was performed in slab gels (18.3 × 19.3 cm) with 1.00 mm thickness. The electrophoresis was conducted at 200 V for 40 min. The gels were stained with ProteoSilver Plus silver stain overnight to visualize the protein bands.

The bands from the 1-D gels were excised and cut into ~1 mm^3 cubes, and the Coomassie blue was washed away with 200 mM NH_4HCO_3 and 40% acetonitrile (2 × 30 min). For silver stained spots, the ProteoSilver Plus stain was removed using commercial destaining solutions.[238] After destaining, the gel pieces were dried in a vacuum centrifuge, rehydrated with 20 µL of 0.4 µg trypsin in 40 mM ammonium bicarbonate containing 9% acetonitrile to completely cover the gel pieces. The tubes were then incubated overnight at 37 °C.[239, 240] Peptides were extracted from gel pieces with 60% acetonitrile and 40% water containing 0.1% trifluoroacetic acid (TFA).

Liquid chromatography electrospray tandem mass spectrometry (LC MS/MS) was used to analyze the peptides extracted from the gel spots. The tryptic peptides were dissolved in 50 µl of 0.1% formic acid (FA) and 10 µl of the digest was injected with a microplate autosampler onto a 0.3 × 1 mm trapping column (PepMap C_{18}) on a nano LC system at a flow rate of 10 µL/min. After sample loading, the trapping column was washed with 0.1 % FA at a flow rate of 5

μL/min for an additional 5 min. The peptides were then eluted onto a 75 μm × 15 cm C_{18} column (Biobasic Vydac) and separated using a gradient of 5–40% B over 60 min with a flow rate of 200 nL/min. Solvent A was 95% water and 5% acetonitrile containing 0.1% FA. Solvent B was 80% acetonitrile and 20% water containing 0.1% FA

The effluent was directed to the quadrupole time-of-flight mass spectrometer (QSTAR XL) and ionized using a nano electrospray source at a voltage of 2.5 kV. The mass spectrometer was operated in information dependent acquisition (IDA) mode. Three collision energies (25, 38, and 50 eV) were selected for fragmentation of the peptides.

The peptide sequences were generated from MS/MS data using Mascot Distiller software. The software determined amino acid sequences from MS/MS data using peak fitting and isotope distribution.[241, 242] The *de novo* sequencing results were verified manually.

De Novo peptide sequences were subjected to database searching using Mascot[154] and BLAST to find proteins that contain sequences matching with sequences determined by *de novo* sequencing. For BLAST searching, Swiss-Prot and NCBI databases were used to search peptide sequences using all taxonomies. The expect threshold was set to 20,000 and the score matrix used was PAM30.

3.3 Results and Discussion

One-dimensional electrophoresis of the leukocyte extracts revealed six major bands ranging from 10 to 75 kDa. The bands marked in Figure 3-1 were excised, digested with trypsin and analyzed by LC-MS/MS.

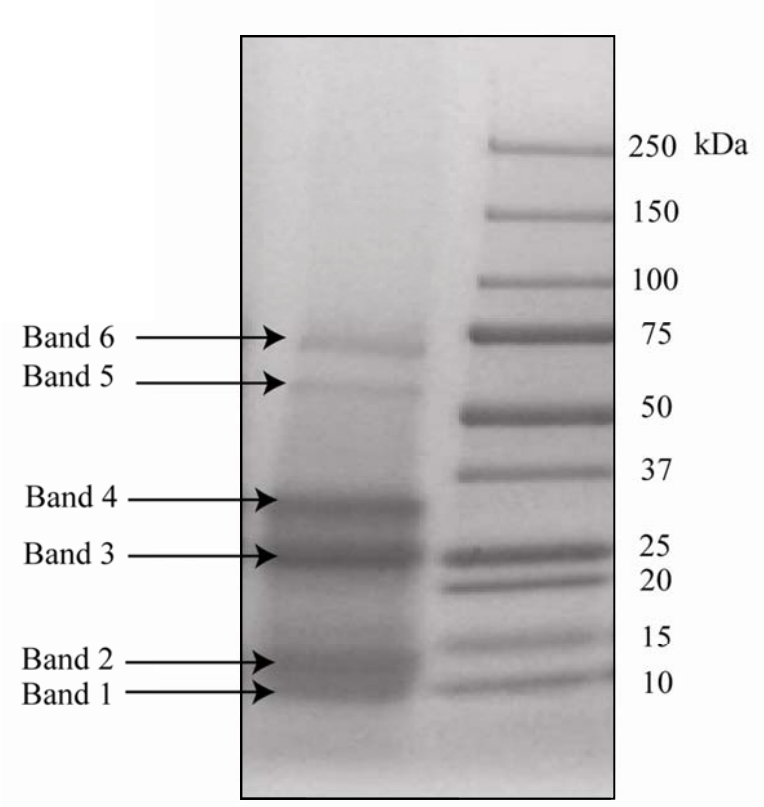


Figure 3-1. Alligator leukocyte extract separated on 1D gel electrophoresis and stained with Coomassie blue. Band numbers are referenced in Tables 1 and 2.

Figure 3-2 shows a two dimensional separation of the leukocyte extracts stained with mass spectrometry compatible silver stain. The number of proteins visible on the 2D gel from the alligator leukocyte was lower than a typical cell extract due to acetic acid denaturation of some proteins. The majority of the protein spots were observed in the pI range of approximately 5–7 and molecular weight range of approximately 20 to 200 kDa. Thirty-seven of the most intense well resolved spots (labeled on the 2D gel in Figure 3-2) were selected for further analysis and excised, digested and analyzed by LC MS/MS.

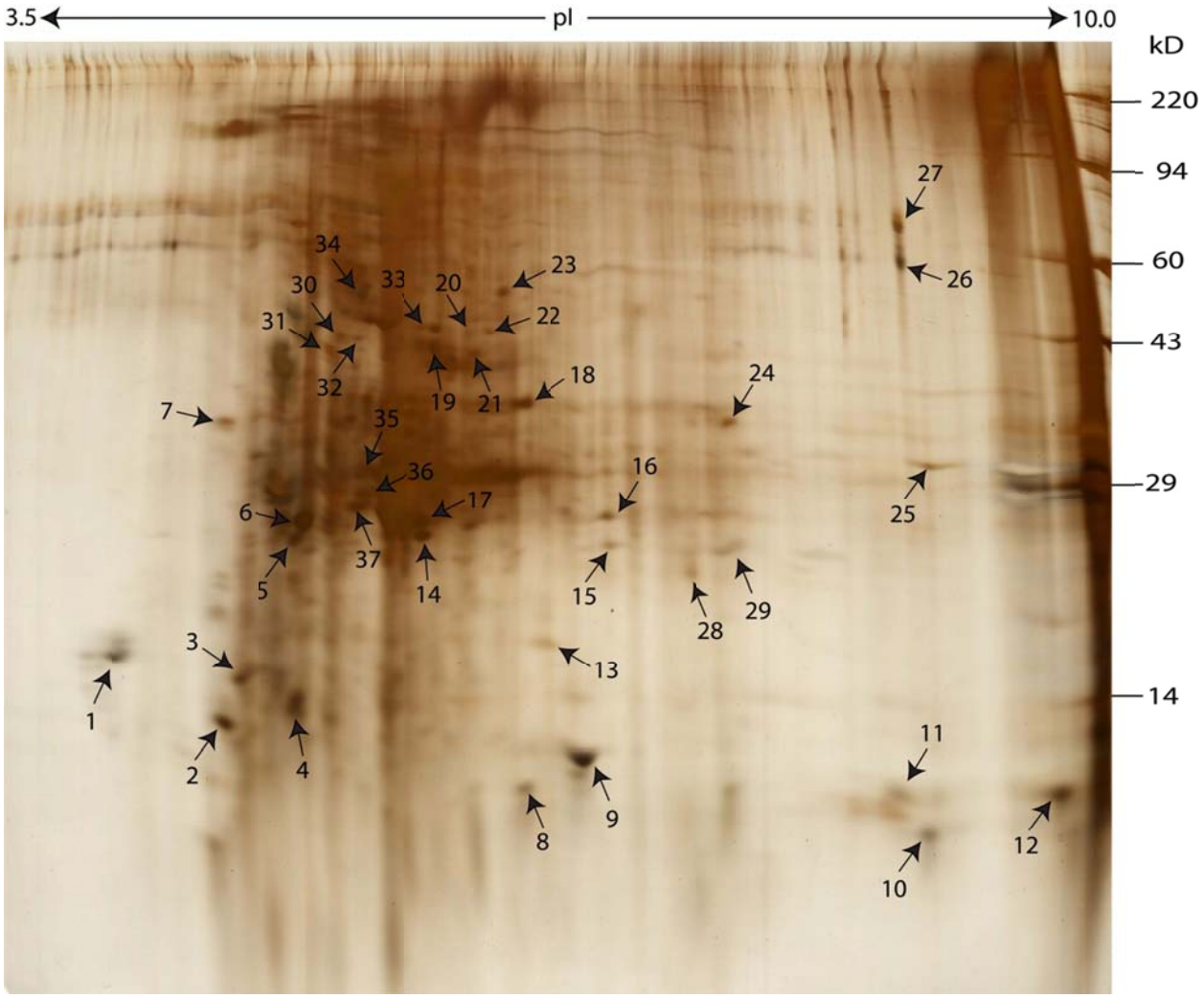


Figure 3-2. 2D separation of alligator leukocyte stained with Proteosilver stain for mass spectrometry analysis. Approximately 400 µg of leukocyte protein was loaded and separated on a 2D large gel format.

Figure 3-3 shows a selected separation and associated mass spectra that illustrates the method of peptide sequences determination. The LC separation in Figure 3-3a shows the total ion chromatogram for the digest of Band 1. Figure 3-3**b** shows an example of a mass spectrum acquired at 33.3 minutes as indicated with an asterisk. Tandem MS data were collected and peptide sequences were determined by *de novo* sequencing. Figure 3-3c illustrates the peptide fragment series that was found using *de novo* from the MS/MS data acquired from Band 1 of

Figure 3-1. A similar strategy was followed for the other labeled bands from Figure 1 as well as the protein spots in Figure 3-2.

For the characterization of peptides and proteins in the gel bands and spots, a three-step strategy was adopted. First, the peptides from the digest were sequenced using *de novo* sequencing as outlined above. Second, the sequences were searched against the full SwissProt and NCBI databases using Mascot to identify similar proteins. Third, a BLAST search was performed for *de novo* peptide sequences that did not match sequences in the database search of step two.

The Mascot search results are summarized in Table 3-1. The matched peptides shown in the table were identified from either the 1D or 2D gel as indicated in the table. The proteins included in the table had at least two peptide matches and sequence coverage ranging from 1 to 37%. The proteins polyubiquitin, hypothetical protein LOC100158585, ubiquitin, hypothetical protein, hypothetical protein 4732456N10, endonuclease P1, myosin alkali light chain, myosin regulatory light chain, actin, β-actin, actin-2, talin 1, chain A, chicken annexin V {complex with Ca^{2+}), alkaline phosphatase-like enzyme, lysozyme C, ras-related protein O-Krev, annexin I, hemoglobin and vimentin-4 proteins were identified.

The *de novo* sequenced peptides that didn't match sequences in the database were subjected to a BLAST search to identify proteins containing similar sequences. The results of the BLAST search are summarized in Table 3-2 (and Appendix A, Table 1).

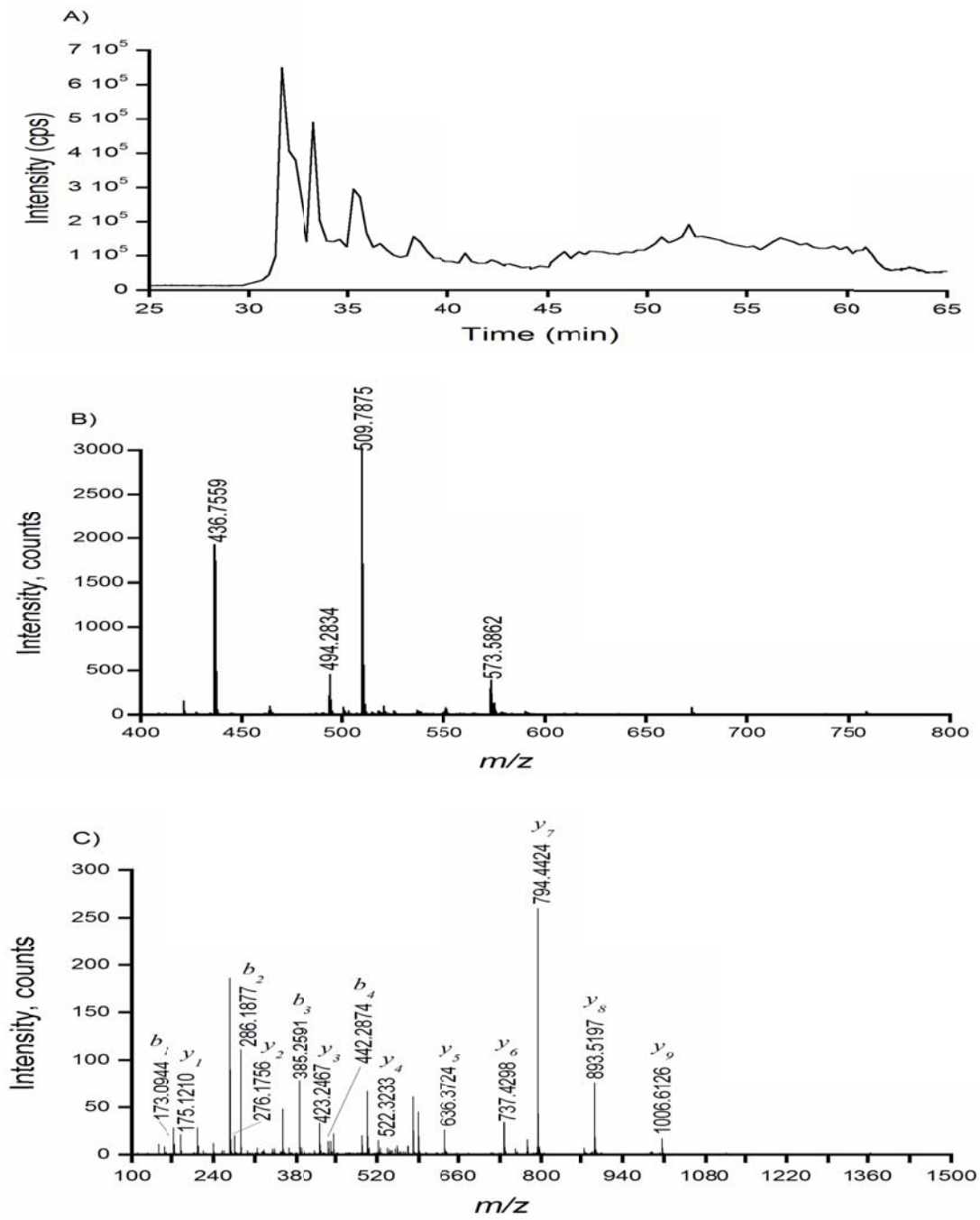


Figure 3-3. Mass spectrometry data for peptide sequencing. Figure **(a)** shows the total ion chromatogram of all the mass spectra recorded during the gradient LC separation as a function of time. The ***** in Figure **(a)** indicates the trace for a particular ion eluted at approximately 33.3 minutes. Figure **(b)** shows the mass spectrum of the peptides eluted during the time indicated in Figure **(a)**. Figure **(c)** shows the MS/MS plot of the peptide ion highlighted in Figure 2 **(b)**. The mass differences between the y-ions series is used to predict the amino acid series. The y-ion series is written in the direction from carboxy to amino terminus direction, going from left to right.

Table 3-1. Proteins identified in leukocyte extract using Mascot search. Proteins were identified by Mascot search of MS/MS data from the leukocyte peptide digest with databases, NCBI, MSDB and SWISS-PROT. NCBI accession numbers are provided in table.

Band/Spot No.	*m/z*	Charge	Matched Peptides	Name	Function	Accession No.	Mascot Score	% Coverage
B1	383.2	+2	MQIFVK	Polyubiquitin	[a]	**BAA23486**	340	8
	520.3	+2	EGIPPDQQR					
	534.8	+2	ESTLHLVLR					
	541.8	+2	TLSDYNIQK					
	508.6	+3	IQDKEGIPPDQQR					
	894.5	+2	TITIEVEPSDTIENVK					
B1	520.3	+2	EGIPPDQQR	Ubiquitin	[a]	**AAG22093**	368	32
	526.3	+2	NSTLHLVLR					
	534.8	+2	ESTLHLVLR					
	541.3	+2	TLSDYNIQK					
	508.6	+3	IQDKEGIPPDQQR					
	895.0	+2	TITNEVEPSDTIENVK					
	902.0	+2	TITKEVEPSDTIENVK					
B3	405.2	+2	LASYLDK	Hypothetical protein LOC100158585	[b]	**NP_001121485**	70	3
	533.3	+2	AQYEDLAKK					
B3	515.3	+2	VLDELTLAR	Hypothetical protein	[b]	**XP_511487**	218	3
	579.3	+2	QGVEADVNGLR					
	651.4	+2	ALEEANADLEVK					
	453.3	+3	QSVEADINGLRR					
B2	565.8	+2	NLASAVSDLLK					
	871.8	+3	AVSSAIAHLLGEVAQGNENYTGIAAR					
3	639.3	+2	LALDIEIATYR	Hypothetical protein 4732456N10	[b]	**NP_808385**	187	6
	651.8	+2	SLNLDSIIAEVK					
	738.4	+2	FLEQQNKVLETK					
1	518.9	+3	LIGGHALSDAESWAK	Endonuclease P1	[c]	**P24289**	82	11
	565.6	+3	LANWINEIHGSEIAK					
2	518.9	+3	LIGGHALSDAESWAK	Endonuclease P1	c	**P24289**	255	25
	565.6	+3	LANWINEIHGSEIAK					
	567.9	+3	VSDSSLSSENHAEALR					
	851.5	+2	VSDSSLSSENHAEALR					
	665.5	+4	FLVHFIGDMTQPLHDEAYAVGGNK					
4	518.9	+3	LIGGHALSDAESWAK	Endonuclease P1	[c]	**P24289**	72	11
	565.6	+3	LANWINEIHGSEIAK					
5	518.9	+3	LIGGHALSDAESWAK	Endonuclease P1	[c]	**P24289**	100	11
	565.6	+3	LANWINEIHGSEIAK					
7	518.9	+3	LIGGHALSDAESWAK	Endonuclease P1	[c]	**P24289**	72	11
	565.6	+3	LANWINEIHGSEIAK					
9	518.9	+3	LIGGHALSDAESWAK	Endonuclease P1	[c]	**P24289**	381	20
	565.6	+3	LANWINEIHGSEIAK					
11	518.9	+3	LIGGHALSDAESWAK	Endonuclease P1	[c]	**P24289**	72	11
	565.6	+3	LANWINEIHGSEIAK					
12	518.9	+3	LIGGHALSDAESWAK	Endonuclease P1	[c]	**P24289**	98	11
	565.6	+3	LANWINEIHGSEIAK					
2	498.3	+2	HVLVTLGEK	Myosin alkali light chain	[d]	**AAA48979**	181	30
	513.2	+2	EAFQLFDR					
	686.8	+2	ALGQNPTNAEVMK					
	575.2	+3	VFDKEGNGTVMGAEIR					
3	518.2	+2	ELLTTMGDR	Myosin regulatory	[e]	**1805343A**	189	18

[a] Removal of damaged or unnecessary proteins and controlling of other cellular processes
[b] Unknown function
[c] Hydrolyzes only single stranded DNA and RNA without specificity for bases
[d] Binds actin and modulates myosin motor function
[e] Affect myosin motor function and kinetics

Table 3-1 cont'd.

				light chain				
	614.8	+2	LNGTDPEDVIR					
	708.3	+2	FTDEEVDELYR					
6	499.7	+2	DLTDYLMK	Actin	[f]	**ACD44586**	188	14
	895.9	+2	SYELPDGQVITIGNER					
14	527.3	+2	TALAPSTMK	Actin	[f]	**ACF32719**	79	4
17	566.7	+2	GYSFTTTAER	β-actin	[f]	**ABB91894**	255	37
	581.3	+2	EITALAPSTMK					
	895.9	+2	SYELPDGQVITIGNER					
	855.4	+3	GYSFTTTAER					
19	473.2	+2	AVFPSIVGR	β-actin	[f]	**CAA27396**	81	7
	652.0	+3	VAPEEHPVLLTEAPLNPK					
33	400.2	+2	AVFPSIVGR	Actin-1	[f]	**P02578**	275	18
	505.9	+3	IWHHTFYNELR					
	652.0	+3	VAPEEHPVLLTEAPLNPK					
	499.7	+2	DLTDYLMK					
	566.7	+2	GYSFTTTAER					
	581.3	+2	ELTALAPSTMK					
20	473.2	+2	AVFPSIVGR	β-actin	[f]	**ABD48797**	139	10
	505.9	+3	IWHHTFYNELR					
	652.0	+3	VAPEEHPVLLTEAPLNPK					
35	499.7	+2	DLTDYLMK	Actin	[f]	**AAA20641**	103	13
	652.0	+3	VAPEEHPVLLTEAPLNPK					
	855.7	+3	LCYDALDFEQEMQTAASSSSLEK					
36	499.7	+2	DLTDYLMK	Actin-2	[f]	**P45885**	164	14
	505.9	+3	IWHHTFYNELR					
	652.0	+3	VAPEEHPVLLTEAPLNPK					
	654.3	+3	YPIEHGIITNWDDMEK					
37	505.9	+3	IWHHTFYNELR	Actin	[f]	**Q99023**	119	9
	652.0	+3	VAPEEHPVLLTEAPINPK					
	499.7	+2	DLTDYLMK					
37	378.2	+2	LLLAVVK	Chain A, Crystal Structures of Chicken Annexin V in Complex with Ca2+	[g]	**1YII_A**	102	5
	501.3	+2	VLTEILASR					
B2	565.8	+2	NLASAVSDLLK	Talin 1	[h]	**XP_002192479**	141	1
	871.8	+3	AVSSAIAHLLGEVAQGNENYTGIAAR					
9	539.9	+2	GTSIFGLAPSK	Alkaline phosphatase-like	[i]	**NP_001096792**	114	4
	714.9	+2	DKQNLVQAWQAK					
12	877.4	+2	NTDGSTDYGILQINSR	Lysozyme C	[j]	**LZFER**	114	10
13	747.9	+2	INVNEIFYDLVR	Ras-related protein O-Krev	[k]	**P22123**	95	6
18	701.4	+2	GVDEATIIEILTK	Annexin I	[l]	**CAA39971**	141	8
	851.9	+2	GLGTDEDTLNEILASR					
31	578.3	+2	FGGPGTASRPS	Vimentin-4	[m]	**P24790**	151	6
	627.7	+2	TRSYVTTSTRTYSLG					
	561.2	+2	DLYEEEMR					
12	510.8	+2	VLASFGEAVK	Hemoglobin β chain	[n]	**P02130**	131	19
	560.3	+2	KFIVDLWAK					
	566.7	+2	FHVDPENFK					

[f] Highly conserved proteins that are involved in various types of cell motility and are ubiquitously expressed in all eukaryotic cells

[g] Exocytosis

[h] Cytoskeletal protein in lymphocyte involved in connections of major cytoskeletal structures to the plasma membrane

[i] Causes the chemical reactions that remove phosphates from a number of different molecules

[j] Cell derived leukocyte that is part of the immune system to fight against infectious diseases and foreign material

[k] Controls a spectrum of diverse cellular processes

[l] Ca^{2+} dependent phospholipid binding proteins with potential anti-inflammatory activity

[m] Filament protein

[n] Involved in oxygen transport from the lung to the various peripheral tissues

Table 3-2. Proteins identified at single peptide level using BLAST. *De novo* sequenced peptides obtained from a gel digest were searched using BLAST and proteins were matched based on sequence similarity and evolutionary relationship using the NCBI database. The top E-values with the closest related organisms are reported.

Band Number	*m/z*	Charge	*De novo* Sequence	Protein	Function	Accession Number	Protein Length	Organism	E-Value
B4	1060.6	+2	EINLSPDSTSAVVSGLMVATK	Fibronectin	[a]	**P11722**	1256	Gallus gallus	2e-11
B4	536.3	+2	YEVSVYALK	Fibronectin	[a]	**P11722**	1256	Gallus gallus	0.33
B4	701.3	+2	YQINQQWER	Fibronectin	[a]	**Q91740**	2481	Xenopus laevis	0.013
B1	613.9	+2	LIALLEVLSQK	Filamin-C	[b]	**Q14315**	2725	Homo sapiens	0.019
B1	487.3	+2	VYGPGVE	Filamin-C	[b]	**Q14315**	2725	Homo sapiens	36
B2	736.4	+2	AQQVSQGLDLLTAK	Vinculin	[c]	**P12003**	1135	Gallus gallus	5e-05
B2	739.4	+2	LGQMTDQ+ADLR	Vinculin	[c]	**P12003**	1135	Gallus gallus	0.006
B2	655.9	+2	TVTAMDVVYALK	Histone H4	[d]	**P70081**	103	Gallus gallus	6e-04
B2	628.7	+2	LITKAVSASK	Histone H1	[e]	**P09987**	218	Gallus gallus	0.49
B6	833.9	+2	FSGSGSGTDFTFTIS	Ig K chain V-I region Lay	[f]	**P01605**	108	Homo sapiens	5e-06

[a] Cell adhesion, cell motility, opsonization, wound healing & cell shape maintenance
[b] Plays a central role in muscle cells, probably by functioning as a large actin-cross-linking protein
[c] Involved in attachment of the actin-based microfilaments to the plasma membrane; may also play roles in cell morphology & locomotion
[d] Plays a central role in transcription regulation, DNA repair, DNA replication and chromosomal stability; a mixture of histones H2B and H4 has antimicrobial activity against Gram-positive bacterium M.luteus
[e] Condensation of nucleosome chains into higher order structures
[f] Used by the immune system to identify and neutralize foreign objects, such as bacteria and viruses

Table 3-2 cont'd.

B6	751.4	+2	VFGGGTKLTVL QPK	Ig Λ chain V-III region LOI	[g]	**P80748**	111	Homo sapiens	7e-05
B2	510.8	+2	VLASFGEAVK	Hemoglobin, β	[h]	**P02130**	146	Alligator mississippiensis	0.27
B2	500.8	+2	PISGDPK	Microtubule-actin cross-linking factor 1	[i]	**Q9UPN3**	5430	Homo sapiens	54
B6	522.8	+2	IMSIVDPNR	α-actinin-1	[j]	**P05094**	893	Gallus gallus	0.1
B5	506.2	+3	QEYDESGPSIVHR	γ-Actin	[k]	**Q5JAK2**	375	Rana lessonae	3e-05
B1	683.9	+2	STDYGILQINSR	Lysozyme C	[l]	**P00698**	147	Gallus gallus	6e-04
B1	503.3	+2	WDAW+ALK	Acyl-CoA-binding protein	[m]	**Q9PRL8**	86	Gallus gallus	1.7
B4	509.8	+2	MLPMQK	Myeloid protein 1	[n]	**P08940**	326	Gallus gallus	36
B1	565.3	+2	LVTDVQEAVR	Proactivator polypeptide (Containing Saposin-A, B, C & D)	[o]	**O13035**	518	Gallus gallus	0.084
B2	646.9	+2	ISMPDFDLNLK	Neuroblast differentiation-associated protein AHNAK	[p]	**Q09666**	5890	Homo sapiens	0.001
B2	772.0	+2	AVASAAAALVLK	Talin-1	[q]	**P54939**	2541	Gallus gallus	0.026

[g] Activates the alternative complement pathway by binding to the short consensus repeat domain 3 (SCR3) of factor H
[h] Involved in oxygen transport from the lung to the various peripheral tissues
[i] Cross-linking actin to other cytoskeletal proteins
[j] Anchor actin to a variety of intracellular structures
[k] Highly conserved protein involved in various types of cell motility
[l] Enhances the activity of immunoagents
[m] Displace diazepam from the benzodiazepine recognition site located on the GABA type A receptor. Diazepam binding inhibitor (32-86) has antibacterial properties
[n] Cell derived leukocyte that is part of the immune system to fight against infectious diseases and foreign material
[o] Lysosomal degradation of sphingolipids takes place by the sequential action of specific hydrolases
[p] Required for neuronal cell differentiation
[q] Cytoskeletal protein in lymphocyte involved in connections of major cytoskeletal structures to the plasma membrane

Table 3-2 cont'd.

B2	525.8	+2	LGTFLEN	Histone acetyltransferase p300	[r]	**Q09472**	2414	Homo sapiens	36
B4	821.9	+2	V+QQQADDAEER	Putative tropomyosin α-3 chain-like protein	[s]	**A6NL28**	223	Homo sapiens	0.006
B6	587.3	+2	PP+PPARAA	Protein bassoon	[t]	**Q9UPA5**	3926	Homo sapiens	17
B5	506.8	+2	PSLPSGVD	Stromelysin-1	[u]	**P08254**	477	Homo sapiens	12
B6	583.8	+2	AAGEIIAIPRR	Envelope glycoprotein gp160	[v]	**Q89607**	852	HIV-2 B_EHO	0.011
B1	831.4	+2	NSWGTSWGEDGYFR	Dipeptidyl-peptidase 1 (Cathepsin C)	[w]	**Q60HG6**	463	Macaca fascicularis	5e-07
B5	684.9	+2	FRTTMLQDSI	Origin recognition complex subunit 2	[x]	**Q75PQ8**	576	Rattus norvegicus	0.014
B6	586.8	+2	LP EQGT SSR	Cadherin EGF LAG seven-pass G-type receptor 2	[y]	**Q9QYP2**	2144	Bos Taurus	0.36
B6	593.3	+2	AILYNYWDK	Complement component c3	[z]	**CAC69535**	401	Crocodylus niloticus	0.50
B6	506.8	+2	NEALIALLR	Plastin-2	[aa]	**P13796**	627	Homo sapiens	0.79

[r] Functions as histone acetyltransferase and regulates transcription via chromatin remodeling
[s] Binds to actin filaments in muscle and non-muscle cells
[t] Involved in the organization of the cytomatrix at the nerve terminals active zone (CAZ) which regulates neurotransmitter release
[u] Can degrade fibronectin, laminin, gelatins of type I, III, IV, and V; collagens III, IV, X, and IX, and cartilage proteoglycans
[v] The surface protein gp120 (SU) attaches the virus to the host lymphoid cell by binding to the primary receptor CD4
[w] Plays a role in the generation of cytotoxic lymphocyte effector function; the central coordinator for activation of many serine proteases in immune/inflammatory cells
[x] Component of the origin recognition complex (ORC) that binds origins of replication
[y] Receptor that may have an important role in cell/cell signaling during nervous system formation
[z] Plays a central role in the complement system and contributes to innate immunity
[aa] Actin-binding protein; plays a role in the activation of T-cells

Peptide sequences found using BLAST was used in a search of both the SwissProt and NCBI databases. Of the peptides found by *de novo* sequencing, 81% matched similar proteins in the two databases. The proteins were identified at single peptide level by observing the E-value, which is a statistical value indicating sequence similarity and potential evolutionary relationship.[243] A low E-value indicates a statistically significant match. The peptides listed in Table 3-2 are those returned from the BLAST search having the lowest three E-values (the complete Table can be found in Appendix A). Notably, the organisms associated with the low E-value peptides are close in homology to alligators, for example *Chelonia caretta* (turtle), *Caiman crocodilus* (caiman) and *Gallus gallus* (chicken). Eleven of the thirty-two *de novo* sequences had chicken protein sequences in the top three hits. Also, there were several identified proteins that are associated with leukocytes and the immune system. However, the identification of hemoglobin β suggests contamination from red blood cells during isolation of the leukocytes from the whole blood. Other peptides in the table matched proteins that have not, to date, been associated with the immune system.

To assess the quality of the protein matches, several factors were considered. First, the number of peptides matching the protein was considered. For example, three of the *de novo* sequences in Table 3-2 matched the protein fibrinonectin, which increases the confidence level that a similar protein is present in the alligator leukocyte. The percentage of matching residues was also considered. The *de novo* sequenced peptides matching fibrinonectin had 100% sequence alignment with that protein. However, the sequence coverage was only 3%.

Last, a BLAST search was performed using a limited taxonomy containing birds and reptiles. Limiting the taxonomy search reduced the random matches and resulted in lower E-values for the protein matches compared to the full database search. This data is shown in the supplementary material (Appendix A, Table 2).

The proteins were grouped based on their functionality. Protein functions were divided into six groups: other, cytoskeletal proteins, immune proteins, enzymes, DNA/synthesis proteins, and unknown. Proteins involved in the cytoskeletal system, immune system, and other make up the three most abundant groups of peptides with approximately 37%, 23% and 23% respectively.

De novo sequencing coupled with Mascot and BLAST searches aided in the identification of 43 proteins that are similar to the proteins isolated from the alligator's leukocyte extract. To identify and characterize the function of the proteins derived from the *de novo* sequenced peptides, two approaches were employed. First, if a protein was identified from more than one *de novo* sequenced peptide in the same gel band, the protein and its function were assigned. Second, *de novo* sequenced peptides with only one peptide matching a protein, the phylum of the matched protein was considered. From the three highest E-values from the BLAST search, seen in Table 3-2 (for the complete table, refer to Appendix A), E-values less than one were first considered and the organisms associated with the protein hit were classified by biological phylum and only animals were considered. Matched proteins that are from closely related species, e.g. reptiles and birds, were moved to the top of the list. Of the *de novo* sequenced proteins, 29% matched invertebrates 39% mammals, 18% birds, 6% amphibians, 5% fish and 3% reptiles. Although proteins from invertebrates are highly represented, it should be noted that 81% of the database proteins are associated with invertebrates. The taxonomic distribution of vertebrate proteins in the Swiss-Prot database is 50% mammal, 25% fish, 10% bird, 9% amphibian, and 6% reptile.[244] This can be compared to 55% mammal, 27% bird, 8% amphibian 6% fish, and 4% reptile, thus bird proteins are overrepresented in the matched proteins and fish proteins are underrepresented. This is consistent with evolutionary reports that birds have a close phylogenetic relationship to crocodilians.[245]

A putative function was assigned to the *de novo* sequenced proteins identified based on the function of the matched proteins. The majority of the proteins identified in the alligator leukocyte are common to proteins identified in leukocytes of other class of animals.

A range of proteins related to the immune system were identified. The identified immune system proteins were fibrinonectin, histone H4, Ig kappa chainV-1 region, Ig lambda chain V-III region, lysozyme C, acyl-CoA-binding protein, myeloid protein, cathepsin C, complement component c3, ubiquitin, and polyubiquitin. The protein ubiquitin was identified with 32% sequence coverage, the largest observed. This is consistent with previous reports demonstrating that ubiquitin is a highly conserved protein in eukaryotes.[246] Ubiquitin has several functions such as the removal of damaged or unnecessary proteins and controlling other cellular processes such as antigen processing, apoptosis, immune response and inflammation.[247] Although it has several functions, it is related to the immune system. A de novo sequenced protein matched lysozyme (*Chelonia mydas*) from residue 50 to 61. Lysozyme was one of the first antibacterial proteins isolated from insects and is also a component of the humans innate immune system.[248] The complement C3 protein also plays a role in innate immunity in organisms[248, 249] and plays a major role in the activation of the complement system, which helps to remove pathogens.[249] A *de novo* sequenced protein matched histone (*Gallus gallus)*, a class of proteins that are involved in DNA packaging; however, they have been reported to behave as antimicrobial agents.[250]

Several cytoskeletal proteins, myosin alkali light chain, myosin regulatory light chain, β-actin, actin-2, actin-1, annexin V, annexin I, ras-related protein O-krev, vimentin-4, talin 1, filamin-C, vinculin, microtubule-actin cross-linking factor 1, α-actinin-1, actin, plastin-2, and putative tropomyosin γ-3 chain-like protein were identified. These were the major group of proteins identified and play an important role in the function of leukocytes. Actin is the most

abundant protein in all eukaryotic cells[251] and is responsible for muscle contraction and cell motility and works in conjunction with myosin protein through actin-myosin interactions.[251, 252] Examples of cell movements include invasion of tissue by white blood cells to fight infection and migration of cells involved in wound healing.[252] Actin has multiple functions but is also related to the immune system; it regulates the integrin function, which is important for inflammation and immunity. Other identified cytoskeletal proteins, talin, α-actinin, plastin-2, and filamin-C, also play a role in the actin cytoskeletal system.[253-256] Variations of myosin and tropomyosin proteins work along with actin to assist in phagocytosis of microorganisms.[257] Annexins have various functions such as regulation of membrane organization and of Ca^{2+} concentrations within cells as well as participation in phagocytosis.[258] Vinculin, another cytoskeleton protein, is necessary for cell movement and changes in cell shape.[259] Vimentin, a filament protein, is involved in structural processes such as wound healing and is also found to be secreted by macrophages.[260] Also identified in this study was a ras-related protein; these proteins are GTPases that are a part of the Ras superfamily of proteins and are involved in regulating cell behavior such as cell growth and action.[261]

A class of enzymes known as endonucleases was also identified with high Mowse scores. These enzymes are commonly found in bacteria and behave like an immune system: bacteria use these enzymes to cut DNA from foreign material destroying them.[262] Although endonucleases are typically found in bacteria, alligator leukocyte may contain proteins that are similar to endonuclease protein and play a similar role. Alkaline phosphatase is a key enzyme involved in the process of dephosphorylation. It is also found to be involved in the protection against endotoxins, which may be lethal after an infection.[263]

Three matched proteins are hypothetical proteins from the western clawed frog, (*Xenopus tropicalis)*, house mouse (*Mus musculus*), and chimpanzee (*Pan troglodytes*). The *Pan*

troglodyte's hypothetical protein had four matching peptide sequences suggesting that a similar protein is expressed in the alligator leukocyte. However, the function of these hypothetical proteins is not known and their relationship to the immune system cannot be determined.

There were proteins identified from the Mascot search that had a score greater than 57 but only one matching peptide: lysozyme C and ras-related protein O-krev (Table 3-2). These proteins cannot be identified with the same degree of confidence as the other proteins that had more than one peptide matching the protein. The proteins characterized as other have not been previously identified in leukocytes of other organisms. However, the E-values are high which suggests limited sequence identity and evolutionary relationship.

3.4 Summary

In the work described in this Chapter, proteins from the leukocytes of the American alligator (*Alligator mississippiensis)* were identified and their functions were characterized by adopting a general proteomics approach. A three-step strategy was performed to identify similar proteins in the gel bands and spots: *de novo* sequencing, Mascot search, and BLAST search. Mascot search results identified eighteen proteins with sequence similarity to the alligator leukocyte proteins. The identified proteins are common among eukaryotes and are associated with the immune system. The BLAST search results showed that 81% of *de novo* sequenced proteins matched similar proteins in the database. Among the identified proteins, the majority were related to the cytoskeletal system.

The BLAST search identified proteins most closely related to vertebrates. Proteins originating from birds, particularly chicken, were overrepresented in the matched proteins and correlated with immune related proteins reported for other species. Fish species was the least represented species among the matched proteins.

CHAPTER 4. SMALL MOLECULE INTERFERENCES IN *ALLIGATOR MISSISSIPPIENSES* LEUKOCYTE EXTRACT

In this chapter, experiments in which small molecule interferants in the leukocyte extracts from *Alligator mississippiensis* were separated and purified using reversed phase liquid chromatography are described. Seven bacterial species including *Escherichia coli, Enterobacter cloacae, Shigella flexneri, Klebsiella oxytoca, Salmonella, Staphylococcus aureus, and Staphylococcus epidermidis* were assessed for antimicrobial activity. The active fractions were characterized using tandem mass spectrometry and X-ray crystallography. The results from the mass analysis and X-ray crystallography suggest that the component exhibiting antimicrobial activity was ethylenediaminetetraacetate (EDTA), the anticoagulant agent used during the collection of the alligator blood. Tandem mass spectrometry data along with accurate mass measurements suggests that the other small molecule interferant is spermine.

4.1 Introduction

Antimicrobial drug resistance is a growing problem. Numerous bacteria such as *Staphylococcus aureus*, *Salmonella*, *E. coli*, *Pseudomonas aeruginosa,* and *Streptococcus pneumoniae* have developed resistance to antibiotics.[264] One of the most resistant bacteria affecting hospital settings is *Staphylococcus aureus*: over 50% of its bacterial strains have become resistant to penicillin.[264, 265] Due to this problem, new antibiotics with broad spectra of activity must be developed.

Studies have shown that when extracts of alligator leukocytes from blood are exposed to bacterial pathogens, the growth of the bacteria is inhibited.[266] Several anticoagulant agents that prevent blood clotting are available, such as EDTA, heparin, citrate, and oxalate. EDTA is a chelating agent that synergistically inhibits growth of bacteria with antimicrobial agents[267] such as sulfamethoxazole, trimethoprim,[268] penicillin G, ampicillin, tetracycline, and

chloramphenicol.[269] Studies have shown that EDTA can also inhibit bacterial growth independently.[270] EDTA enhances the effectiveness of antibiotics against bacteria, particularly gram-negative bacteria, and chelates Ca^{+2} and Mg^{+2} which disrupts the bridge between them and lipopolysaccharide (LPS) in the bacterial outer membrane.[267, 271]

Peptides and small molecules have the potential for therapeutic use. Small molecules are less costly and easy to synthesize and are stable, therefore longer half-life.[272] However, small molecules can potentially interact with multiple targets and accumulate in tissues causing toxicity.[272] Unlike small molecules, peptides are more specific and can be chemically modified to make the peptide less susceptible to proteolysis *in vivo*.[272]

Another small molecule that exhibits antimicrobial activity is spermine. Spermine is a biological polyamine that is found to be widely distributed in animal tissues.[273] It has antibacterial activity, providing resistance to a wide range of microorganisms such as *S. aureus, S. albus,* and *B. anthracis*.[273, 274] Spermine was first isolated from semen and exhibits antimicrobial activity against gram-positive microorganisms.[275]

In this chapter, we report that EDTA and spermine are present and isolated from leukocyte extracts of alligators. These molecules can potentially cause interferences with the activity measurements.

4.2 Experimental

Leukocyte extracts were obtained from the blood of alligators as described in Section 2.1. In this study, which was the first studies performed on isolation of peptides from alligator leukocytes, 0.1 mM EDTA was used as the blood anticoagulant.

A Bradford protein assay was used to determine the concentration of proteins in the alligator leukocyte extracts. A 2 mg/mL quantity of bovine serum albumin (BSA) was used to prepare five BSA dilutions from 125 to 2000 µg/mL. A 0.1% acetic acid buffer was used to

perform the dilutions. Each standard and leukocyte protein extract was pipette into clean test tubes and mixed with the protein dye reagent. The protein samples were assayed in triplicate and incubated for 5 min before UV-Vis analysis. Absorbance was measured at 595 nm for both the protein standard and leukocyte protein extracts. The concentration of the leukocyte protein mixture was determined using a standard Bradford curve.[276]

The leukocyte extract was tested for antimicrobial activity before chromatographic separation. A radial diffusion assay was used to monitor bacterial growth inhibition.[277] Leukocyte extracts and chromatographic fractions of the leukocyte extracts were pipetted into wells on the activity plate, incubated overnight at 37ºC, and zones of bacterial growth inhibition were measured the following day.

Alligator leukocyte extracts were separated by reverse phase liquid chromatography (RP-LC) and peak fractions were monitored with a photodiode array (PDA) detector and manually collected in 2 minute intervals. For PDA detection, the UV spectrum was recorded between 190 and 400 nm. The fractions were tested for antimicrobial activity and the active fractions were lyophilized and re-suspended in 0.01% (v/v) aqueous acetic acid. The fractions that exhibited antimicrobial activity were injected onto the C_{18} reversed phase (RP) column using a gradient of 10–90% acetonitrile over 40 minutes for further purification and analyzed using ESI-TOF mass spectrometry.

The purified fraction that exhibited antimicrobial activity was submitted for structural characterization using X-ray crystallography. Tandem mass spectrometry (MS/MS) using the nano ESI-Q-TOF was also performed to obtain structural information on the molecules.

4.3 Results and Discussion

4.3.1 Separation of Leukocyte Extract

The alligator leukocyte was separated into 35 fractions using C_{18} reversed-phase chromatography (Figure 4-1). The fractions from the leukocyte extract were collected and the initial fractions collected within the first six minutes (labeled F2 and F3) were tested for antimicrobial activity.

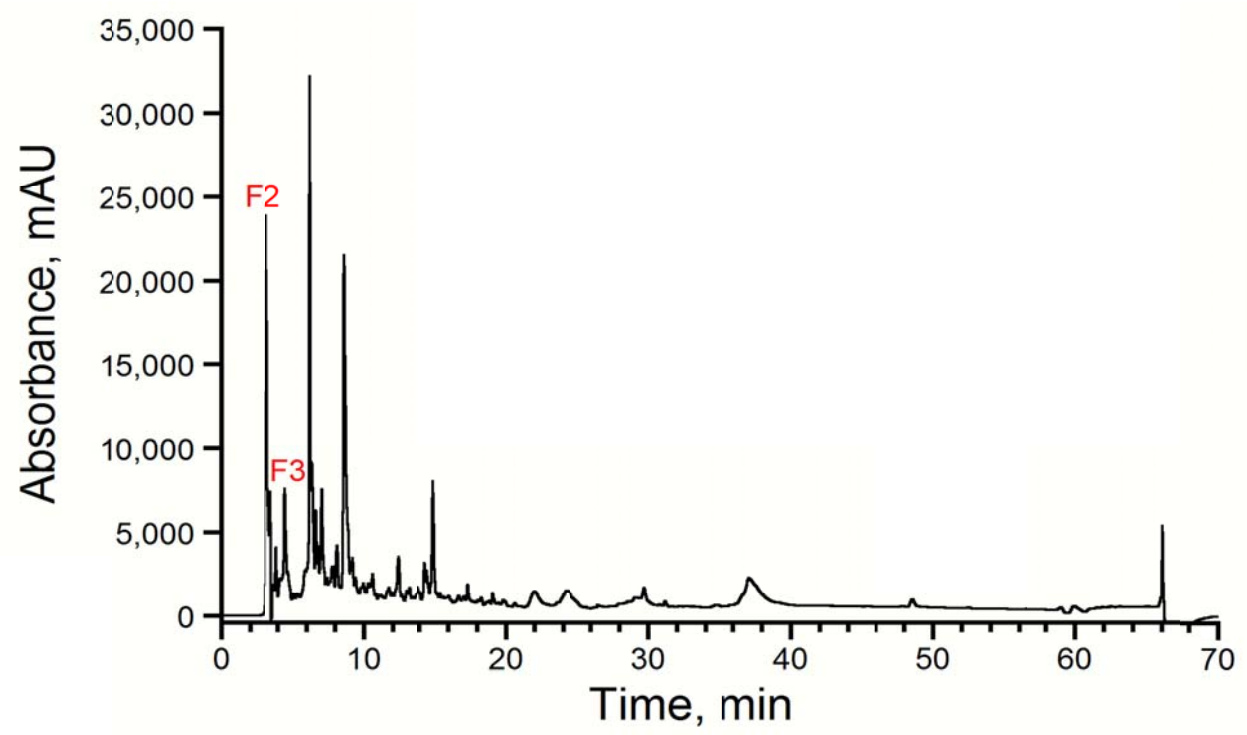


Figure 4-1. Separation of alligator leukocyte using reversed-phase separation and monitored between 190 and 400 nm.

The collected fractions F2 and F3 were further purified using RP chromatography as seen in Figures 4-2 (A) and (B).

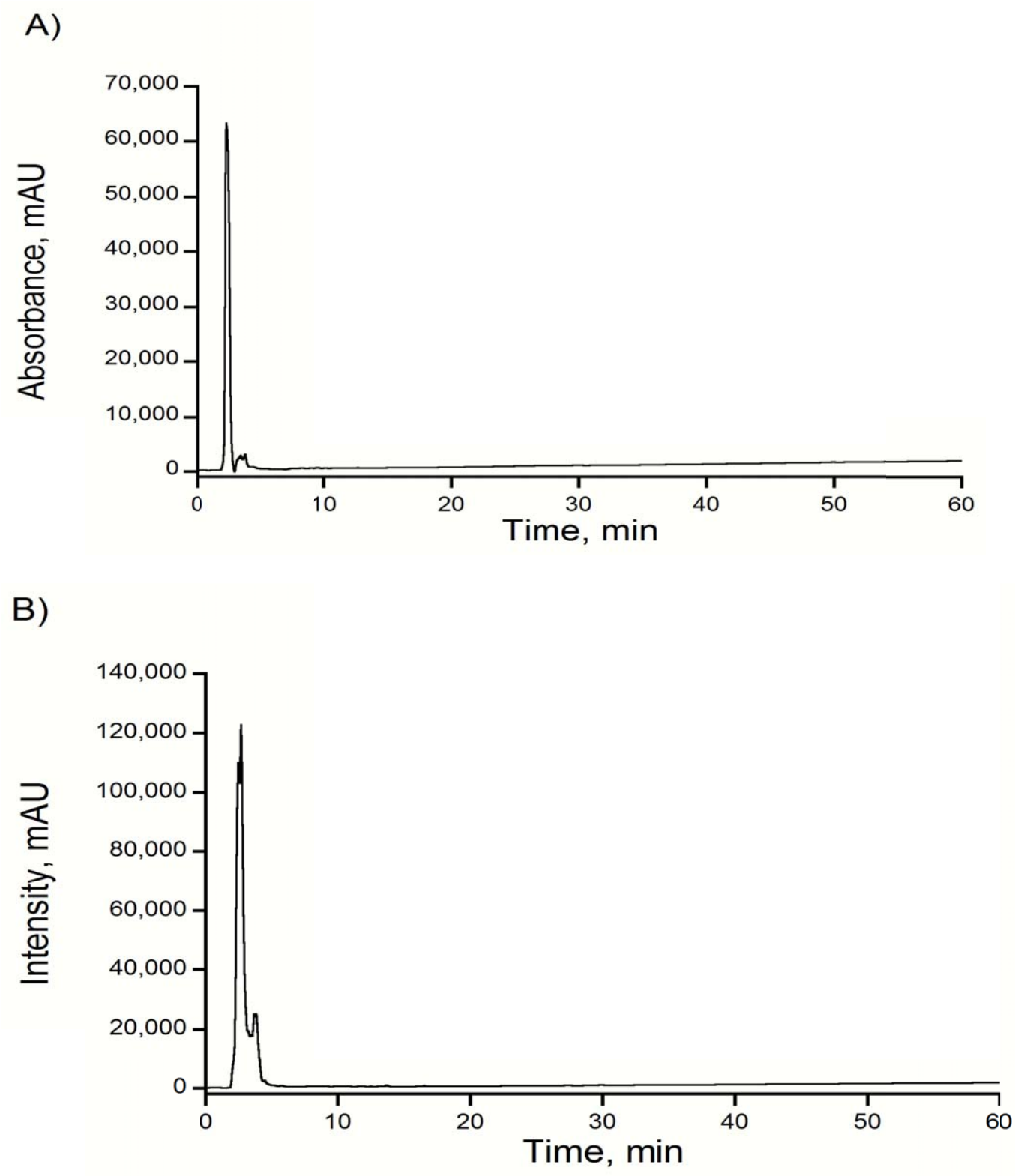


Figure 4-2. (A) Purification of Fractions 2 and (B) Fraction 3 using reversed-phase separation.

4.3.2 Antimicrobial Activity

Results from this study showed that Fractions 2 and 3 exhibited antimicrobial activity against *Escherichia coli, Shigella flexneri, Salmonella, Staphylococcus aureus*, and *Staphylococcus epidermidis* (Table 4-1). When tested against *Klebsiella oxytoca,* a gram-negative bacterium, neither of the fractions exhibited activity. However, growth inhibition was observed for Fraction 2 against *Enterobacter cloacae*, a gram-negative bacterium, but not for Fraction 3. The largest antimicrobial activities from Fractions 2 and 3 were observed for *Staphylococcus epidermidis*, a gram-positive bacterium; zones of inhibition were 17 and 14 mm respectively. The lowest activity was observed in *Salmonella*, a gram-negative bacterium, at 5 and 4 mm respectively.

Table 4-1. Antibacterial activity of Fractions 2 and 3 from leukocyte extract against bacterial species tested using radial diffusion assay.

Bacterial Strain	**Zone of Inhibition (mm) (F2)**	**Zone of Inhibition (mm) (F3)**
Escherichia coli	10.0	5.0
Enterobacter cloacae	7.0	0
Shigella flexneri	6.0	5.0
Klebsiella oxytoca	0	0
Salmonella	5.0	4.0
Staphylococcus aureus	15.0	10.0
Staphylococcus epidermidis	17.0	14.0

Both fractions exhibited antimicrobial activity towards both gram-negative and gram-positive bacteria; however growth inhibition was highest for gram-positive bacteria.

The negative control, 0.01% acetic acid, showed no antibacterial activity and the positive controls, polymyxin B and synthesized EDTA and spermine showed antibacterial activity against the tested microbes.

4.3.3 ESI-MS Analysis

ESI-MS was used to analyze fractions F2 and F3 as seen in Figure 4-3 (A) and (B). Because of the high mass accuracy of the ESI-TOF instrument it was possible to assign chemical formulas to the peaks present in the mass spectrum. The ion at *m/z* 203.2250 was observed in both fractions F2 and F3 and was the most abundant ion present. The molecular formula, $C_{10}H_{26}N_4$, for the molecule measured at *m/z* 203.2250 was confirmed with a mass accuracy of 6 ppm.

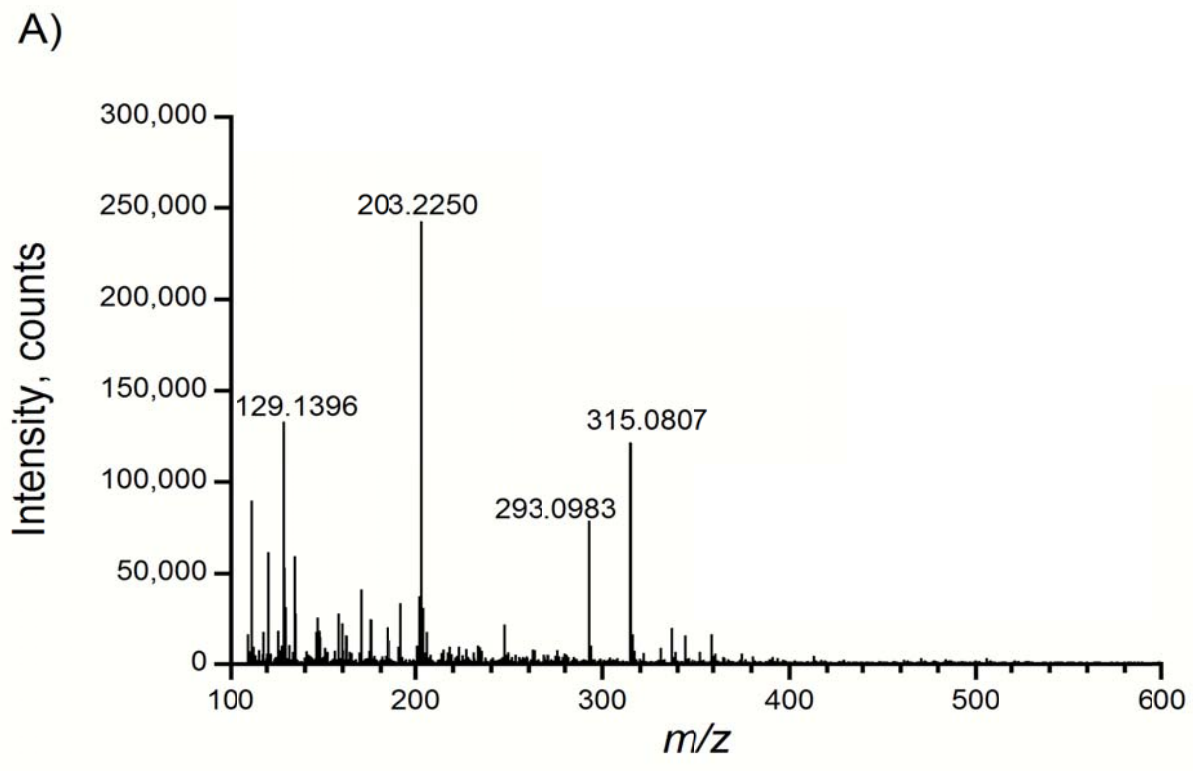
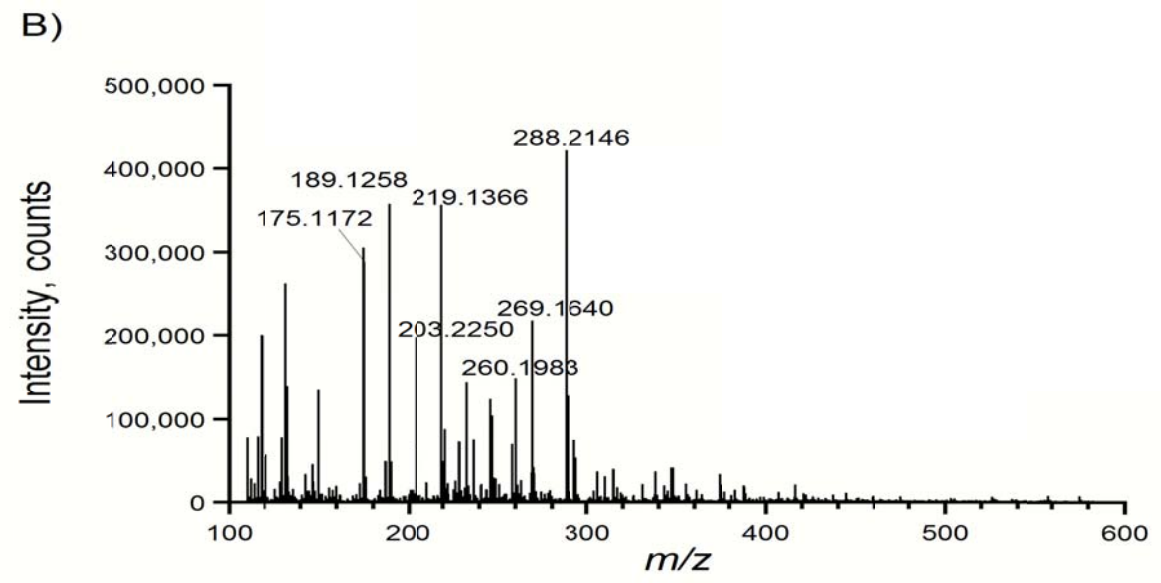


Figure 4-3. ESI-MS of fractions (A) F2 and (B) F3 isolated from alligator leukocyte extract.

Figure 4-3. cont'd.



Another intense peak in Fraction 2 was at *m/z* 315.0807 with a mass accuracy of 3 ppm, corresponding to molecular formula $C_{10}H_{16}N_2O_8Na$. However, this molecule was absent in Fraction 3. This suggests that this molecule could have contributed to the antimicrobial activity of Fraction 2 because growth inhibition was observed against *Enterobacter cloacae* from Fraction 2 but not from Fraction 3.

Previous studies have shown that antimicrobial components isolated from plants, insects and mammals are often peptides with 12–60 amino acids and have masses less than 10 kDa.[24] The results from this study show that the antibacterial properties of the fractions are either 1) produced by small molecules with masses less than 400 Da, 2) produced by larger molecules that are present at concentrations below the mass spectrometer detection limit, or 3) produced by small and large molecules acting synergistically (as noted above) with the larger molecule not detected in the mass spectrometry analysis.

Tandem MS of the small molecules identified from MS analysis was performed to identify their structure. Figure 4-4 shows the MS/MS data for the product ions 315.1 and 203.2Da.

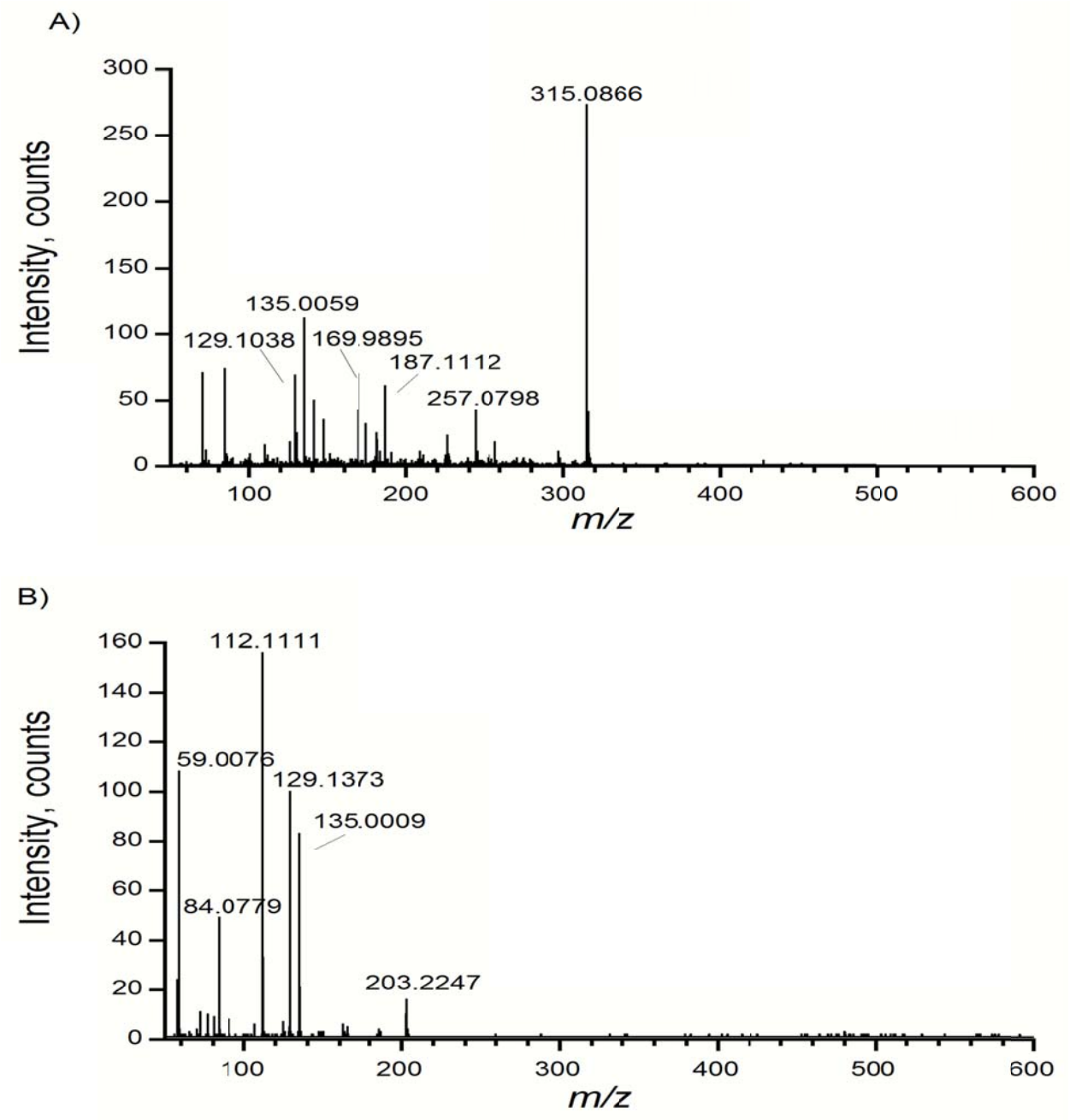


Figure 4-4. Tandem mass spectrometry data of (A) the *m/z* 315.1 ion from Fraction 2 and (B) the 203.2 ion from Fractions 2 and 3.

The MS/MS results for the ion at *m/z* 203.2250 showed fragments previously reported for spermine, $C_{10}H_{26}N_4$, by Q-TOF analysis.[278] These fragments are based on the cleavages of the carbon and nitrogen chain, as seen on Figure 4-5.

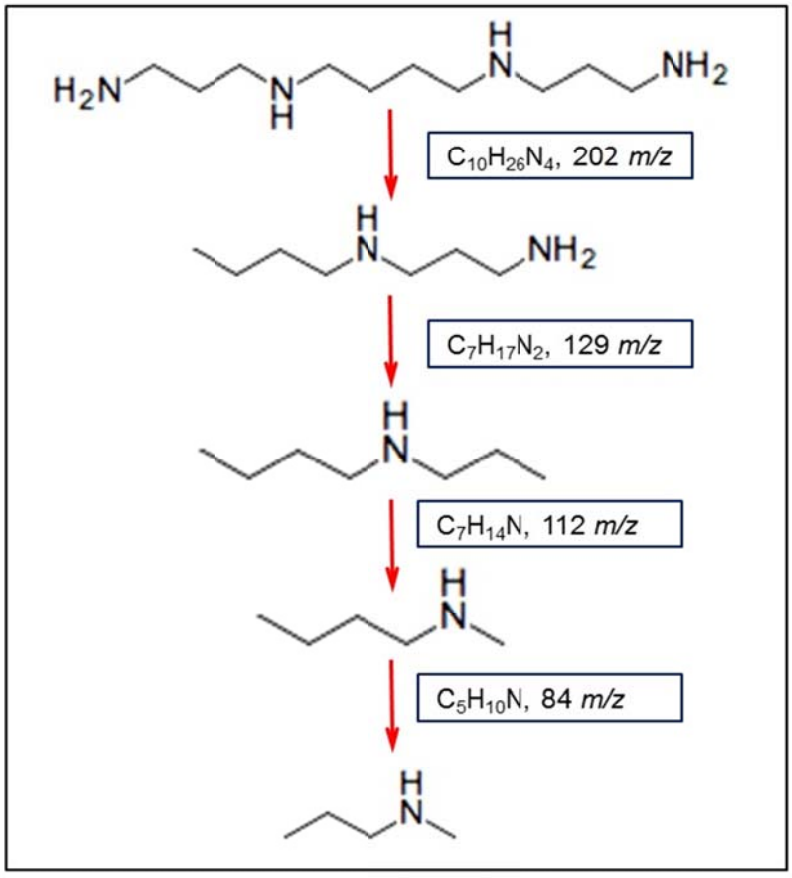


Figure 4-5. Fragmentation of spermine.

The fragments observed from the ion at *m/z* 203.2250 in Figure 4-4 (B), for example 84.1 112.1, and 129.1, are characteristic of the fragmentation pattern of spermine ion. The tandem MS data are consistent with the molecular formula obtained from the MS accurate mass results. Results from the MS/MS spectrum of the ion at *m/z* 315.1 ion did not allow for structure assignment.

4.3.4 X-ray Crystallography

Fraction 2 was dried and formed crystals that were analyzed using X-ray crystallography. The crystallographic data was searched in the Cambridge Structural Database for organic and organometallic compounds and eight hits of possible structures were generated, including phenylammonium nitrate ($C_6H_8N_2O_3$), EDTA disodium salt $(C_{10}H_{18}N_2Na_2O_{10})^n$, 3a,4,7,7-Tetramethyl- (3aR,4S,6aR,10aS)-perhydronaphtho(8a, 1-b)furan-2-one ($C_{16}H_{26}O_2$), 7-Iodo-2,9-dimethyl-4,4-ethylenedioxy-11-oxatricycloundec-2-ene ($C_{14}H_{19}I_1O_3$), 7-Methoxy-4,9,13-trimethyl-6-oxa-1,13:10,11-diepoxytricyclotridec-3-en-5-one ($C_{16}H_{20}O_5$), catena-((μ_2-Benzene-1,3-dicarboxylato)-aqua-(propane-1,3-diamine)-copper(ii)monohydrate) ($C_{11}H_{16}Cu_1N_2O_5$)n,n(H_2O)), catena-(tetra-sodium bis(μ_2-methylenediphosphonato)-di-copper(ii)dehydrate) (($C_2H_4Cu_2O_{12}P_4^{4-}$)n,4n(Na^{1+}),2n(H_2O)), and 3-(4-Methoxybenzylidene)-2-oxo-2,3-dihydro-1H-imidazo(1,2-a)pyridiniumbromide ($C_{15}H_{13}N_2O_2$). Based on the crystal structure, seen on Figure 4-6, and the crystal's space group values, the structure was a best match for EDTA disodium salt.

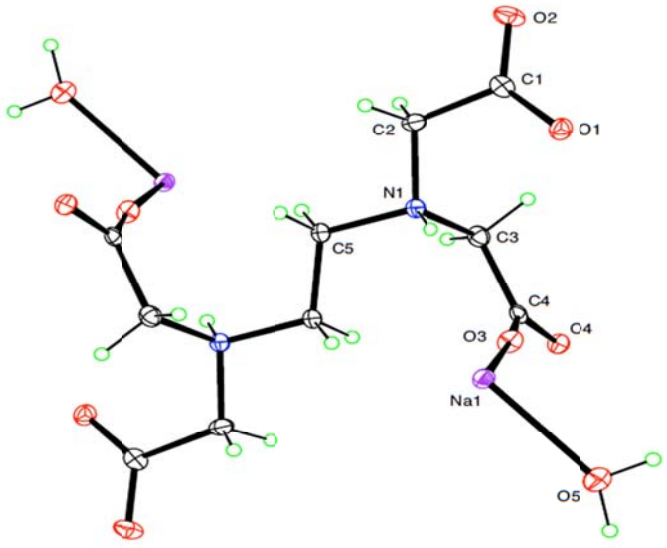


Figure 4-6. Crystal structure of EDTA disodium salt obtained from alligator leukocyte extract.

4.4 Summary

In the work described in this Chapter, leukocyte extracts were subjected to purification and characterization by LC ESI-MS, MS/MS and X-ray crystallography analysis. The results from this study suggest that EDTA, a blood anticoagulant, in the presence of a protein might contribute to inhibition of bacterial growth. In addition, previous studies have shown that EDTA contributes to bacterial growth inhibition at both high and low concentrations.[270] An alternate explanation is that the antimicrobial activity resulted from a molecule that was not detected in the mass spectrometry analysis. RP separation was capable of reducing the complexity of the leukocyte protein mixture; however, the chromatographic fractions were still not sufficiently pure to identify the specific molecules that contributed to antimicrobial activity. Additional off-line RP separations assisted in small molecule purification and antimicrobial activity was identified for fractions collected within the first six minutes. The masses of the small molecules detected in the active fractions were measured by ESI-MS and the small molecules were identified by empirical formula calculations based on accurate mass measurement. A molecule, from the active fractions, F2 and F3 at *m/z* 203.2 had the formula, $C_{10}H_{26}N_4$ with 6 ppm mass error. In addition, tandem MS analysis revealed fragmentation patterns for this ion that matched the molecule spermine. Another molecule from fraction F2 at *m/z* 315.1 (3 ppm accuracy) had the formula $C_{10}H_{16}N_2O_8Na$ and X-ray crystallography was used for further characterization where results from the structural database identified the structure from the active fraction as EDTA disodium salt.

The results from this study suggest that EDTA may not be a suitable anticoagulant agent for isolation of blood proteins when testing for antimicrobial activity. As a result, for future studies the anticoagulant agent heparin was used as a substitute. The other small molecule that was identified was spermine, which is typically found in all animal cells and is usually excreted

from semen and pancreatic tissue. Blood contains various proteins that are secreted or shed from cells and tissues throughout the body.[78] Therefore, it may not be possible to eliminate spermine as an interferant, hence for the studies described in the following chapters, small molecules eluting within the first six minutes from the leukocyte extracts were not considered as potential candidates for antimicrobial molecules.

CHAPTER 5. ISOLATION AND *DE NOVO* SEQUENCING OF ANTIMICROBIAL PEPTIDES FROM LEUKOCYTES OF THE AMERICAN ALLIGATOR (*ALLIGATOR MISSISSIPPIENSIS)*

The goal of the research in this chapter was to isolate novel peptides with antimicrobial activity from American alligator blood (*Alligator mississippiensis).* Reversed-phase high performance liquid chromatography (RP-HPLC) was used to separate and purify peptides from leukocyte extracts and the HPLC fractions were checked for activity. A partially purified fraction containing two major peptides at 4.7 and 4.9 kDa was tested for antimicrobial activity against *E. coli*, *E. cloacae*, and *K. oxytoca.* The peptides were analyzed using ion mobility spectrometry mass spectrometry. Matrix assisted laser desorption ionization time-of-flight/time-of-flight mass spectrometry (MALDI-TOF/TOF) was used to generate MS/MS data for the peptides and *de novo* sequencing was used to determine their sequence. The 4.7 kDa peptide was aligned to other antimicrobial peptides and showed 36% sequence identity with Pilosulin 4 and the 4.9 kDa peptide showed 33% sequence identity to SpStrongylocin 2.

5.1 Introduction

Increasing bacterial resistance to antibiotics has become the driving force behind the development of new anti-infectives to replace or supplement conventional antibiotics.[279] Antimicrobial proteins and peptides from natural sources show promise as new antibiotics. Antimicrobial peptides (AMP) are important components of the innate immune system that have been isolated in plants,[280] vertebrates,[281] and invertebrates[282] and have potential for pharmaceutical use. Most antimicrobial peptides are small (<10 kDa), cationic, lysine and arginine containing molecules; however, some anionic peptides have also been reported.[25] Cationic proteins and peptides have an affinity for microbe cell surfaces, which have negatively charged groups. This attraction is dependent on electrostatic interactions and their tertiary

structure.[283] Cationic proteins and peptides vary in their size, structure, and mechanism of action.[283]

In some vertebrates such as mammals and reptiles, antimicrobial proteins and peptides are found in the blood, particularly in the neutrophils. There are several classes of antimicrobial proteins and peptides associated with neutrophils, such as lactoferrin, cathelicidins, calprotectin and defensins.[283] Defensins are small cationic peptides around 4 kDa with a broad antimicrobial activity against bacteria, fungi, viruses, and parasites.[284] β-defensins serve as effector molecules of innate immunity, providing an initial defense against infectious agents,[285] and have been identified in bovine neutrophils,[286] turtle[33] and avian leukocytes,[287] and human plasma.[288] Leukocytes in the blood of crocodilian species have a range of antimicrobial, antifungal, and antiviral activity,[1, 289, 290] but little is known about antimicrobial peptides in *Alligator mississippiensis*.

There is limited information about reptile antimicrobial peptides. To date there are very few reptilian antimicrobial peptide sequences identified among the nearly 400 that have been reported from other sources.[33, 36] The *Alligator mississippiensis* leukocytes have a broad range of activity towards bacteria, viruses and fungi.[1] Some studies have been performed to characterize the proteins and peptides from alligator leukocytes;[291] however, none have been sequenced.

De Novo sequencing is a common approach for mass spectrometry based sequencing of peptides and proteins that are not in a protein sequence database. *De Novo* sequencing entails identifying peptide sequences without any previous knowledge of the protein sequence.[164]

In this study, we isolated two novel peptides that exhibit antimicrobial activity from the leukocytes of the American alligator (*Alligator mississippiensis)* using reversed-phase chromatography separation. The peptides were sequenced using *de novo* sequencing and compared to other known antimicrobial peptides.

5.2 Experimental

5.2.1 Isolation and Extraction of Antimicrobial Peptides

Blood was drawn from farm and wild alligators ranging in size from 5 to 12 feet for isolation of leukocytes using an approach described in Section 2.1.

5.2.2 Microbial Strains

All bacterial strains were obtained from American Type Culture Collection (ATCC). The following bacterial strains were used: *Escherichia coli* (ATCC 43827), *Enterobacter cloacae* (ATCC 23355), and *Klebsiella oxytoca* (ATCC 33496).

5.2.3 Antimicrobial Activity

The antimicrobial activity of the leukocyte extracts and the peptide fractions from HPLC eluents were determined using an antimicrobial assay.[278] Trypticase soy broth containing 1% agarose was sterilized and 15 mL aliquots were transferred to centrifuge tubes. The nutrient broth was inoculated with 20 µL of a log phase culture of bacteria and added to petri dishes. Wells were cut into the petri dish using a sterilized pipette with a vacuum line attached for suction. Five microliters of leukocyte extract dissolved in aqueous 0.1% acetic acid (v/v) was added to each well and allowed to diffuse at room temperature for 3 hours. The plates were then incubated at 37^{o}C overnight and the zones of bacterial growth inhibition were measured. The antibiotic polymyxin B at 2×10^{-5} M was used as a positive control and aqueous 0.1% acetic acid (v/v) was used a negative control.

5.2.4 Peptide Separation

Leukocyte extract in 0.1% acetic acid was centrifuged to remove any insoluble residue. A 10 µL volume of concentrated sample was injected with an autosampler system onto a 0.3 × 1 mm trapping column on a LC system (Agilent) at a flow rate of 0.3 ml/min. The sample was eluted onto a 2.1 × 30 mm, 3.5 µm particle size C_{18} column (Zorbax) and separated using a

gradient of 5 to 60% B (80% acetonitrile (v/v) and 20% water (v/v) containing 0.1% formic acid (v/v)) over 70 min with a flow rate of 0.3 mL/min. The eluent was monitored by UV absorbance from 190 to 480 nm. A fraction collector was used to collect the eluent at timed intervals for further characterization and activity studies.

5.2.5 Ion Mobility Mass Spectrometry

The peptide samples were diluted in 0.1% formic acid and injected on the nanoAcquity ultra pressure liquid chromatography (UPLC) system at a flow rate of 500 nL/min. The peptides were separated using a gradient of 3–40% B over 30 minutes. The peptides were ionized using a nano-electrospray source, operated in positive ion mode and analyzed via a quadrupole mass analyzer before being introduced to the ion mobility spectrometry (IMS) on the ESI-Q/IMS/TOF MS. The quadrupole was operated in RF mode in order to record full mobility spectra and the data was processed using MassLynx software. A pressure of 0.5 bar of nitrogen was maintained in the ion mobility (IM) cell.

5.2.6 MALDI Mass Spectrometry

Alligator leukocyte fractions were mixed on a stainless steel plate using a 1:1 ratio of 1 µL sample and 1 µL saturated matrix including α-cyao-4-hydroxycinnamic acid (CCA) and 2-hydroxy-5-methoxybenzoic acid (sDHB) matrix solution containing 0.1% v/v TFA for molecular mass determination. The sample and matrix solution was air-dried at room temperature and analyzed in TOF mode on the MALDI TOF/TOF MS. Mass spectra were recorded in reflector positive ion mode with external calibration using a standard calibration mixture. Spectra were 500 shot averages collected using a mass range from 420–5000 Da. Masses were determined using FlexAnalysis software. MS/MS data was obtained by LIFT/CID fragmentation. MS/MS experiments were carried out in reflectron mode and more than 5000 laser shots were collected. The MS/MS data were processed in FlexAnalysis.

5.2.7 ESI Mass Spectrometry

The fraction containing the two peptides of interest was also analyzed on the ESI-QTOF. The peptide sample was diluted twofold with 50% ACN solution containing 0.1% FA (v/v). MS/MS experiments were performed on the selected parent ion and the collision energies were varied from 25 to 60 eV in 5 eV intervals.

5.2.8 *De Novo* Sequencing and Sequence Analysis

De Novo sequencing of the peptide fragments generated using MALDI MS/MS was performed with BioTools. The *de novo* sequencing results were verified manually. The peptide sequences produced by *de novo* sequencing were aligned with β-defensins peptide sequences from other vertebrates for sequence comparison.

The alligator antimicrobial peptide sequence was compared to other antimicrobial peptides in the Antimicrobial Peptide Database.[292] The alligator antimicrobial peptide was also aligned to find the most similar peptides in the database.

5.3 Results and Discussion

5.3.1 Antimicrobial Activity of Leukocyte Extracts

The antimicrobial activity of the alligator leukocyte extracts was tested against three species of bacteria. The leukocyte extracts exhibited antimicrobial activity against gram negative bacterium *E.coli*, *K. oxytoca,* and *E.cloacae* with approximately 0.5 mm zone of inhibition.

5.3.2 Peptide Purification

Several RP separations were run to separate and collect the active peptides. The separation of the peptide and protein mixture from the alligator leukocytes produced more than 12 fractions (Figure 5-1). The fraction collected at 15 min at 20% acetonitrile showed antimicrobial activity against the tested bacteria. The deconvoluted masses from the ESI data

(Figures 5-1B and 5-1C) showed at least three peptides present in the active fraction at 4618.3, 4745.4, and 4961.9 *m/z*.

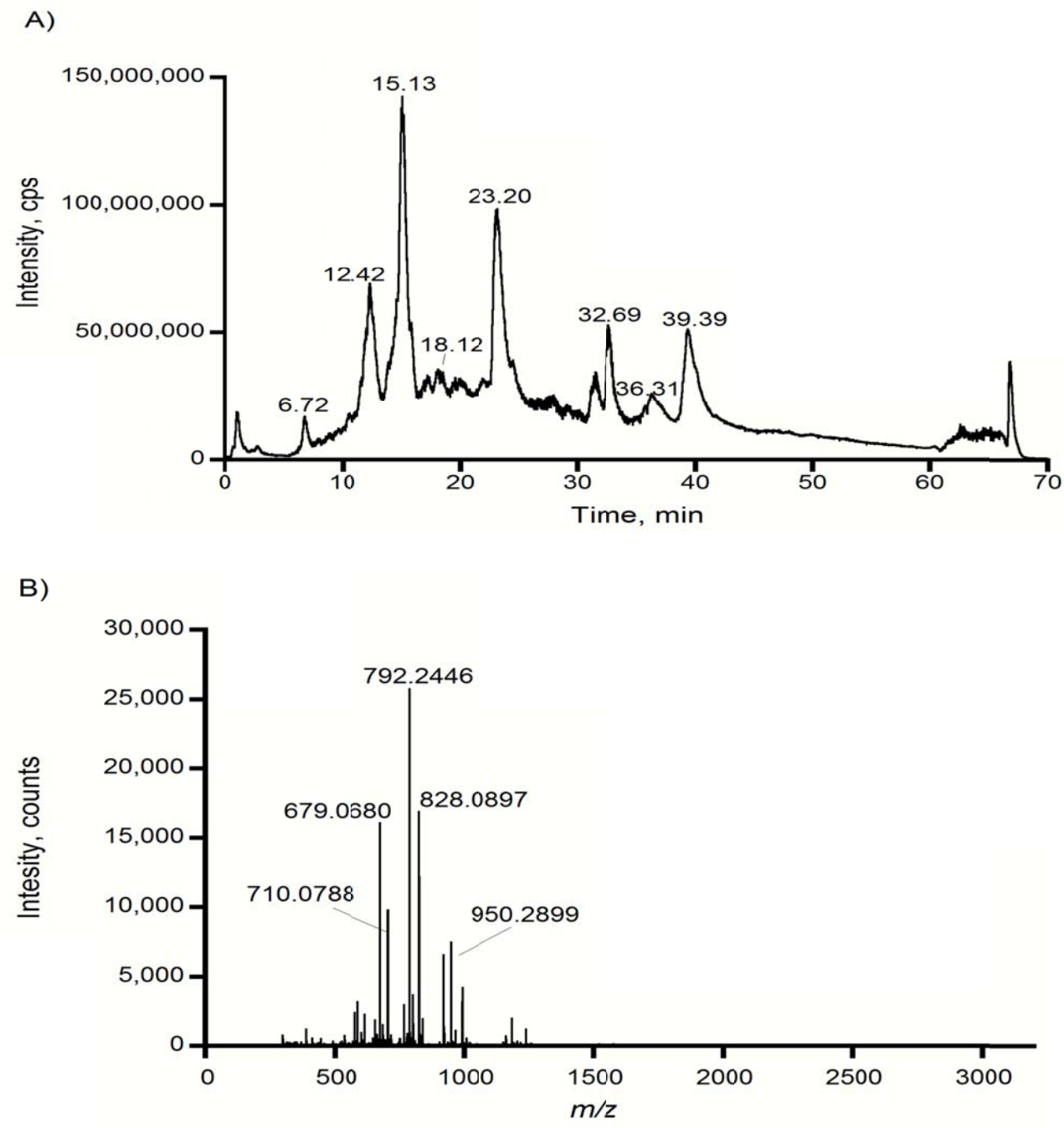
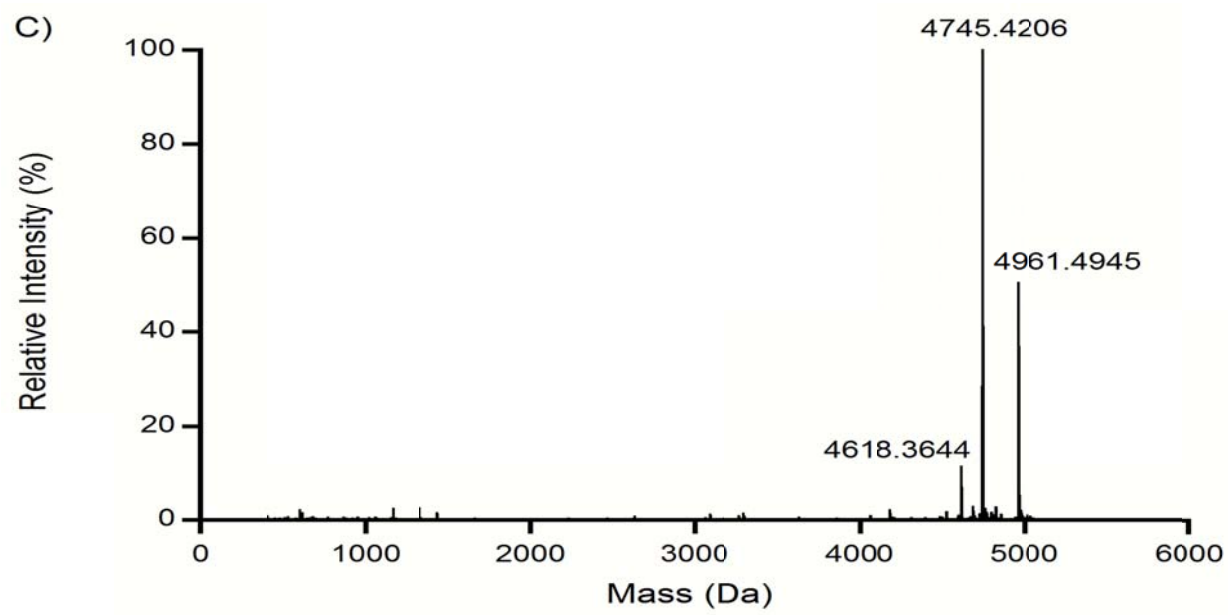


Figure 5-1. (A) The leukocyte extract in 0.1% acetic acid was separated using RP-HPLC on a C_{18} Vydac column. (B) Mass spectrum of the biologically active fraction eluted at 20% acetonitrile. (C) Deconvoluted mass spectrum of the peptides eluted at 15% acetonitrile.

Figure 5-1. cont'd.



In addition to the major peptides identified in the active fraction, there were other minor impurities in the lower mass range which may suggest the presence of other antimicrobial peptides in lower quantities.

5.3.3. Ion Mobility Mass Spectrometry of the Active Peptides

The two peptides from the alligator leukocyte extract at 4745.4 and 4961.9 Da have a mass difference of ~200 Da; therefore, nano liquid chromatography-ion mobility spectrometry mass spectrometry (LC-IMS) was used to separate them based on their charge, size and shape. The base peak chromatogram seen in Figure 5-2 corresponds to multiple peptides with similar retention times, indicating that these two peptides coelute. Figures 5-3 and 5-4 show plots of *m/z* versus drift time and *m/z* versus retention time of the ions separated by ion mobility.

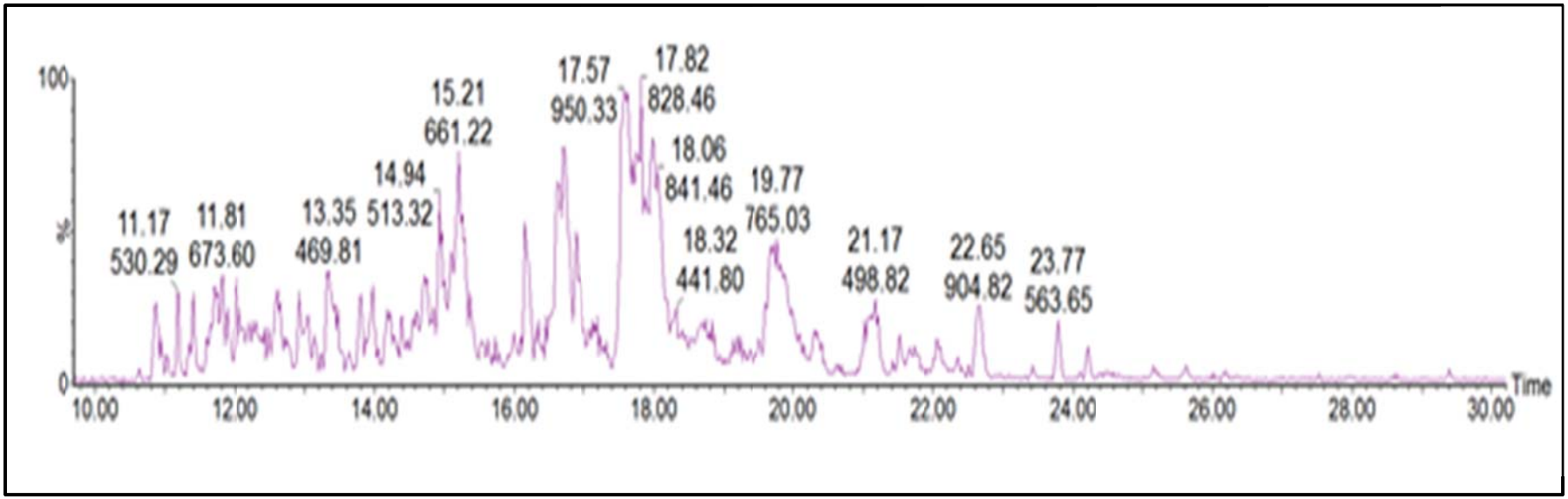


Figure 5-2. Base peak chromatogram of the nanoflow LC-IMS experiment for a mixture of peptides from alligator leukocyte extracts. The values above the peaks represent both the elution time and *m/z* value for ions measured at the respective times.

(A)

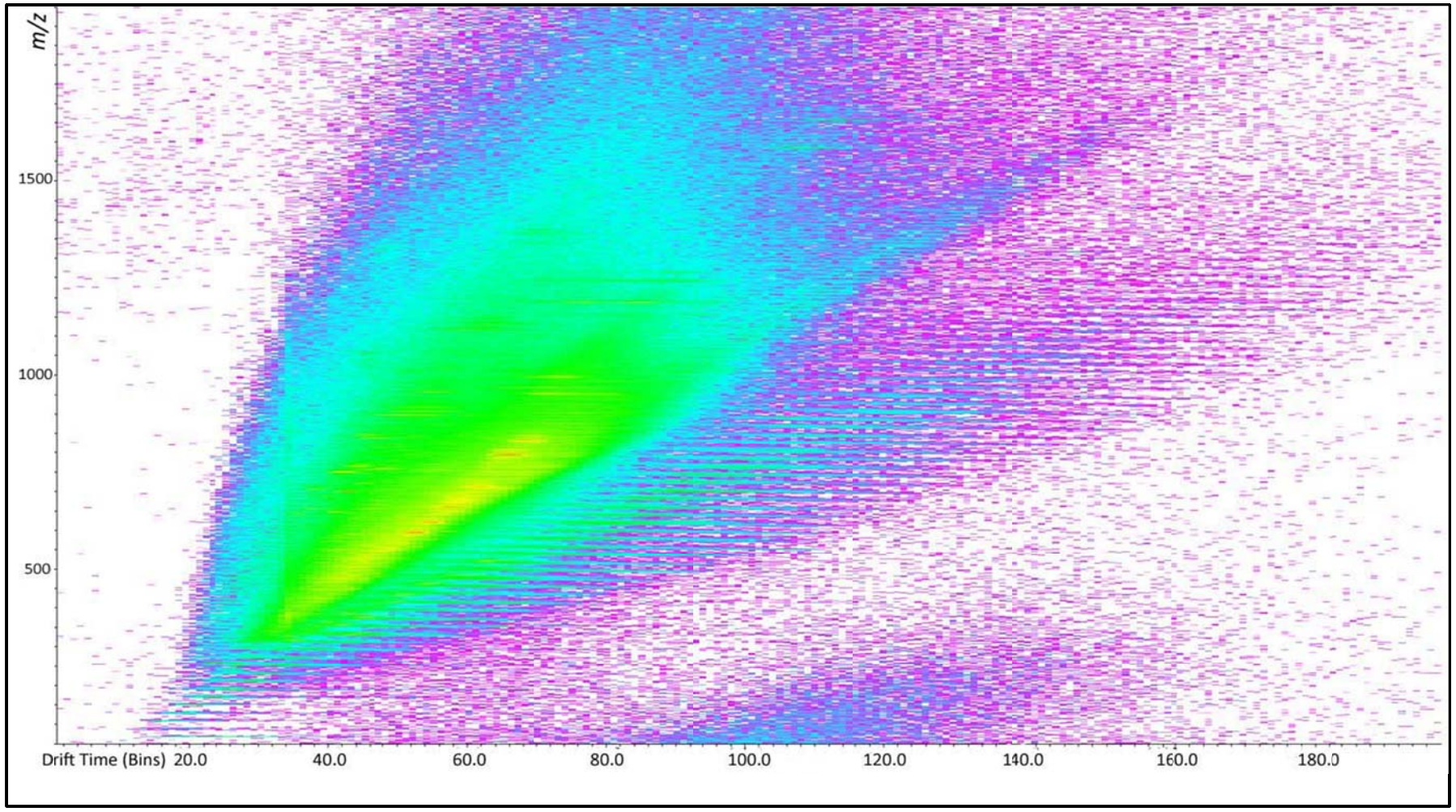
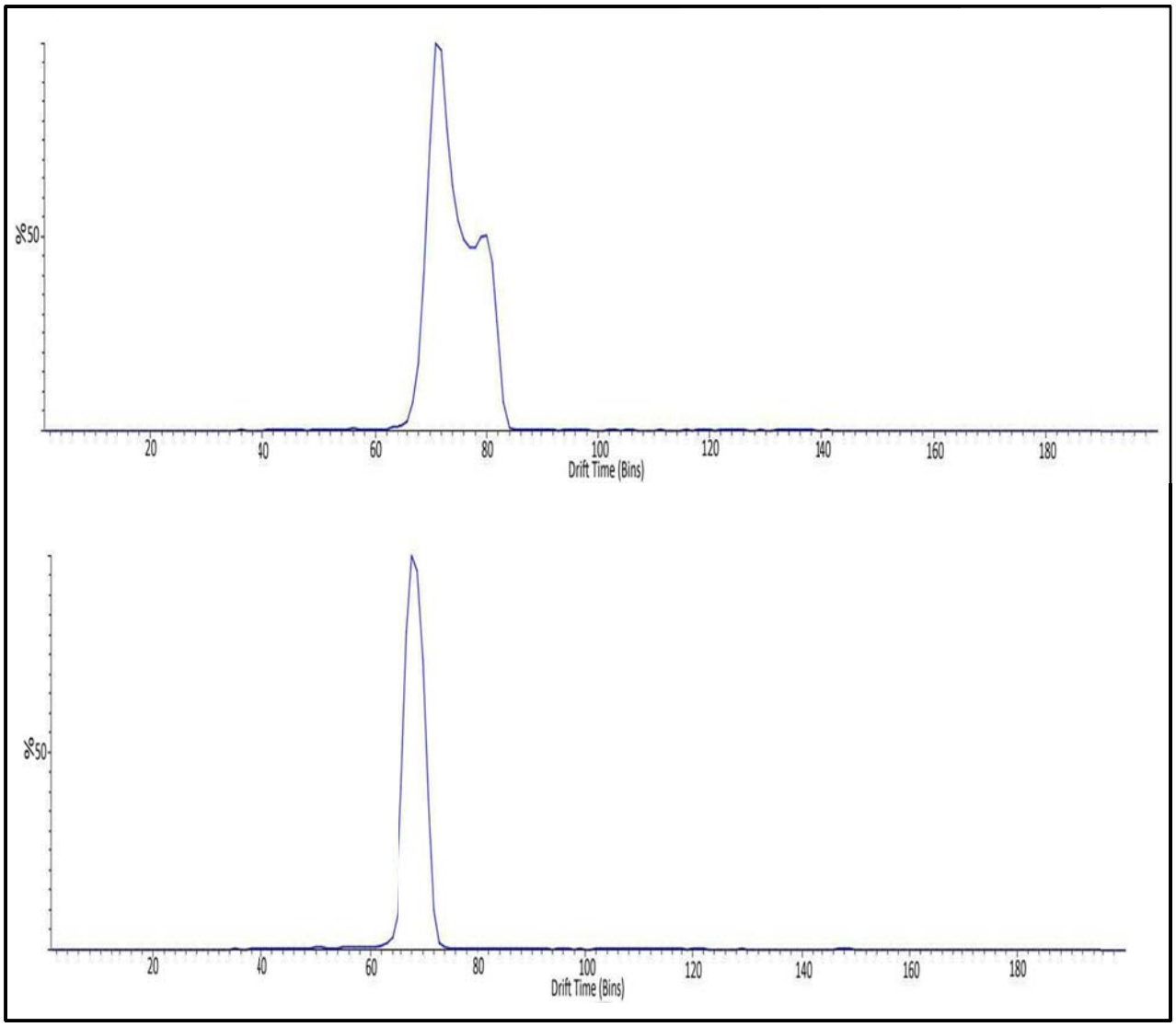


Figure 5-3. (A) 2D Plot of *m/z* versus ion arrival time for active peptides from leukocyte extract. (B) Ion mobility drift times.

Figure 5-3. cont'd.

(B)



The two peptides, from the RP fraction exhibiting antimicrobial activity, with similar retention times between 17.5 and 18 min (seen in Figure 5-4B and C) were separated in the IM cell based on their ion mobility drift times as seen in Figure 5-4A. The peptide ions were measured at *m/z* = 827.8 with a charge state of +6 and *m/z* = 949.9 with a charge state of +5. These results indicate that these two peptides can be separated on the basis of their net charge.

(A)

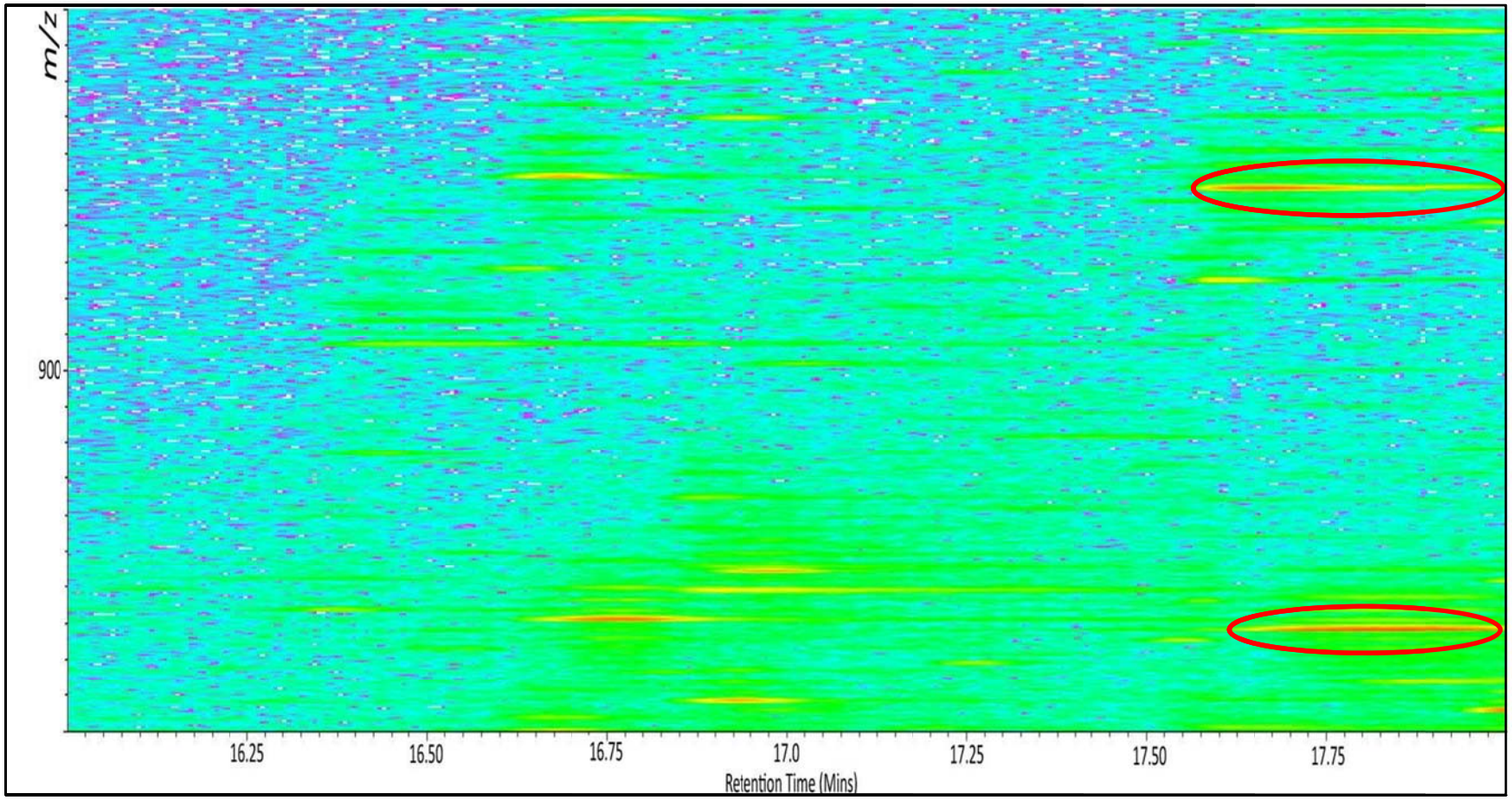
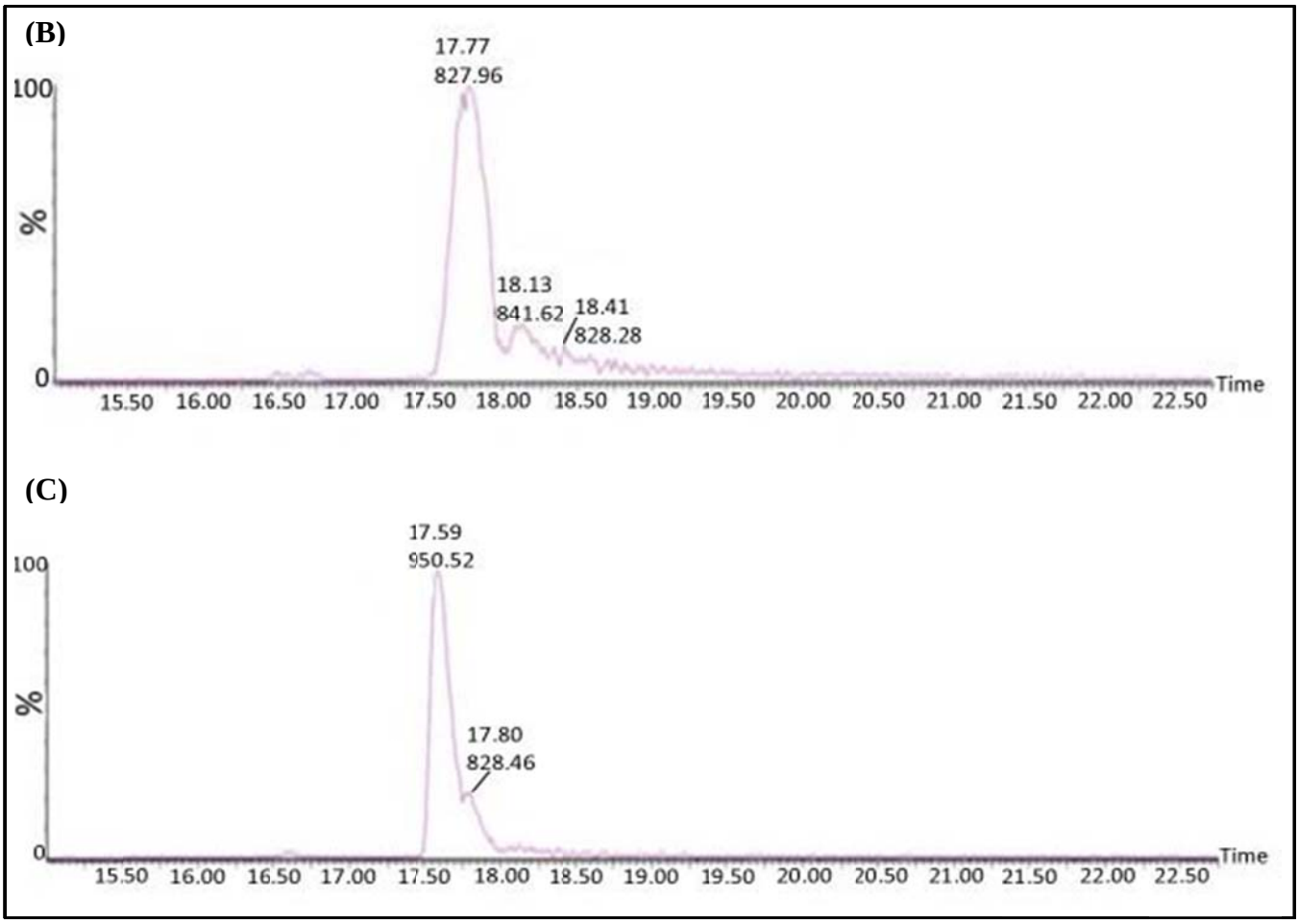


Figure 5-4. (A) 2D plot of *m/z* versus retention time representing the coupling of nanoLC to ion mobility separation. Circled in red is the separation of the 4.7 (top) and 4.9 kDa (bottom) peptide. (B) Extracted ion chromatograms recorded for the 4.7 kDa peptide and (C) 4.9 kDa peptide.

Figure 5-4. cont'd.



The peptides measured at 4745.4126 and 4961.4771 Da that were separated using IMS were subjected to MS fragmentation. However, insufficient fragments from the parent ions were generated; therefore, the MS data could not be used for *de novo* sequencing.

5.3.4 MALDI MS/MS of Peptides

Due to limited fragmentation from the ion-mobility spectrometry-mass spectrometry analysis, the leukocyte fraction exhibiting antimicrobial activity was analyzed using MALDI MS/MS. The two major peaks were observed at 4.7 and 4.9 kDa (Figure 5-5A), which were in agreement with the ESI-MS data (Figure 5-1). The peptides at 4.7 and 4.9 kDa were selected for

MALDI LIFT-MS/MS analysis (Figures 5-5B and 5-6C) and the primary structure was obtained for the 4.7 kDa peptide via *de novo* sequencing. The peptide sequence RDHRTKKLKKFDKSMMPECEKCKESKDFKKNKKTWECGA was determined for the 4.7 kDa peptide using both automated and manual *de novo* sequencing. Fragment ions predicted for the 4.7 kDa peptide sequence were confirmed in the mass spectrum shown in Table 5-1.

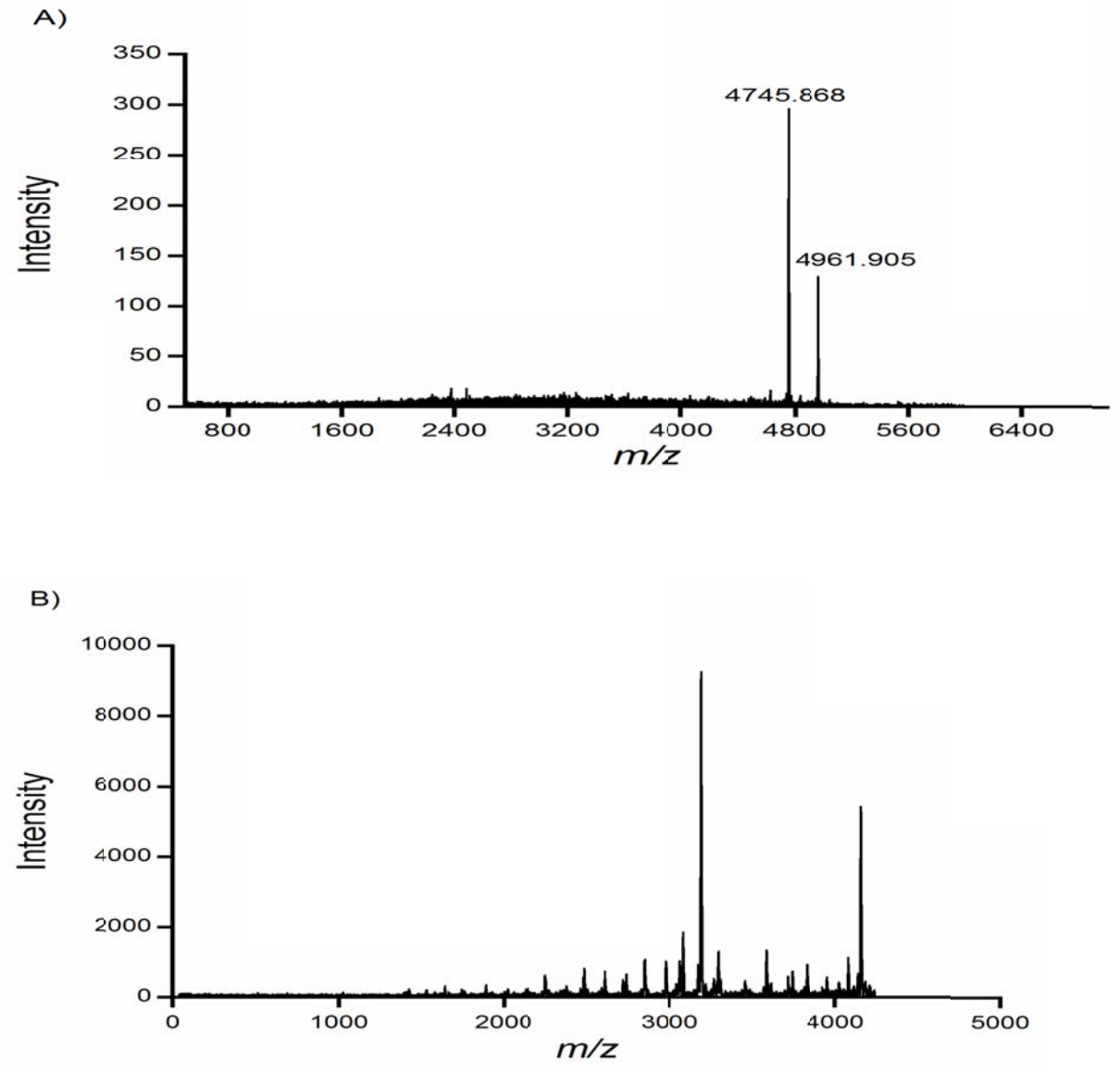


Figure 5-5. (A) MALDI spectrum of leukocyte fraction exhibiting antimicrobial activity. (B) MALDI LIFT-TOF/TOF spectrum of the 4.7 kDa peptide and (C) 4.9 kDa peptide.

Figure 5-5. cont'd.

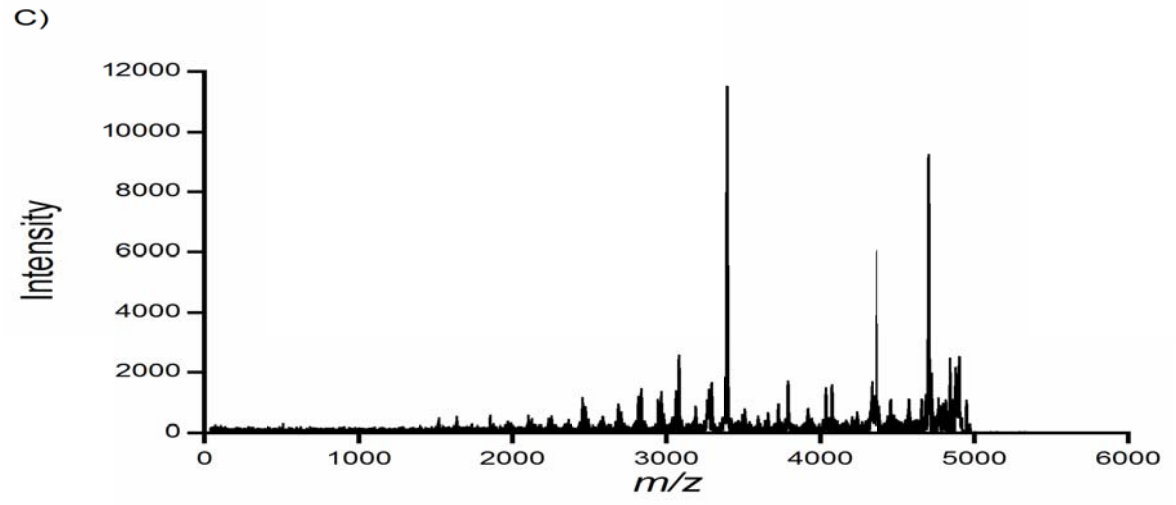


Table 5-1. Fragment ions predicted and confirmed in the mass spectrum for the 4.7 kDa peptide.

N-term	Residue	y-series	b-series	C-term
1	R	**4746.3921**	157.1084	39
2	D	**4590.2910**	272.1353	38
3	H	**4475.2641**	409.1942	37
4	R	**4338.2052**	565.2954	36
5	T	**4182.1041**	666.3430	35
6	K	**4081.0564**	794.4380	34
7	K	**3952.9614**	922.5330	33
8	L	**3824.8665**	1035.6170	32
9	K	**3711.7824**	1163.7120	31
10	K	**3583.6874**	1291.8069	30
11	F	**3455.5925**	1438.8754	29
12	D	**3308.5241**	1553.9023	28
13	K	**3193.4971**	1681.9973	27
14	S	**3065.4022**	1769.0293	26
15	M	**2978.3701**	1900.0658	25
16	M	**2847.3296**	2031.1103	24
17	P	**2716.2892**	**2128.1630**	23
18	E	2619.2364	**2257.2056**	22
19	C	**2490.1938**	2360.2148	21
20	E	2387.1846	**2489.2574**	20
21	K	**2258.1420**	2617.3524	19
22	C	**2130.0471**	**2720.3615**	18
23	K	2027.0379	**2848.4565**	17
24	E	1898.9429	**2977.4991**	16
25	S	1769.9003	**3064.5311**	15
26	K	1682.8683	**3192.6261**	14
27	D	1554.7733	**3307.6530**	13
28	F	1439.7464	**3454.7215**	12
29	K	1292.6780	**3582.8164**	11
30	K	1164.5830	**3710.9114**	10
31	N	1036.4880	**3824.9543**	9
32	K	922.4451	**3953.0493**	8
33	K	794.3502	**4081.1442**	7
34	T	666.2552	**4182.1919**	6
35	W	565.2075	**4368.2712**	5

Table 5-1. cont'd.

36	E	379.1282	**4497.3138**	4
37	C	250.0856	**4600.3230**	3
38	G	147.0764	**4657.3445**	2
39	A	90.0550	**4728.3816**	1

Mass tolerance: 0.5 Da
MS/MS tolerance: 0.5 Da
Bold represents ions found in the mass spectrum.

Peptides in the low mass range of the spectrum were absent due to inefficient fragmentation of the 4.7 kDa peptide. However, the fragments in the higher *m/z* range correlated with the "b" and "y" ions in the mass spectrum, and allowing the determination of the peptide sequence from the C- to N-terminus.

The 4.9 kDa peptide was sequenced in a similar manner to the 4.7 kDa peptide and the sequence obtained for this peptide is

MGAKTKWKRFRGLDTCVFFLLSCKKHCCLLHHTKGRNKTKAAV.

The peptide primary structure was determined based on the fragment ions in Table 5-2. Fragments in the mass spectrum that were not present in either the b- or y-ion series for the 4.9 kDa peptide were predicted by the *de novo* software.

Table 5-2. Fragment ions predicted and confirmed in the mass spectrum for the 4.9 kDa peptide.

N-term	Residue	y-series	b-series	C-term
1	M	4961.6449	132.0478	43
2	G	4830.6044	189.0692	42
3	A	**4773.5829**	260.1063	41
4	K	**4702.5458**	388.2013	40
5	T	**4574.4509**	489.2490	39
6	K	4473.4032	617.3439	38
7	W	**4345.3082**	803.4233	37
8	K	4159.2289	931.5182	36
9	R	**4031.1339**	1087.6193	35
10	F	3875.0328	1234.6877	34
11	R	**3727.9644**	1390.7889	33
12	G	3571.8633	1447.8103	32
13	L	**3514.8418**	1560.8944	31
14	D	**3401.7578**	1675.9213	30
15	T	**3286.7308**	1776.9690	29
16	C	**3185.6832**	1879.9782	28
17	V	**3082.6740**	1979.0466	27
18	F	**2983.6056**	2126.1150	26
19	F	**2836.5371**	2273.1834	25
20	L	**2689.4387**	2386.2675	24
21	L	2576.3847	2499.3516	23
22	S	**2463.3006**	**2586.3836**	22
23	C	2376.2686	**2689.3928**	21
24	K	2273.2594	**2817.4877**	20
25	K	2145.1644	**2945.5827**	19

Table 5-2. cont'd.

26	H	2017.0695	**3082.6416**	18
27	C	1880.0106	**3185.6508**	17
28	C	1777.0014	**3288.6600**	16
29	L	1673.9922	**3401.7440**	15
30	L	1560.9081	**3514.8281**	14
31	H	1447.8241	3651.8870	13
32	H	1310.7651	**3788.9459**	12
33	T	1173.7062	3889.9936	11
34	K	1072.6586	4018.0886	10
35	G	944.5636	**4075.1100**	9
36	R	887.5421	4231.2111	8
37	N	731.4410	**4345.2541**	7
38	K	617.3981	4473.3490	6
39	T	489.3031	**4574.3967**	5
40	K	388.2554	**4702.4917**	4
41	A	260.1605	**4773.5288**	3
42	A	189.1234	**4844.5659**	2
43	V	118.0863	4943.6343	1

Mass tolerance: 0.5 Da
MS/MS tolerance: 0.5 Da
Bold represents ions found in the mass spectrum.

5.3.5 ESI MS/MS of Peptides

ESI MS/MS experiments were performed to collect corresponding sequence information on the peptides from the fraction exhibiting antimicrobial activity, to overlap with the peptide sequences obtained from the MALDI TOF/TOF experiments. The ion measured at *m/z* 940.5 (5+ charge), was subjected to MS/MS fragmentation and data was acquired using different collision energies. The spectrum generated using 45 eV collision energy shows the following fragmentation pattern in Figure 5-6.

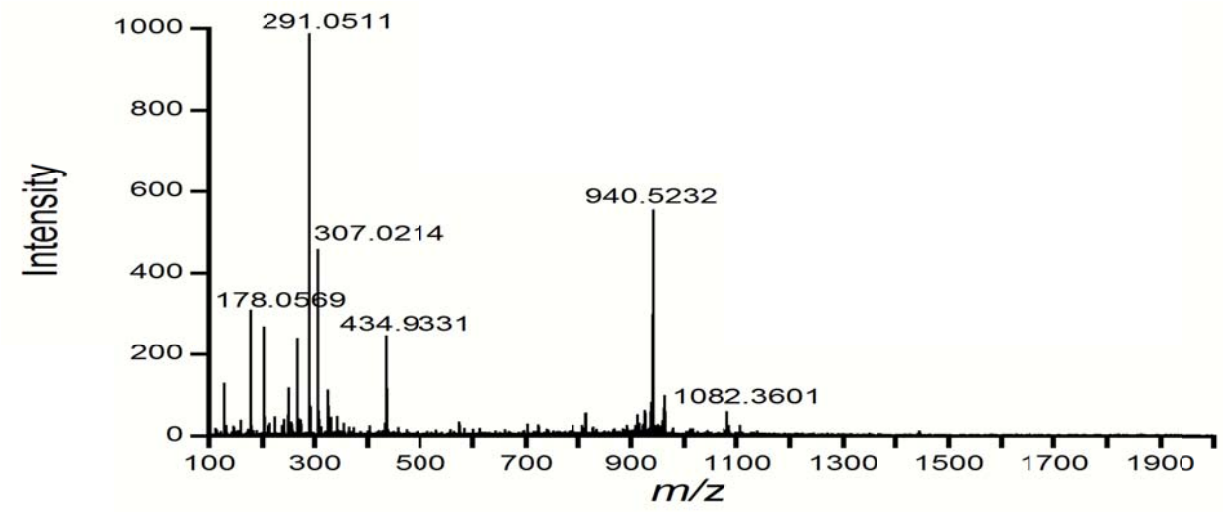


Figure 5-6. MS/MS of on the ion measured at *m/z* 940 with a 5+ charge.

Incomplete sequence information was determined from the spectrum due to the lack of fragmentation of the parent ion. Fragmentation with lower collision energies (e.g. 25 eV) and higher collision energy (e.g. 60 eV) were the same. This data suggests that 1) the size of the intact peptide is too large to fragment using low energy CID and 2) disulfide bond(s) may be present which will stabilize the peptide structure and make it difficult to fragment.

5.3.6 Alligator Peptide Sequence Analysis

The alligator peptide sequence generated from the 4.7 kDa peptide was searched against the antimicrobial peptide database for sequence similarity and it showed 36% sequence similarity to Pilosulin 4, an antimicrobial peptide from an Australian ant and 33% sequence homology to NA-CATH, a snake cathelicidins. The similarities between the antimicrobial peptides are seen in Figure 5-7.

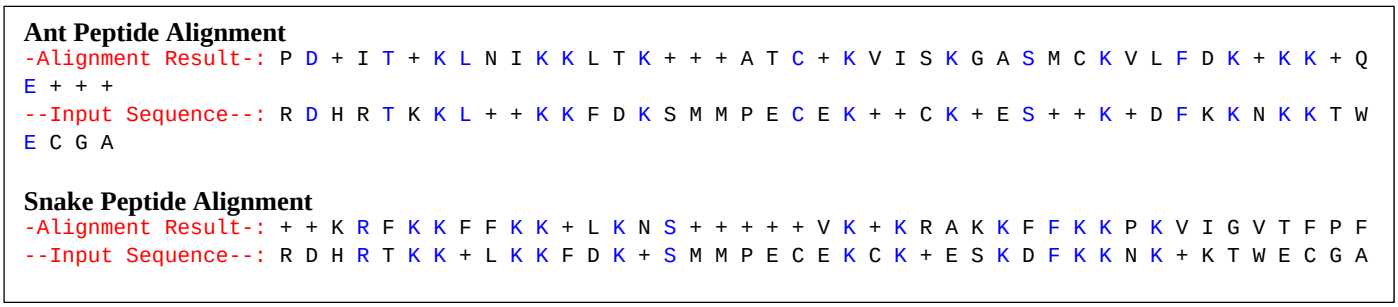


Figure 5-7. Sequence comparison among antimicrobial peptides with alligator 4.7 kDa peptide. Residues highlighted in blue represent identity in the sequence.

The alligator peptide sequence has low sequence homology when compared to antimicrobial peptides from other species, including snakes, from the antimicrobial peptide database. This suggests that the alligator sequence may function as an antimicrobial peptide based on its antimicrobial characteristics but its sequence does not share strong homology to other reptilian antimicrobial peptides.

The 4.7 kDa peptide has a theoretical isoelectric point (pI) of 9.63 which is close to the average pI of antimicrobial peptides which is 9.26.[293] However, a high positive net charge is also required for antimicrobial peptides to carry out their function. The 4.7 kDa peptide sequence has a net charge of +8 at pH 5.5, indicating that it will also be cationic under *in vivo* conditions near physiological pH. The cationic nature of antimicrobial peptides is conserved and it may be essential for the initial electrostatic attraction to the phospholipid membrane of bacteria and fungi.[24]

There are seven hydrophobic residues at positions 8, 11, 15, 16, 28, 35, and 39 and the peptide may form an alpha helix. The total net charge, hydrophobic ratio, and Boman index for the peptides were calculated using the antimicrobial peptide prediction server. The antimicrobial peptide prediction server predicts whether a new peptide sequence has the potential to be antimicrobial based on principles such as size, net charge, hydrophobic percentage and residue composition. The system can also perform simple structural predictions, for example if hydrophobic residues occur at every two to three residues in the peptide sequence an amphipathic helix will be predicted.[292]

The Boman index is an estimate of the antimicrobial peptide's potential to bind to other proteins.[294] A lower index value (≤1) indicates that the peptide will most likely have more antibacterial activity with little side effects *in vivo* whereas a higher index value indicates that the peptide is multifunctional.[294] The sequence determined for the peptide measured at 4.7 kDa has a Boman index of 3.8 kcal/mol. It also has an even number of cysteines which suggests that it may form a disulfide-bond linked beta structure,[292] similar to defensins. The percent hydrophobicity of this peptide is 25%, which is similar to that for defensins;[295] suggesting that the peptide may have antimicrobial activity.[296]

The sequence similarity search for the 4.9 kDa peptide showed 33% sequence identity to SpStrongylocin 2 from sea urchin and Brevinin-2PTa from frog. The sequences of the peptides can be seen in Figure 5-8.

Sea urchin Peptide Alignment

```
-Alignment Result-: W N P F R K L Y R K E C N D V T S C D T V S G V K T C T K K N C C + + + H R
K F F G K T I L K A P E C T V I S
--Input Sequence--: M G A K T K W K R F R G L D + T + C + + V F F L L S C + K K H C C L L H H T
K G R N K T + + K A + + + A V + +
```

Frog Peptide Alignment

```
-Alignment Result-: + G A I K D A L K + + + G A A K T + V A V E L L + + K K A Q C + + + + + K +
L E K T + + + C
--Input Sequence--: M G A + K T K W K R F R G + L D T C V + F F L L S C K K H C C L L H H T K G
R N K T K A A V
```

Figure 5-8. Sequence similarity between antimicrobial peptides with alligator 4.9 kDa peptide. Highlighted in blue represent identical peptides.

Like the 4.7 kDa peptide, the sequence similarity with the alligator 4.9 kDa peptide to other peptides in the database is low.

The 4.9 kDa peptide sequence has a net charge of +13 at pH 5.5 and a theortical pI of 10.26. The peptide has 19 hydrophobic residues, 2 Val, 5 Leu, 3 Phe, 4 Cys, 1 Met, 3 Ala, and 1 Trp and the percent hydrophobicity of the peptide is 44%. The peptide sequence of the 4.9 kDa peptide contains an even number of cysteines; therefore it may form a disulfide-bond linked beta structure similar to β-defensins. The Boman index for this peptide is 1.68 kcal/mol, which is similar to defensins,[295] suggesting that the peptide may have antimicrobial activity.[294]

5.4 Summary

Peptides and proteins were extracted from alligator leukocyte and the mixture was tested for antimicrobial activity. The extract was separated, each fraction tested for antimicrobial activity, and the peptides from the active fraction were sequenced using mass spectrometry and *de novo* sequencing. Based on ESI and MALDI MS measurements, the masses of the predominant peptides in the active fraction were found to be 4.7 and 4.9 kDa. The peptides were

not separated by RP-HPLC; hence ion-mobility mass spectrometry was used to separate the peptides by their size, mass, and charge for MS analysis and sequencing. Due to limited fragmentation from IM-MS, MALDI MS/MS was used with *de novo* sequencing to determine the sequence of the 4.7 kDa peptide. The putative sequence has 39 residues and a net charge of +8 at pH 5.5 and the 4.9 kDa peptide has a putative sequence containing 43 residues and a net charge of +13 at pH 5.5. Based on the peptide's primary structures, size, cationicity, and its predicted antimicrobial activity; the 4.7 and 4.9 kDa peptide properties are consistent with the characteristics of antimicrobial peptides.

CHAPTER 6. ISOLATION AND DETERMINATION OF THE PRIMARY STRUCTURE OF A LECTIN PROTEIN FROM THE SERUM OF THE AMERICAN ALLIGATOR (*ALLIGATOR MISSISSIPPIENSIS*)*

In this chapter mass spectrometry in conjunction with *de novo* sequencing for determining the amino acid sequence of a 35 kDa lectin protein isolated from the serum of the American alligator (*Alligator mississippiensis*) that exhibits calcium-dependent binding to mannose is described. The protein was isolated from alligator serum using mannose affinity chromatography and the N-terminal sequence was determined using Edman degradation. Enzymatic digestion with trypsin, α-chymotrypsin, Lys-C, Glu-C, and Asp-N proteases was used to generate peptide fragments for analysis by liquid chromatography tandem mass spectrometry (LC-MS/MS). Separate analysis of the protein digests with multiple enzymes extended the protein sequence coverage. *De novo* sequencing was accomplished using Mascot Distiller and PEAKS software and the sequences were searched against the NCBI database using Mascot and BLAST to identify homologous peptides. A 35 kDa protein was identified as an alligator lectin.

6.1 Introduction

Animal lectins comprise three major families: S-type lectins, P-type lectins, and C-type lectins.[297] S-type lectins (also known as galectins) are found in the cytoplasm and nucleus of cells and sometimes on cell surfaces.[298] This family of lectins is involved in cell adhesion and plays an important role in normal development of multicellular animals.[298] P-type lectins target lysosomal enzymes in their subcellular compartment,[299] and they differ from other lectins in that they have the ability to recognize phosphorylated mannose residues. The C-type lectins bind to carbohydrates in a calcium-dependent manner and are grouped into three classes: endocytic lectins, collectins, and selectins.[298]

*The work reported in the chapter has been submitted to *Comparative Biochemistry and Physiology – Part B: Biochemistry & Molecular Biology.* The protein sequence data reported in this paper will appear in the UniProt Knowledgebase under the accession number P86928.

Endocytic lectins are found in cell membranes and are specific for different sugars, for example mannose, which is found on the cell surface of microorganisms.[298] Endocytic lectins act as opsonins, binding to the cell surface carbohydrates on microorganisms, stimulating phagocytisis by leukocytes, and eventually death. Collectins function in a similar way to endocytic lectins; they bind to microorganisms that have mannose glycans on their surfaces, leading to lysis of the pathogen and activation of the innate immune system.[297] Selectins direct the movement of leukocytes to sites of inflammation.[300] The importance of animal lectins is increasingly recognized and studied because of their critical roles in host defense and potential for novel medical applications.[301]

Among the family of animal lectins, the C-type lectin known as mannose-binding lectin (MBL) has been well characterized in humans and some other animals.[302, 303] Mannose-binding lectin is a calcium-dependent protein found in the serum of mammals.[304] This protein, which is generated in the liver, tags the surface of bacteria making them susceptible to destruction.[304] The C-type lectin consists of a collagen-like domain which interacts with cells that have been targeted by the innate immune system.[305] Binding of MBL to target cells activates the serum complement component of the immune system, which is an important element of innate immunity[305] and consists of a set of plasma proteins that attack extracellular pathogens.[306] Complement proteins play an important role in innate immunity. The innate immune system can identify foreign material by recognizing pattern receptors, such as mannose, that are found on the surface of pathogens.[306, 307]

Another family of animal lectins that bind to carbohydrates in a calcium-dependent manner is intelectins.[307] Intelectins exhibit high sequence homology and have been identified in many organisms including human,[308] mice,[309] frogs,[310] trout,[311] and ascidians.[312] The principal function of intelectins is to participate in cell differentiation,[313] apoptosis,[314] and recognize tumor

antigens.[315] However, they also recognize pathogens and bacterial components and have roles in innate immunity.[307] Hence, they behave similarly to C-type lectins.

Reptilian lectins such as those from lizards[316] and snakes[317] have been isolated and characterized. However, limited information is available for alligator lectins. Alligators have a potent immune system and their serum produces a broader antibacterial activity compared to other vertebrates such as humans.[318] Hence, characterization of lectins from alligator serum may provide a better understanding of the structure-function relationship of their immune system.

Lectins from different species such as *Turbo (Lunella) coreensis,*[319] *Bubalus bubalis,*[320] *Halocynthia roretzi,*[312] and *Salmo salar*[321] have been isolated and identified using Edman degradation and cDNA cloning.[322-324] However, mass spectrometry is a more commonly used tool for protein sequencing and identification.[325] Tandem mass spectrometry (MS/MS) in conjunction with separation techniques is a commonly used approach to sequence and characterize proteins derived from complex mixtures.[325] A recent study showed the use of mass spectrometry to characterize the primary sequence of a galactose-binding lectin from the seeds of a legume plant.[326] In addition to this study, several other studies have been conducted that demonstrate the utility of mass spectrometry for identifying lectin sequences.[327-329]

One of the major challenges in the study of alligator proteins is the limited information available on the reptilian genome and proteome. For proteomic analysis of species with limited proteomic data, *de novo* sequencing can be used to determine protein sequences.[330] In typical *de novo* sequencing methods, proteins are isolated and enzymatically digested, and the resulting peptides are analyzed by tandem mass spectrometry. The sequence is determined by observing the mass differences between fragment peaks in the tandem mass spectra that are generated by fragmentation of the peptide of interest.[331] *De novo* sequencing was previously used to sequence peptides from proteins of the American alligator leukocytes.[332]

In the present study, mannose affinity separation was used to isolate a 35 kDa lectin protein from the serum of the American alligator (*Alligator mississippiensis*). Liquid chromatography ESI-MS and MALDI-MS was used to obtain the mass of the intact protein. The protein was digested using several different proteases. Peptide sequences were determined using *de novo* sequencing. Database searching and BLAST search was also used to match peptides with more than 50% homology. Using mass spectrometry and different enzymes independently, peptides of different lengths were generated and aligned to determine the sequence of the American alligator lectin.

6.2 Experimental

Proteins were isolated from alligator serum with affinity chromatography. Mannan agarose (3 mL) was equilibrated with 3 mL of mixing buffer (10 mM Tris-HCl, pH 7.8, 1.25 M NaCl) and allowed to settle into a column prepared from a Pasteur pipette plugged with cotton. Fresh alligator serum (5 mL) was mixed with 5 mL of loading buffer (20 mM Tris-HCl, pH 7.8, 2.5 M NaCl). The diluted serum was allowed to filter through the column and the column was washed with 10 volumes of loading buffer. The proteins were eluted with 5 mL of elution buffer (10 mM Tris-HCl, pH 7.8, 1.25 M NaCl, 2 mM EDTA). The isolated lectin protein was transferred to 10 kDa centrifugal concentrator tubes and centrifuged at 7500 g to concentrate and desalt the sample. A portion of the concentrate was used for mass analysis before enzymatic digestion.

The collected lectin samples were analyzed by reverse phase liquid chromatography (RP-LC) ESI time-of-flight mass spectrometry (Agilent 6210) and matrix-assisted laser desorption ionization time-of-flight/time-of-flight tandem mass spectrometer to obtain the mass of the intact protein. For LC/MS, 10 µg of sample obtained from affinity separation was dissolved in 10 µl of water containing 0.1% formic acid (FA) and loaded onto a C_{18} reverse phase column 5 µm 250 ×

mm (Vydac) which had been equilibrated with 0.1% FA in water. A linear gradient from 0% to 60% of B over 60 minutes was used for protein separation. Solvent A was water containing 0.1% FA and solvent B was acetonitrile containing 0.1% FA. The eluent from the LC column was directed towards the mass spectrometer and mass spectra were acquired from 600 to 3000 *m/z* using ESI at 4200 V. The spectra were extracted and deconvoluted using Analyst software.

The protein samples for MALDI analysis were deposited on a ground-steel target using a 1:1 ratio (v/v) of sample and saturated matrix solution. The matrices used were a 9:1 mixture of 2,5-dihydroxybenzoic acid and 2-hydroxy-5-methoxybenzoic acid (sDHB) and 2,4-dimethoxy-3-hydroxycinnamic acid (sinapinic acid, SA). The samples were analyzed in positive ion mode and 500 shots were collected. The mass spectra were collected using a mass range from 20 to 90 kDa. Exact masses from the spectra were determined using Bruker FlexAnalysis software.

One-dimensional SDS-PAGE was performed using pre-cast 4–20% polyacrylamide gradient tris-HCl gels on the small format gel electrophoresis system. The samples were mixed 1:1 v/v with sample buffer containing 62.5 mM tris-HCl at pH 6.8, 2% SDS, 25% glycerol (w/v), 0.01% w/v bromophenol blue, and, in both the presence and absence of 10% 2-mercaproethanol (v/v), heated at 95ºC in a water bath for 5 minutes before loading on the gel. Samples were run at 100 V for 1 h using tris-glycine (pH 8.3) as the gel running buffer. Gels were stained with Coomassie blue for 1 h and rinsed twice with distilled water for 10 min (each time) to remove the stain. The gel images were captured using the Gel Doc XR System and with BioRad Quantity One 1-D analysis software. Samples were excised from the SDS polyacrylamide gel and submitted for N-terminal sequencing by Edman degradation using a protein sequencer (Model 494, Perkin Elmer Applied Biosystems).

A purified alligator lectin (40 μg) was incubated in the presence of 40 μL of mannan-, mannose-, D-lactose-, N-acetylgalactosmine-, N-acetylglucosamine-, and β-D-glucose-agarose.

After 30-minute incubation at ambient temperature, the samples were centrifuged and a 20 µL aliquot of the supernatant (containing the unbound protein) was subjected to SDS-PAGE. One-dimensional SDS-PAGE was performed and the gels were stained with Coomassie blue. The experiment was repeated three times. The density of the bands, in pixels, was evaluated and compared using ImageJ software.[333]

To obtain maximum sequence coverage, separate aliquots of lectin protein obtained from affinity separation were subjected to several different enzymatic digestions. The protein fractions were dissolved in a 100 mM ammonium bicarbonate buffer adjusted to the pH range in which each endoprotease was most active. The protein was reduced using 50 mM DTT in 25 mM ammonium bicarbonate buffer and incubated at 37ºC for one hour. Following reduction, the protein was alkylated with 50 mM iodacetamide in 25 mM ammonium bicarbonate buffer for 45 minutes at room temperature in the dark. The in-solution digestions were performed separately using trypsin, α-chymotrypsin, Lys-C, Glu-C, or Asp-N according to the manufacturer's protocol. Briefly, digestion with trypsin was performed using a 1:20 ratio of enzyme to substrate in 100 mM ammonium bicarbonate, pH 8.5; with α-chymotrypsin using a 1:20 ratio at pH 8.5; with Lys-C using a 1:20 ratio at pH 8.5; with Glu-C using a 1:20 ratio at pH 7.8; and with Asp-N using a 1:50 ratio at pH 8.5. All digestions were run at 37ºC overnight. The enzymatic digests were dried in a vacuum centrifuge and resuspended in 0.1% FA for LC separation.

Bands from the one-dimensional gels were excised and cut into ~1 mm cubes, and the stain was removed with 200 mM NH_4HCO_3 and 40% acetonitrile (2×30 min). After destaining, the gel pieces were dried in a vacuum centrifuge, rehydrated with 20 µL of 0.4 µg trypsin in 40 mM ammonium bicarbonate containing 9% acetonitrile to completely cover the gel pieces. The tubes were then incubated overnight at 37°C.[334] A similar approach was used for digesting the protein gel pieces using α-chymotrypsin, Lys-C, Glu-C, and Asp-N proteases.[335, 336] Peptides

were extracted from gel pieces with 60% acetonitrile and 40% water containing 0.1% trifluoroacetic acid (TFA). CNBr digestions were performed overnight with 1 M CNBr at room temperature in the dark. The following day the CNBr digestion was dried completely using a speed-vac and resuspended in water containing 0.1% FA for ESI analysis. FA digestions were performed overnight at 37ºC.

Liquid chromatography electrospray tandem mass spectrometry (LC MS/MS) was used to analyze the peptides derived from the in-gel and in-solution digests. The peptides were dissolved in 20 µL of 0.1% FA and 10 µL of the digest was injected onto a 0.3 × 1 mm trapping column (PepMap) on a nano-LC system (Ultimate,) at a flow rate of 10 µL/min. The peptides were eluted onto a 75 µm × 15 cm C_{18} column (Biobasic) and separated using a gradient of 5 - 40% B over 90 minutes with a flow rate of 200 nL/min. The peptides were analyzed via a quadrupole time-of-flight mass spectrometer (QSTAR) and ionized using a nano-electrospray source at a voltage of 2.5 kV. The mass spectrometer was operated in information-dependent acquisition (IDA) mode. Three collision energies (25, 38, and 50 eV) were selected to fragment the peptides.

The peptide sequences were generated from MS/MS data using Mascot Distiller and PEAKS software. The Mascot Distiller software determined amino acid sequences from MS/MS data using-peak fitting and isotope distribution[241] and PEAKS software computes peptides for whose ions matches as many of the most abundant peaks in the spectrum.[167] The *de novo* sequencing results were also verified manually. *De novo* peptide sequences were searched using Mascot and BLAST for protein identification based on homology searches. For Mascot and BLAST search NCBI database were used to compare to peptide sequences derived from all taxonomies.

Multiple sequence alignments were performed using DIALIGN, CLUSTALW, and BLAST described in Section 2.7.5. The 35 kDa alligator lectin sequence was aligned and compared with four other vertebrate lectin proteins.

6.3 Results and Discussion

6.3.1 Isolation of the 35 kDa Alligator Lectin

SDS-PAGE analysis under non-reducing conditions (Figure 6-1a) of the isolated alligator lectin proteins showed five major bands between 35 and 250 kDa, which suggests a mixture of proteins or oligomers. By contrast the SDS-PAGE under reducing conditions showed only two bands that corresponded to 35 and 70 kDa (Figure 6-1b), suggesting that some of the bands were the result of oligomers. The band detected at 70 kDa could be another protein or a dimer of the 35 kDa protein.

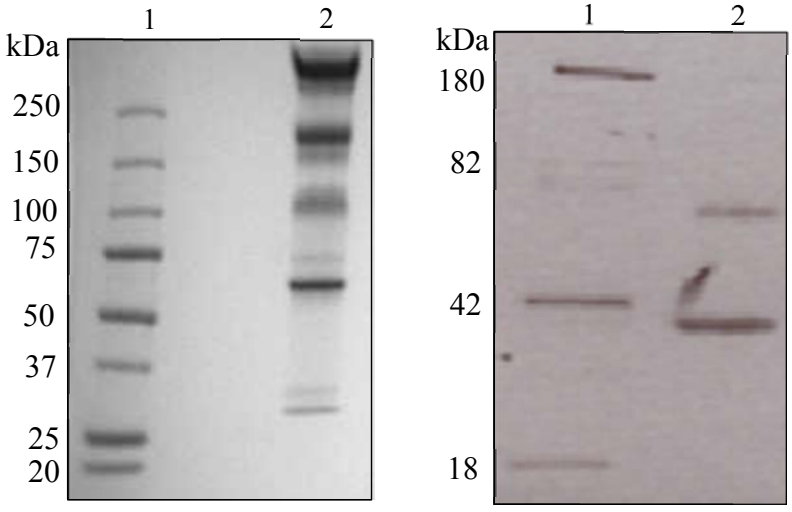


Figure 6-1. SDS-PAGE analysis of non-reduced and reduced alligator lectin protein on a 4–20% gel. (A) Lane 1, molecular weight marker; lane 2, non-reduced alligator lectin. (B) Lane 1, molecular weight marker; lane 2, reduced alligator lectin.

LC ESI-MS and MALDI-MS was performed to obtain the intact mass of the lectin protein and oligomers. Figure 6-2a shows a LC separation of the crude extract of the alligator lectin and Figure 6-2b shows mass spectrum of the major peak that eluted at R_T between 32 and

35 minutes. The deconvoluted mass spectrum of the major peak reveals the average masses of the eluted lectin protein to be 34.6, 69.3, 103.9, and 138.6 kDa, consistent with the monomer, dimer, trimer and tetramer of the 35 kDa protein.

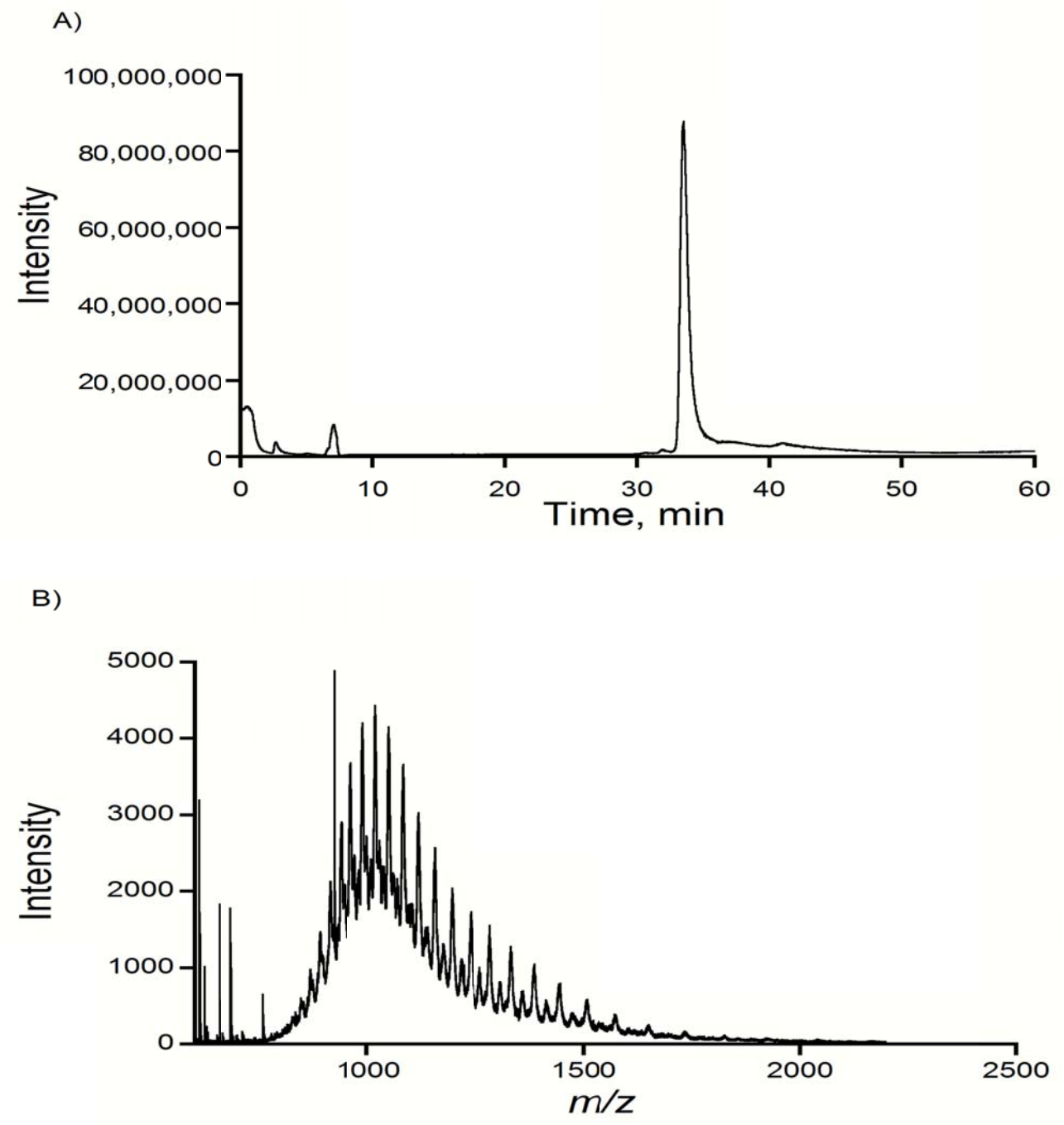


Figure 6-2. (A) The chromatogram of alligator lectin eluted from a C_{18} reversed phase column. The lectin protein was eluted between 32 and 35 minutes. (B) The mass window of the isotopic distribution for the protein masses.

Figure 6-3 shows a MALDI mass spectrum of the crude extract, the observed masses from MALDI-MS analysis were 34.7, 69.3, 104.0, and 138.6 kDa.

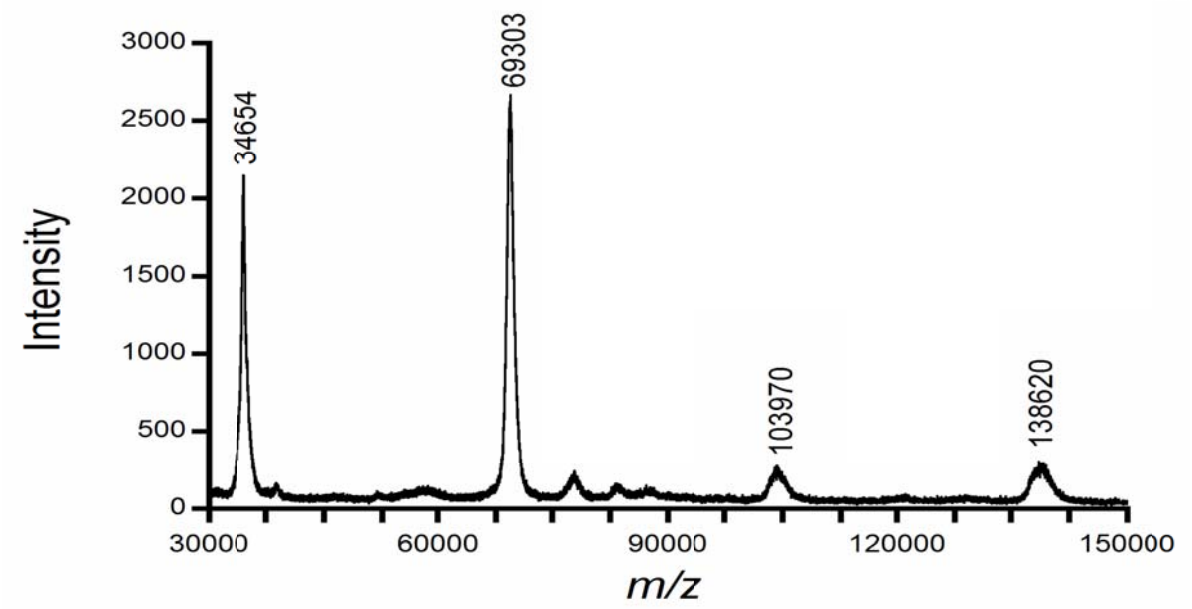


Figure 6-3. Linear mode (+) MALDI mass spectrum showing the average masses of the oligomers from the 35 kDa alligator lectin.

6.3.2 Sequence Determination of the 35 kDa Alligator Lectin.

To sequence the isolated protein, Edman sequencing and mass spectrometry were used. Edman degradation of the 35 kDa molecular weight gel band of the alligator protein produced the N-terminal sequence NNQLALKLAATGGSTNXLPA, where position X was undetermined. The unidentified amino acid at position X is potentially due to the amino acid being glycosylated (O-linked glycosylation) which can not be detected by Edman sequencing.[337]

Peptides were obtained from enzymatic digestion of the 35 kDa gel band and the isolated lectin sample using the different endoproteases as described above. Trypsin produced ten peptides of various lengths from the 35 kDa protein. To collect additional sequence information from overlapping peptides, the endoproteases Lys-C, Asp-N, and Glu-C were used. Endoprotease Lys-C produced five peptides, Glu-C digest produced nine peptides, and seven

peptides were produced from Asp-N digest. A less specific endoprotease, α-chymotrypsin, was also used to obtain additional overlapping peptides. The endoprotease α-chymotrypsin produced seven fragments that overlapped with fragments generated from trypsin, Asp-N, and Lys-C. There was no additional sequence coverage obtained from the CNBr and FA digest.

Peptides obtained from the in-solution and in-gel digests of the 35 kDa band were *de novo* sequenced. Figure 4 shows mass spectra for a Glu-C and tryptic digest of the protein to illustrate peptide sequence determination. Figures 6-4a and 6-4b show examples of mass spectra acquired at 57 and 58 min, respectively, in the separation of the Glu-C and tryptic digests. Tandem MS data were collected and peptide sequences were determined via *de novo* sequencing. Figures 6-4c and 6-4d illustrate the peptide fragment series generated from the MS/MS data acquired from the ion 492.8 in Figure 4a and 743.8 in Figure 6-4 b. The sequences were searched against the NCBI protein database using Mascot and no protein matches were found. The *de novo* sequenced peptides were subjected to a BLAST search to identify proteins with similar sequences. Peptide sequences found with BLAST exhibited strong similarity to intelectin-1 and intelectin-2 proteins from mammals. Some of the lectin peptide sequences showed a high degree of similarity to the intelectin peptides, with few amino acid differences; these regions of similarity appear to be conserved among the lectins. The *de novo* sequenced peptides that did not match peptides in the protein database with a high degree of similarity appear to be regions of the alligator lectin sequence that are unique.

The peptides from the different enzymatic digests produced overlapping data that were aligned to create the primary sequence of the protein (Figure 6-5). Partial protein sequences were generated when peptide sequences from the different digests were overlapped. Multiple alignments of the 35 kDa alligator lectin peptides with peptides of other lectin proteins provided sequence positioning of the alligator lectin peptides.

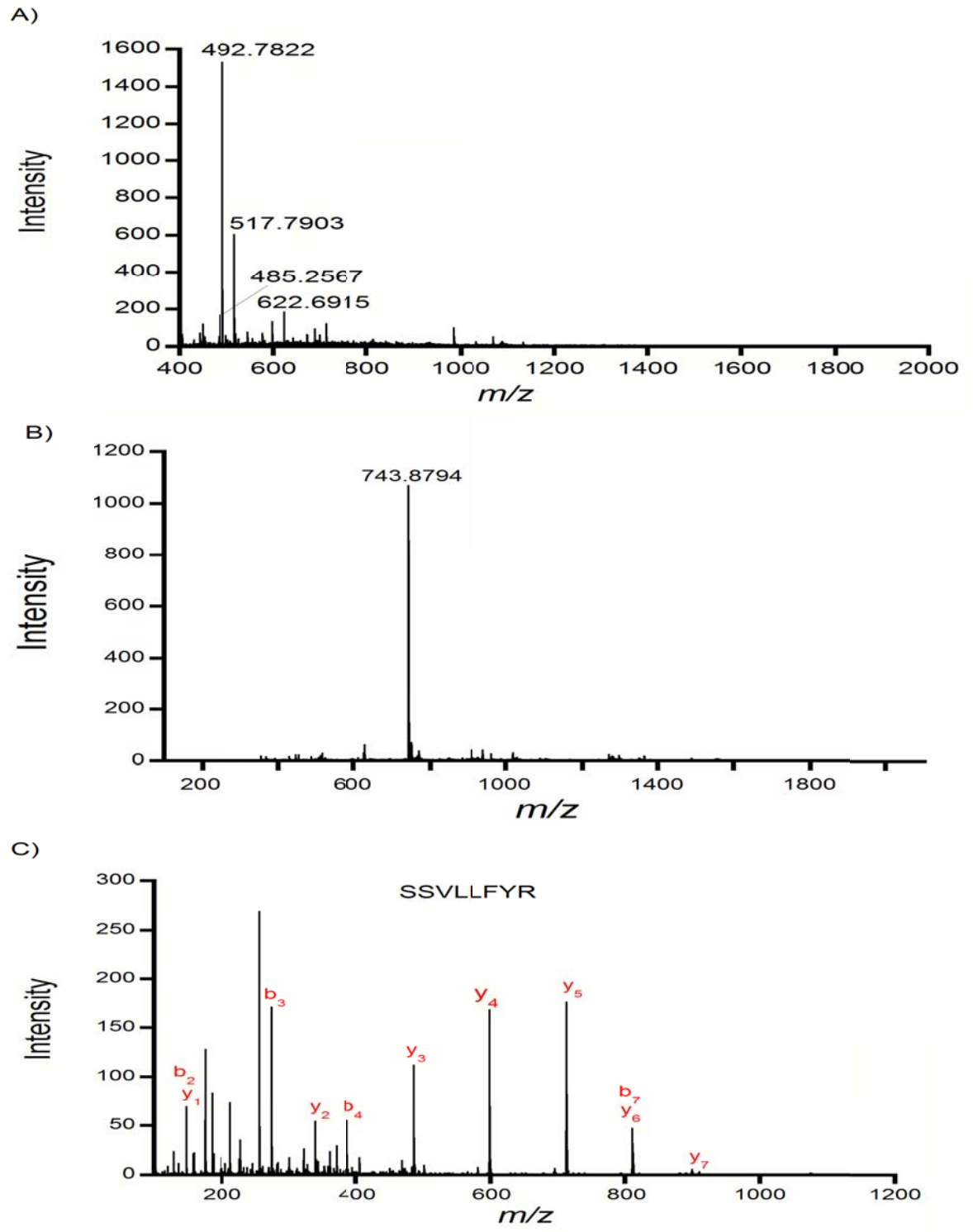
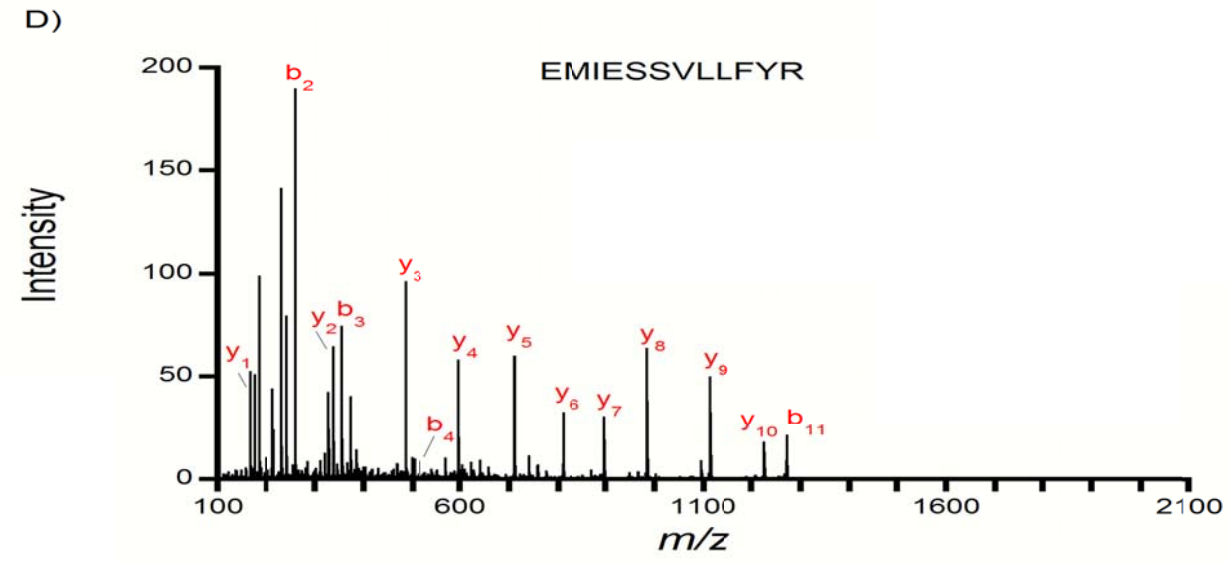


Figure 6-4. Mass spectrometry data for peptide sequencing illustrating (A) an example of peptides eluted from the Glu-C digest at 57 minutes, (B) peptides eluted from trypsin digest at 59 minutes, (C) an MS/MS plot of the ion 492.8 generated from the mass spectrum of the Glu-C digest, and (D) an MS/MS plot of the ion 767.7 generated from the mass spectrum of the tryptic digest.

Figure 6-4. cont'd.



The chymotryptic peptide LAATGGSTNSLPA of the 35 kDa protein determined by mass spectrometry and *de novo* sequencing overlapped with residues 8 to 20 from the Edman sequence and identified the missing residue as serine.

The digests from the 70 kDa gel band generated some peptides that show sequence identity to peptides from the 35 kDa protein. This suggests there is some sequence similarity between the 35 kDa and 70 kDa protein from the SDS-PAGE gel. However, unique peptide sequences were identified, suggesting that this is a different protein rather than the dimer of the 35 kDa protein.

Overall, approximately 98% of the 35 kDa protein was sequenced and found to consist of 313 residues; however, 2% of the protein was not sequenced. The missing sequence may correspond to a short tryptic peptide that was *de novo* sequenced from both the in-gel and in-solution digest. This short tryptic peptide did not overlap with the other enzymatic peptide fragments; hence its position in the sequence could not be determined.

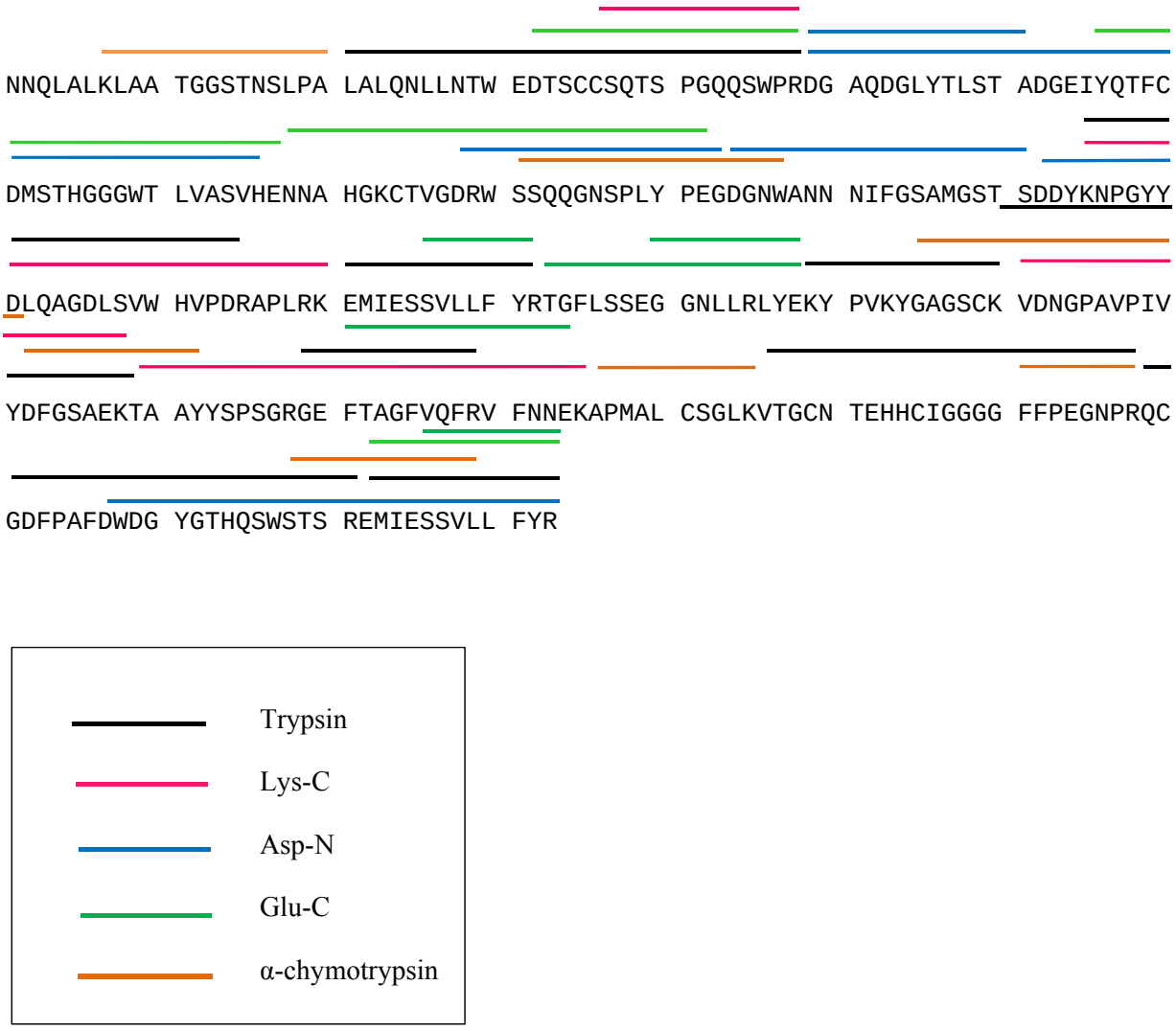
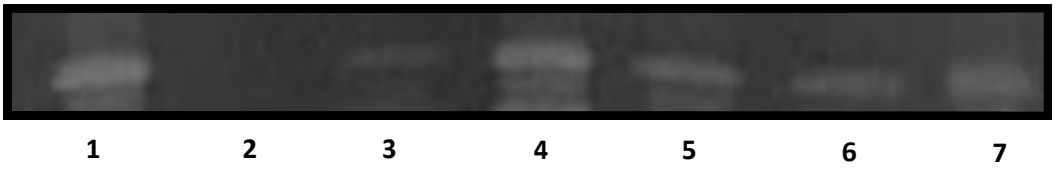


Figure 6-5. Primary structure of the 35 kDa lectin protein isolated from American alligator assembled from different endoprotease digestions. The peptide sequences were generated using ESI-MS/MS. Peptides obtained from the different enzymes was highlighted using different colors.

6.3.3 Binding Activity of Alligator Lectin

The alligator lectin-binding activity was determined for four carbohydrates, and the results are shown in Figures 6-6A and 6-6B.

A)



B)

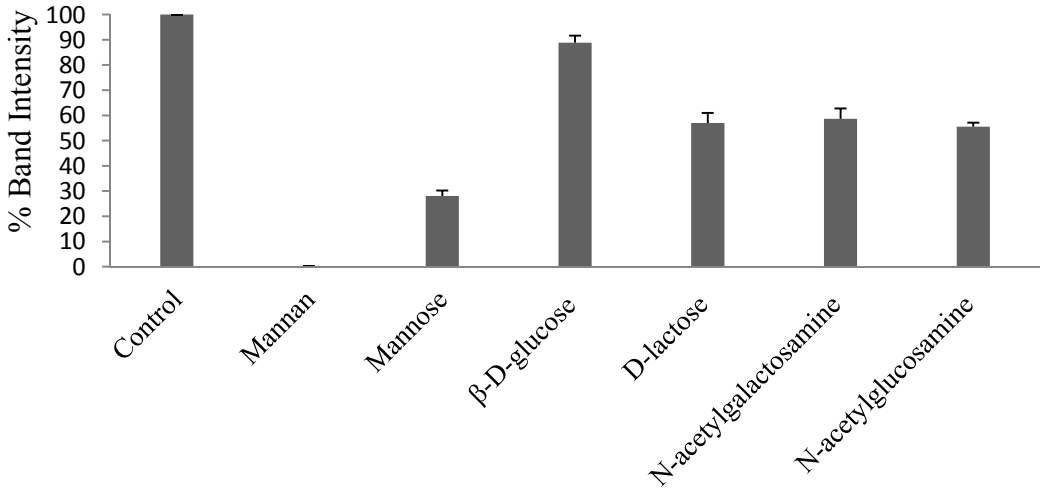


Figure 6-6. (A) The isolated alligator lectin was incubated with agarose-conjugated mannan (Lane 2), mannose (Lane 3), β-D-glucose (Lane 4), D-lactose (Lane 5), N-acetylgalactosamine (Lane 6), and N-acetylglucosamine (Lane 7). The samples were centrifuged and the supernatants were subjected to 1D gel analysis. The control (Lane 1) was alligator lectin with underivatized agarose. (B) The image of the gel bands were quantified in ImageJ, a Java-based image processing program.

The alligator lectin binds strongly to agarose-conjugated mannose and mannan, indicating little or no unbound protein remaining in the supernatant. This produced a relatively low gel band intensity of 28% for mannose and no band was observed in the presence of mannan. Weaker binding activity of the alligator lectin was observed with β-D-glucose, D-lactose, N-acetylglucosamine, and N-acetylgalactosamine; therefore, there was less interaction and a relatively stronger band intensity of 89%, 57%, 56%, and 59% respectively, compared to the control (100%).

6.3.4 Comparison of the 35 kDa Alligator Lectin Sequence with Other Animal Lectins

The alligator lectin sequence was compared with other calcium-dependent animal lectins, specifically mannose-binding lectins and intelectins, from six other species: *Homo sapiens*, *Mus musculus*, *Rattus norvegicus, Pan troglodytes, Gallus gallus, and Xenopus tropicalis* using DIALIGN software. A sequence comparison for the lectin of the above organisms to the alligator lectin sequence resulted in 13%, 3%, 1%, 5%, and 4% sequence similarity, respectively.

Sequence comparison of the alligator lectin sequence to intelectin-1 from *Homo sapiens*, *Mus musculus, Rattus norvegicus,* and *Xenopus tropicalis*, exhibited 57%, 59%, 60%, and 57% similarity respectively (Figure 6-7). A sequence comparison of the 35 kDa alligator lectin protein was also made with intelectin-2 from *Homo sapiens* (58% identity) and *Mus musculus* (59% identity). The degree of similarity of alligator lectin protein sequence to intelectin-2 sequence (Appendix B, Figure 1) is close to that of the intelectin-1. Both mannose-binding lectin and intelectin proteins bind in a calcium-dependent manner.[307] However, sequence comparisons suggest that the lectin protein isolated here is more closely related to intelectins rather than to mannose-binding lectins.

For the human intelectin-1, residue 154 is the start of the consensus sequence (Asn-X-Thr/Ser) indicating a site of N-linked glycosylation. However, this site is not present in the alligator lectin sequence, indicating that it lacks N-glycosylation at this site. There are several possible sites of glycosylation in the alligator lectin sequence, for example at the serine and threonine residues, suggesting that O-linkage is possible. BLAST search of the alligator lectin sequence showed a conserved fibrinogen-related domain near the N-terminus region of the sequence between residues 50 to 80. The fibrinogen related domain has been shown to be similar to the calcium-dependent carbohydrate-recognition domains.[338] Both mouse and human intelectin-1 sequences have a fibrinogen domain between residues 32 to 251 and 32-255

respectively. A comparison of the alligator lectin sequence to both mouse and human intelectin-1 fibrinogen domain sequences showed a high degree of sequence similarity (63% and 61% respectively).

```
Mouse           MTQLGFLLFIMVATRG----------------------CSAAEENLDTNRWGNSFFSSL 37
Rat             MTQLGFLLFLIVATRG----------------------GSAAKEDLETNKGTHSFFDSL 37
Human           MNQLSFLLFLIATTRG----------------------WSTDEANTYFKEWTCSSSPSL 37
Frog            MLVYSLLVFALGFPAGHAGSCEQASISEKKEKILNLLACWTEDKTDNSGSRFSGSPTGDM 60
Alligator       NNQLALKLAATGGSTNX----------------------LPALALQNLLNTWEDTSCCSQ 38
                    .: :     . .                          .    :   .    :   .

Mouse           P---RSCKEIKQEHTKAQDGLYFLRTKNGVIYQTFCDMTTAGGGWTLVASVHENNMRGKC 94
Rat             S---RSCKEIKEENTGAQDGLYFLRTENGVIYQTFCDMTTAGGGWTLVASVHENNMGGKC 94
Human           P---RSCKEIKDECPSAFDGLYFLRTENGVIYQTFCDMTSGGGGWTLVASVHENDMRGKC 94
Frog            SYGYRSCNEIKSSDPSAPDGIYTLATEHGESYQTFCDMTTNGGGWTLVASVHENNMFGKC 120
Alligator       T----SPGQQSWPRDGAQDGLYTLSTADGEIYQTFCDMSTHGGGWTLVASVHENNAHGKC 94
                .    *  : .     * **:* * * .*  *******:: *************:  ***

Mouse           TVGDRWSSQQGNRADYPEGDGNWANYNTFGSAEAATSDDYKNPGYFDIQAENLGIWHVPN 154
Rat             TVGDRWSSQQGNRADYPEGDGNWANYNTFGSAEGATSDDYKNPGYFDIQAENLGIWHVPN 154
Human           TVGDRWSSQQGSKAVYPEGDGNWANYNTFGSAEAATSDDYKNPGYYDIQAKDLGIWHVPN 154
Frog            TVGDRWSTQQGNTLQNPEGDGNWANYATFGLPEGATSDDYKNPGYYDIQAKNLALWHVPN 180
Alligator       TVGDRWSSQQGNSPLYPEGDGNWANNNIFGSAMGSTSDDYKNPGYYDLQAGDLSVWHVPD 154
                *******:***.    *********   ** . .:**********:*:** :*.:****:

Mouse           KSPLHN--WRKSSLLRYRTFTGFLQHLGHNLFGLYKKYPVKYGEGKCWTDNGPALPVVYD 212
Rat             NSPLHS--WRNSSLLRYRTFTGFLQHLGHNLFGLYQKYPVKYGEGKCWTDNGPALPVVYD 212
Human           KSPMQH--WRNSSLLRYRTDTGFLQTLGHNLFGIYQKYPVKYGEGKCWTDNGPVIPVVYD 212
Frog            NTPMFN--WRNSSILRYRTQNSFFTEEGGNLFELYKKYPVKYDIGKCLADNGPAVPVVYD 238
Alligator       RAPLRKEMIESSVLLFYR--TGFLSSEGGNLLRLYEKYPVKYGAGSCKVDNGPAVPIVYD 212
                .:*:     ..* :* **  ..*:   * **: :*:******. *.* .****.:*:***

Mouse           FGDARKTASYYSPSGQREFTAGYVQFRVFNNERAASALCAGVRVTGCNTEHHCIGGGGFF 272
Rat             FGDAQKTASYYSPYGQKEFTAGFVQFRVYNNERAASALCAGMKVTGCNSEAHCIGGGGFF 272
Human           FGDAQKTASYYSPYGQREFTAGFVQFRVFNNERAANALCAGMRVTGCNTEHHCIGGGGYF 272
Frog            LGSAEKTSSLYSPNGRNEFTAGFVQFRVVNTEKAALALCPGVKVKGCNAEHHCIGGGGYF 298
Alligator       FGSAEKTAAYYSPSGRGEFTAGFVQFRVFNNEKAPMALCSGLKVTGCNTEHHCIGGGGFF 272
                :*.*.**:: *** *: *****:***** *.*:*. ***.*::*.***:* *******:*

Mouse           PEGNPVQCGDFASFDWDGYGTHNGYSSSRKITEAAVLLFYR 313
Rat             PEGNPLQCGDFGAFDWNGYGTHIGYSSSREITEAAVLLFYR 313
Human           PEASPQQCGDFSGFDWSGYGTHVGYSSSREITEAAVLLFYR 313
Frog            PEGSPRQCGDFSAFDWDGYGTHAGWSASKEITEAAVLLLYR 339
Alligator       PEGNPRQCGDFPAFDWDGYGTHQSWSTSREMIESSVLLFYR 313
                **..* ***** .***.***** .:*:*::: *::***:**
```

Figure 6-7. The *Alligator mississippiensis* 35 kDa lectin amino acid sequence was compared to intelectin-1 protein of *Homo sapiensis*, *Mus musculus*, *Rattus norvegicus*, and *Xenopus tropicalis*.

Previous reports show that many vertebrate lectins contain homologous amino acid sequences.[303] The 35 kDa alligator lectin sequence shows a high degree of similarity to

intelectins; however, intelectin proteins are animal lectins that typically have an affinity to galactose- and galactofuranose-binding C-type lectins and require calcium for carbohydrate-binding activity.[307] Although intelectins have been previously reported in other organisms, few studies have identified their binding specificity. In this study, the 35 kDa protein identified as alligator lectin had an affinity to mannose with a calcium-binding dependence. A previous report indicated that a rainbow trout intelectin had a binding specificity for both N-acetylglucosamine (GlcNAc) and mannose saccharides.[311] The alligator lectin could be a similar intelectin-like protein that has a binding specificity for mannose residues.

Intelectin proteins are oligomeric; for example, human intelectin-1 is a trimeric protein.[339] ESI and MALDI analysis (Figures 6-2b and 6-3) suggests that the alligator lectin is an oligomeric protein that exists primarily as a 35 kDa monomer *in vitro*.

Mammalian intelectins usually have at least 10 conserved cysteine residues[339] for example as in human intelectin-1 (Figure 6-6); the 35 kDa alligator lectin sequence has nine cysteine residues. It has cysteine residues at positions 35 and 36 at the N-terminus region that are not present in the intelectin sequences of other vertebrates. Also, the alligator lectin sequence lacks a cysteine at position 42 that is conserved in the other organisms. These cysteine residues can influence tertiary protein structure, as well as carbohydrate-recognition specificity and oligomer formation and structure. When comparing the sequence to those derived from other vertebrates, the lack of a cysteine residue at position 41 in the alligator lectin sequence, as well as the different locations of the other cysteine residues may result in lower thermal stability of the protein.

In addition to the presence of the cysteine residues that allow for oligomers to form via disulfide bonding, intelectins also have carbohydrate recognition domains.[340] The calcium-dependent carbohydrate-recognition domains are similar to the fibrinogen-like domain.

Fibrinogen-like proteins typically contain fibrinogen domains at the C-terminal region in both vertebrates and invertebrates.[338, 341] An example of a human fibrinogen-like protein is ficolin. Although ficolin does not show sequence homology to C-type lectins or intelectins, it plays a vital role in innate immunity in a manner similar to the mannose-binding lectin.[342] There are two types of ficolins found in humans, ficolin-α which is a 326 amino acid protein and has a fibrinogen terminal domain at residues 109 to 326, and ficolin-β, a 313 amino acid protein with a fibrinogen terminal domain at residues 96 to 313.[338] The alligator lectin sequence has low sequence similarity to both the fibrinogen domains in human ficolin-α (16% similarity) and ficolin-β (9% similarity) (Appendix B, Figure 2), suggesting that its carbohydrate-recognition domain is different to that found in ficolin proteins. BLAST results for the alligator lectin showed that it has a fibrinogen-related domain between residues 50 and 80, suggesting that its carbohydrate recognition domain is near the N-terminal region of its sequence as previously shown in the *Halocynthia roretzi* lectin from ascidian plasma.[312]

Binding studies were conducted with several carbohydrates leading to the conclusion that the alligator lectin exhibits strong selectivity for both mannan and mannose, which functions as a recognition molecule in the complement system. The alligator lectin showed some binding activity towards D-lactose suggesting that the lectin may also bind to terminal galactose as observed with intelectins from other organisms. The weak binding of the lectin observed in the presence of β-D-glucose, N-acetylglucosamine, and N-acetylgalactosamine, suggests that the lectin has limited binding specificities for these carbohydrates.

6.4 Summary

Sequence data collected from a 35 kDa protein isolated from the serum of *Alligator mississippiensis,* using a mannose affinity column, indicated that it is a lectin protein comprising approximately 313 amino acids. Mass spectrometry analysis of different enzymatic digests, from

both in-gel and in-solution, was used to generate peptide sequences by MS/MS and *de novo* sequencing as well as database searching with Mascot and BLAST. Using sequence alignment tools, peptides from the lectin protein were aligned with lectin proteins from other vertebrates and peptides from the different digests were overlapped to assemble the protein sequence.

The primary structure of the alligator lectin exhibited more than 50% similarity to human, mouse, rat and frog intelectin-1 protein and less than 15% homology to mannose-binding lectin from other vertebrates. Although the alligator lectin showed a higher degree of similarity to intelectin protein, the differences of the conserved cysteine residues between the alligator and other vertebrate intelectins may result in structural differences. Strong homology to the fibrinogen-like domain of the alligator lectin was observed near the N-terminus region. There were no putative N-glycosylation sites identified in the alligator lectin sequence; however it has potential O-linkages. The isolated alligator lectin exhibited strong binding affinities toward mannose and mannan as compared to other tested carbohydrates.

CHAPTER 7. A MASS SPECTROMETRY APPROACH FOR THE STUDY OF DEGLYCOSYLATED PROTEINS*

The work described in this chapter focuses on preparatory steps to remove salts and buffers for mass spectrometry (MS) analysis after enzymatic or chemical deglycosylation. In this work, the glycosylated protein fetuin and a lectin protein isolated from the serum of *Alligator mississippiensis* were used to evaluate methods for desalting samples after an enzymatic or chemical deglycosylation. Precipitation and dialysis were used to prepare the deglycosylated samples for MS analysis. Both the precipitation and dialysis methods were suitable for sample preparation prior to analysis by matrix assisted laser desorption ionization (MALDI) MS.

7.1 Introduction

Proteomics is the study of the structure and functions of proteins in an organism or tissue, and mass spectrometry is now a routinely used method for protein identification in complex mixtures.[125] Prior to a typical mass spectrometry analysis, separation techniques such as two-dimensional gel electrophoresis and high performance liquid chromatography (HPLC) are used to reduce the complexity of protein mixtures.[125, 325]

Some proteins are post-translationally modified to help regulate gene expression, protein turnover, and cellular structure.[343] Glycosylation is a common post-translational modification that is involved in biological functions such as immune recognition and inflammation.[344] There are two major types of protein carbohydrate linkages: N- and O-linked glycosylation. With N-glycosylation, the glycan is attached to an asparginine residue followed by any amino acid other than proline, which is linked to a serine or threonine residue. During O-glycosylation, the glycan is attached to a serine or threonine residue. Glycoproteins have been characterized by mass spectrometry, but this is challenging due to the heterogeneity of the carbohydrate moieties and

*Reprinted by permission of the Elsevier. The work reported in this chapter has been published in the *Microchemical Journal*.[4]

the complexity of the resulting mass spectra.[345] Glycosylation of the protein may also limit the extent of proteolytic digestion. To avoid this, the N- and/or O-linked carbohydrates can be removed prior to analysis.[346, 347]

Enzymatic or chemical removal of the glycan can be used for analysis of the protein without the attached oligosaccharides. An enzymatic approach using peptide: N-glycosidase (PNGase F)[204] and a chemical approach using trifluoromethanesulfonic acid (TFMS) have been used to remove the oligosaccharides.[205, 348] PNGase F is an enzyme that removes N-linked oligosaccharides, leaving both the protein and oligosaccharides intact for further analysis. There is no enzyme that cleaves O-linked oligosaccharides. TFMS is a non-specific deglycosylating agent that cleaves both N-linked and O-linked oligosaccharides to leave the intact protein.[349]

The proteins used in this study were enzymatically or chemically deglycosylated followed by matrix assisted laser desorption ionization (MALDI) MS. The steps for both chemical and enzymatic deglycosylation have been reported previously,[205, 350, 351] but the deglycosylated product requires cleanup prior to MALDI mass spectrometry. We have developed important sample desalting procedures that follow deglycosylation prior to mass spectrometry analysis, which are not clearly outlined in many procedures. This study focuses on the treatment of deglycosylated proteins for mass spectrometry analysis.

7.2 Experimental

7.2.1 Materials

Fetuin, a plasma glycoprotein produced by the liver, was obtained from Sigma-Aldrich and the lectin protein was isolated from the blood plasma of *Alligator mississippiensis*. Blood from seven American alligators was collected, heparin added for anticoagulation, and the blood allowed to clot overnight at ambient temperature. The serum was collected and the lectin was isolated using mannan-agarose affinity chromatography.

7.2.2 Enzymatic Deglycosylation

Proteins containing oligosaccharides attached through N-linked glycosylation were deglycosylated as described in Section 2.9. Approximately 100 µg of protein was dissolved in 18 µl of nanopure water. Denaturing buffer (2 µl), containing 0.6 mg/mL SDS and 0.5 mg/mL DTT, was added and heated at 100ºC for 10 minutes to increase the rate of deglycosylation. The sample was incubated on ice at 0ºC for five minutes and centrifuged for 10 sec. To the denatured sample, a non-ionic detergent (NP-40) was added to prevent loss of enzyme activity. The PNGaseF enzyme was added to the sample and it was incubated at 37ºC for four hours.[204, 352] The intact protein was isolated from the mixture via protein precipitation, which removed the buffers that can interfere with the MALDI signal. Acetone or trichloroacetic acid (TCA) was used to precipitate the protein.[353]

7.2.3 TCA Precipitation

TCA precipitation was accomplished by the addition of one volume of 60% TCA. The sample was gently vortexed and incubated on ice overnight. The sample mixture was centrifuged at 10,000 g for 30 min and the supernatant was removed. Chilled acetone was added to wash the precipitate and it was centrifuged again at 10,000 g.[80] This procedure was repeated twice. The precipitate was air-dried and suspended in aqueous 0.1% trifluoroacetic acid (TFA) or 0.1% formic acid (FA) for MALDI analysis.

7.2.4 Acetone Precipitation

Acetone precipitation was achieved by the addition of four volumes of acetone to the protein sample, followed by an overnight incubation on ice. The sample was centrifuged at 10,000 g for 30 min at 4ºC. The supernatant was removed and the precipitate was air dried for 30 min, then suspended in the same buffer as used above in the TCA precipitation.

7.2.5 Chemical Deglycosylation

In-solution deglycosylation was performed using a deglycosylation kit (Prozyme) containing trifluoromethanesulphonic acid (TFMS). TFMS removes most glycan types from glycoproteins and leaves the protein intact.[205] Briefly, approximately 0.3 mg of protein in a vial was dried overnight using a lyophilizer. The dried protein was capped with a septum and placed on a dry ice-ethanol bath for 20 seconds to reduce the temperature of the reaction. A mixture of TFMS and toluene solution was prepared and 50 µl was slowly added using a gas-tight syringe. Precautions were taken to prevent the dried protein from being exposed to moisture due to the potential for hydrolysis of the peptide bonds.[205, 354] The vials containing the protein with TFMS mixture were incubated at -20ºC for five minutes then shaken briefly to help dissolve the protein. The solvation process was repeated and followed by a four-hour incubation at -20ºC.

7.2.6 Dialysis

Dialysis was performed prior to mass spectrometry analysis. After the four-hour incubation, the vials containing the samples were placed in a dry ice-ethanol bath for 20 seconds, and then 150 µl of pyridine solution was added to each vial. The vials were incubated on dry ice for five minutes, and then in a dry ice-ethanol bath for 15 minutes. A 400 µl quantity of 0.5% ammonium bicarbonate was added to each vial to neutralize the sample. Following deglycosylation, the protein was isolated using dialysis. Each sample was added to a dialysis cassette (Thermo Scientific, Rockford, IL, USA) and dialyzed against 10 mM ammonium formate for two hours at room temperature. The buffer was changed and dialysis was continued for an additional two hours. After a final buffer change, the sample was dialyzed overnight at room temperature.

7.2.7 Mass Spectrometry

Glycoproteins and deglycosylated proteins were analyzed in linear MS mode with the Bruker UltrafleXtreme MALDI mass spectrometer. The glycoproteins and the deglycosylated proteins (after dialysis or precipitation) were mixed on a steel target using a 1:1 ratio of sample and saturated matrix solution. The matrices used were 2-hydroxy-5-methoxybenzoic acid (sDHB) and 2,4-dimethoxy-3-hydroxycinnamic acid (sinapininc acid, SA). The samples were analyzed in positive ion mode and 500 shot averages were collected using a mass range from 30 to 100 kDa. Masses were determined using FlexAnalysis software.

7.3 Results and Discussion

Two methods were used for deglycosylation: PNGaseF and trifluoromethanesulfonic acid. The N- and O-linked glycoprotein protein, fetuin, was enzymatically deglycosylated to remove N-linked oligosaccharides and chemical deglycosylation was used to remove both N- and O-linked oligosaccharides. Fetuin is a glycoprotein with 359 amino acids containing N-linked glycosylation sites at Asn 99, 156 and 176, and numerous O-linked glycosylation sites.[355] As seen in Figure 7-1a, a mass of 45.3 kDa was measured for fetuin before deglycosylation which is consistent with previous studies.[356] The deglycosylated fetuin was analyzed with MALDI after deglycosylation with PNGaseF and no further cleanup. No signal was observed possibly due to buffers used during the deglycosylation. Using acetone and TCA precipitation independently, the protein was separated from the PNGaseF released oligosaccharides. The deglycosylated protein from the pellet of fetuin gave a 39.5 kDa in the mass spectrum and was compared with the glycosylated fetuin using MALDI and a mass shift of 5.8 kDa was observed in the mass spectrum (Figure 7-1b). Both acetone and TCA precipitation yielded sufficient protein recovery for MALDI analysis and the MALDI results were comparable.

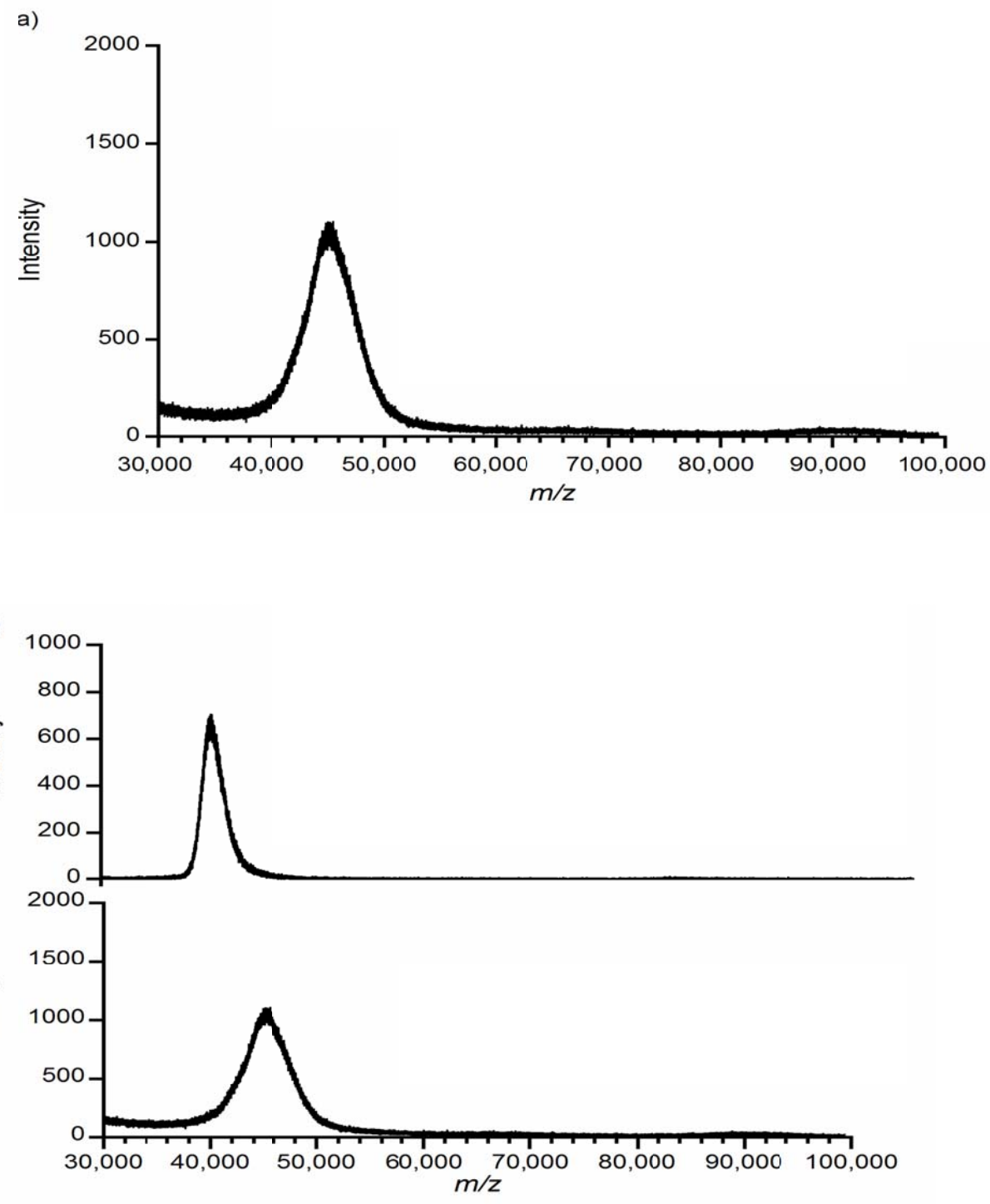
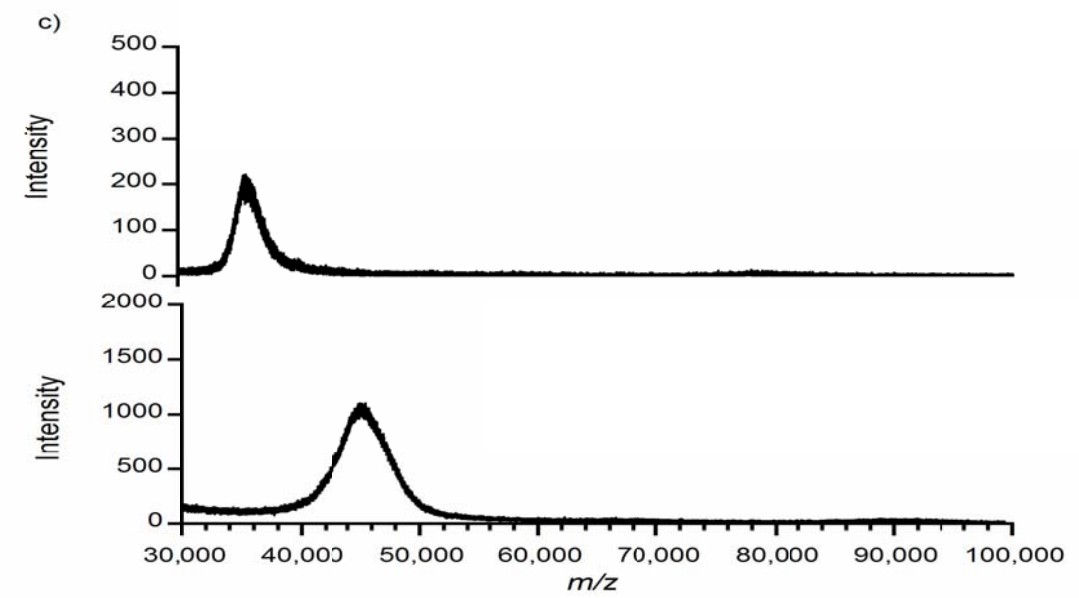


Figure 7-1. MALDI mass spectrometry of fetuin using SA matrix: (a) MALDI spectrum of intact fetuin, (b) MALDI spectra comparing glycosylated and deglycosylated fetuin after acetone precipitation, (c) fetuin after TFMS deglycosylation followed with dialysis.

Figure 7-1. cont'd.



The chemical deglycosylation method using trifluoromethanesulfonic acid was also used to treat the fetuin protein. Following the deglycosylation, a MALDI analysis showed no signal in the protein mass range. However, following dialysis (which was used to desalt and concentrate the deglycosylated protein) a signal was observed. Compared to the precipitation methods, dialysis was more time-consuming and the protein signal observed was lower (Figure 7-1c). However, it removed interfering salts and produced a MALDI signal, although it was not as intense as the signal obtained from the precipitated deglycosylated protein.

This approach was applied to an alligator lectin protein. The non-specific deglycosylation method using TFMS as described above was used to remove oligosaccharides from the protein. The intact protein mass was measured at 34.6 kDa before deglycosylation. After deglycosylation and dialysis a mass of 32.2 kDa was observed in the MALDI mass spectrum (Figure 7-2).

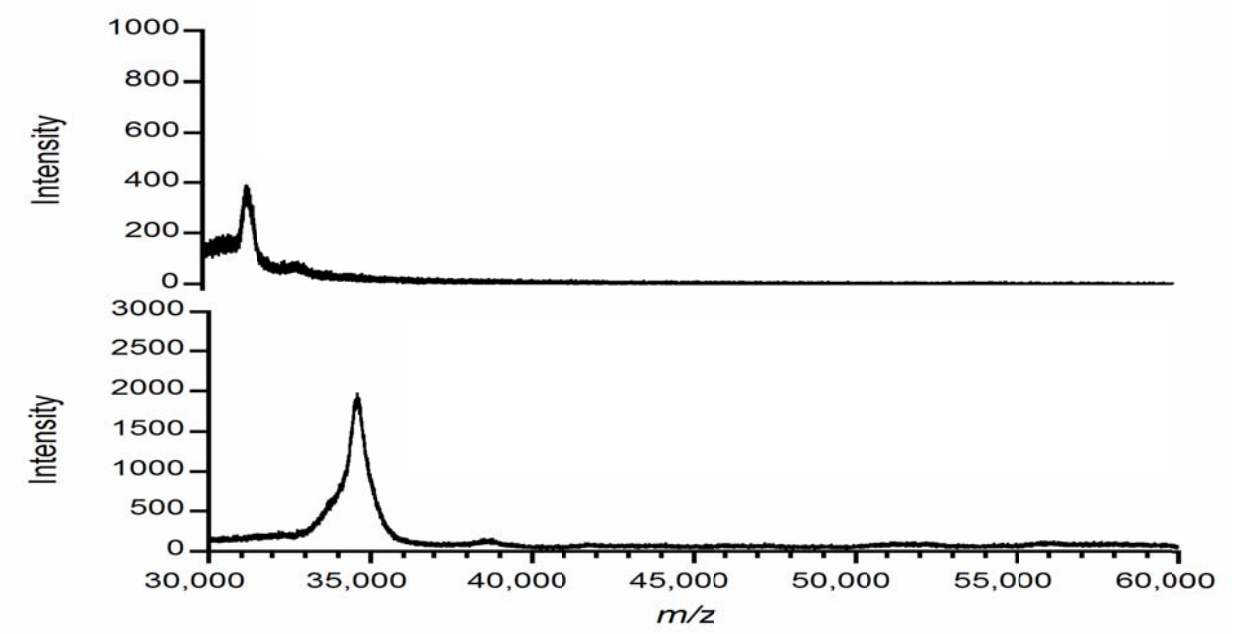


Figure 7-2. Mass spectrometry data before (above) and after (below) deglycosylating the lectin protein using SA matrix.

The mass shift suggests the loss of oligosaccharides and the less intense signal indicates lower protein recovery from the dialysis procedure.

7.4 Summary

Appropriate sample preparation is necessary to obtain mass spectrometry data after deglycosylation. To demonstrate the necessity for sample cleanup prior to mass spectrometry, we have used a glycosylated protein, fetuin, and a lectin protein isolated from *Alligator mississippiensis* serum. Protein precipitation with acetone and TCA was used following deglycosylation to concentrate and desalt the proteins. Alternately, dialysis was used to purify the protein from the other unwanted substances, as demonstrated with the chemical deglycosylation procedure. The precipitation and dialysis methods resulted in efficient desalting and protein concentration. However, the precipitation methods produced a higher protein signal as compared to dialysis.

CHAPTER 8. CONCLUSIONS AND FUTURE DIRECTIONS

In this dissertation, mass spectrometry based proteomics was used to study the alligator blood proteome. Proteins and peptides found in the leukocyte extracts from the alligators were identified using gel electrophoresis techniques, liquid chromatography, and mass spectrometry sequencing, including bottom-up and top-down sequencing. Several peptides and proteins with immunological functions, including lectin and antimicrobial peptides were isolated and identified. The significance of this work lies in that it provides insight on the function and evolution of both innate and adaptive immune systems of crocodilians.

In Chapter 3, the proteome of the leukocytes was investigated because they are one of the most important cells in the body to participate at the site of an infection and play an important role in innate immunity. A 1- and 2-D gel electrophoresis system was used to separate leukocyte proteins, and enzymatic digest of the protein bands and spots using trypsin was applied to cleave the proteins to peptides for peptide sequencing. Using Mascot, a comparison was made between the experimentally observed fragment ions and the predicted fragments for all hypothetical peptides of a particular mass. From the MASOT search, 17 proteins were identified based on sequence homology. For MS/MS data that did not match any proteins in the database, peptide sequences were determined using *de novo* sequencing. Using the *de novo* sequencing results, 23 proteins were identified based on peptide matching using BLAST.

To further investigate peptides from the acidic leukocyte extracts that exhibited antimicrobial activity, a series of separations and antimicrobial tests were performed. Using reversed-phase chromatography, the leukocyte extracts (which consisted of a mixture of peptides and proteins) were separated, collected, and tested for antimicrobial activity. The initial eluents collected within the first 10 min exhibited activity and were further characterized using mass spectrometry and X-ray crystallography for identification, which was described in Chapter 4.

Small molecules less than 500 Da were identified, and the results indicated that the anticoagulant, EDTA, which was used during the blood collection, may have contributed to the antimicrobial activity. Due to the antimicrobial activity interferences from EDTA used in the work in Chapter 4, heparin was used as an alternative anticoagulant for the studies described in Chapter 5.

To investigate the specific peptides and proteins from the alligator leukocytes that exhibit antimicrobial activity, peptides and proteins isolated from alligator leukocyte extracts were separated by reversed phase chromatography as described in Chapter 5. The antimicrobial activity of the peptides from the chromatographic fractions was tested for growth inhibition of various microbes to determine the antimicrobial spectrum of the peptides. Among the microbes studied, growth inhibition was observed for *E. coli, E. cloacae and K. oxytoca* indicating that the antimicrobial peptides from the alligator are active against gram-negative bacteria. Within a fraction that exhibited antimicrobial activity, two peptides at masses 4.7 and 4.9 kDa were identified using mass spectrometry. Due to the close mass of the peptides, ion mobility was used to determine whether the peptides can be separated based on their charge, size and shape. The separated peptide was analyzed using tandem mass spectrometry; however, limited sequence information was determined due to incomplete fragmentation. Therefore, the peptides were subjected to MALDI MS/MS for sequencing. Peptide sequences were determined using *de novo* and manual sequencing. From this work, the first antimicrobial peptides were isolated and partially characterized from *Alligator mississippiensis* and showed to have a net charge of +8 and +13 respectively at pH 5.5, indicating that they are cationic. Based on size and charge, these peptides exhibit strong characteristics similar to previously identified antimicrobial peptides.

In Chapter 6, another class of immunological proteins, lectins, was studied which provides insight on the structure-function relationship of another class of proteins involved

within crocodile's innate immune system. A new calcium-dependent binding lectin protein was isolated from the serum of the *Alligator mississippiensis* by affinity chromatography using mannose columns. The sequence of this novel class of proteins from the blood serum of the alligator was determined with mass spectrometry and Edman sequencing and was used for homologous comparison with other lectins. Molecular masses of the isolated lectin were determined by ESI-MS and were found to be oligomeric which are consistent with the monomer, dimer, trimer and tetramer of the 35 kDa monomer lectin protein. The monomeric protein was enzymatically digested using four different proteases, creating small and large peptides, which were overlapped to enhance protein sequence coverage. The primary sequence of a novel lectin protein isolated from *Alligator miss*issippiensis was determined, and its binding specificity to selected carbohydrates was studied.

An approach to analyze proteins that have been glycosylated also was evaluated. In Chapter 7, preparatory steps used after deglycosylation procedures and before mass spectrometry analysis were described. Here, glycosylated proteins, fetuin and alligator lectin, were deglycosylated and methods for sample desalting and concentrating, such as precipitation and dialysis were used. The dialysis and precipitation methods were both suitable for sample preparation prior to MS analysis. A comparison between precipitation and dialysis procedures showed that precipitation methods produce higher protein signal, indicating a larger protein recovery.

One of the future directions is to investigate other properties of the peptides, other than bacterial properties. Some antimicrobial peptides have shown to have anti-inflammatory, antifungal, antiviral and anti-tumor activity, so these properties should be investigated. Prior work showed that the crude leukocyte extract exhibited antimicrobial activity, so it is hypothesized that the activity may be due to multiple peptides. Therefore, the alligator leukocyte

extract should be further investigated for other antimicrobial peptides or proteins, other than the two peptides isolated in this work and their antimicrobial activity should be characterized, including the minimal inhibitory concentration (MIC) and the antimicrobial spectrum.

The interaction of alligator antimicrobial peptides in the presence of microbes should also be investigated using a technique such as atomic force microscopy (AFM). This study can help provide an understanding of how alligator antimicrobial peptides affect bacterial cells. This effect may be based on the different models proposed as discussed in Section 1.1.1., for example, the carpet model, where peptides cover the membrane like a carpet and destroy the bacteria lipid bilayer.[357] By reducing the AFM cantilever size to approximately 1000 times smaller than the normal size, the imaging speed can be increased without destroying the bacteria.[358] The bacteria can be immobilized on poly-L-lysine coated cover slides (which will provide stronger cell adhesion) and imaged in aqueous solution to keep the cells alive. Alligator antimicrobial peptide can be added to the liquid droplet comprised of bacteria, at a concentration above the minimal inhibitory concentration (MIC) to acquire images.[359] It is also necessary to evaluate the antimicrobial peptides presented in this dissertation for their suitability in commercial products, such as topical creams. They can be added to existing products to enhance the antimicrobial activity of the product and/or they can potentially be used as an antiseptic cream for cuts or an antifungal cream for infections such as athlete's foot.

The research in this dissertation also showed that lectin proteins are present and can be isolated from alligators. Therefore, additional studies should be performed to further characterize the identified lectin in this work, such as investigating the temperature and pH effect, and the effect of the identified alligator lectin on bacteria induced hemagglutination activity; which will indicate whether or not the lectin protects against bacterial induced erythrocyte damage. Because carbohydrates play a vital role in biological processes and diseases, lectins can be used to

identify certain diseases. Depending on the specificity of the alligator lectin it can potentially be used as a means to identify a particular disease state. Finally, antibodies against the alligator lectin discussed in this work can be prepared and used for isolation of lectins from other crocodilian species. This will provide more insight on lectins and their function within crocodilian species.

Lastly, an exhaustive MS-based proteomics study should be performed to map all the proteins involved in the alligator's immune system. The immune system is complex and made of several types of cells and proteins that play different roles in fighting infections. Therefore, a more comprehensive study of the structure-function relationship of the immune related proteins will help provide a better understanding of crocodilian immunity as well as provide more insight on the evolutionary development of reptile's immune system.

REFERENCES

1. Merchant, M. E.; Leger, N.; Jerkins, E.; Mills, K.; Pallansch, M. B.; Paulman, R. L.; Ptak, R. G. Broad spectrum antimicrobial activity of leukocyte extracts from the American alligator (Alligator mississipiensis). *Vet. Immunol. Immunopathol.* **2006,** *110*, 221-228.

2. Merchant, M. E.; Pallansch, M.; Paulman, R. L.; Wells, J. B.; Nalca, A.; Ptak, R. Antiviral activity of serum from the American alligator (Alligator mississippiensis). *Antiviral Res.* **2005,** *66*, 35-38.

3. Merchant, M. E.; Roche, C.; Elsey, R. M.; Prudhomme, J. Antibacterial properties of serum from the American alligator (Alligator mississippiensis). *Comp. Biochem. Physiol. B* **2003,** *136*, 505-513.

4. Darville, L. N. F.; Merchant, M. E.; Murray, K. K. A mass spectrometry approach for the study of deglycosylated proteins. *Microchem. J.* **2011**, 99, 309-311.

5. Ahmed, G. R. V. a. H., *Animal Lectins: A Functional View*. CRC Press Boca Raton, FL, 2008.

6. Darville, L. N. F.; Merchant, M. E.; Hasan, A.; Murray, K. K. Proteome analysis of the leukocytes from the American alligator (Alligator mississippiensis) using mass spectrometry. *Comp. Biochem. Physiol. D* **2010,** *5*, 308-316.

7. Zimmerman, L. M.; Vogel, L. A.; Bowden, R. M. Understanding the vertebrate immune system: insights from the reptilian perspective. *J. Exp. Biol.* **2010,** *213*, 661-671.

8. Hughes, A. L.; Yeager, M. Molecular evolution of the vertebrate immune system. *Bioessays* **1997,** *19*, 777-786.

9. Merchant, M.; Kinney, C.; Sanders, P. Differential protein expression in alligator leukocytes in response to bacterial lipopolysaccharide injection. *Comp. Biochem. Physiol. D* **2009,** *4*, 300-304.

10. Merchant, M. E.; Mills, K.; Leger, N.; Jenkins, E.; Vliet, K. A.; McDaniel, N. Comparisons of innate immune activity of all known living crocodylian species. *Comp. Biochem. Physiol. B* **2006,** *143*, 133-137.

11. Merchant, M.; Thibodeaux, D.; Loubser, K.; Elsey, R. M. Amoebacidal effects of serum from the American alligator (alligator mississippiensis). *J. Parasitol.* **2004,** *90*, 1480-1483.

12. Preecharram, S.; Daduang, S.; Bunyatratchata, W.; Araki, T.; Thammasirirak, S. Antibacterial activity from Siamese crocodile (Crocodylus siamensis) serum. *Afr. J. Biotechnol.* **2008,** *7*, 3121-3128.

13. Preecharram, S.; Jearranaiprepame, P.; Daduang, S.; Temsiripong, Y.; Somdee, T.; Fukamizo, T.; Svasti, J.; Araki, T.; Thammasirirak, S. Isolation and characterisation of crocosin, an antibacterial compound from crocodile (Crocodylus siamensis) plasma. *Anim. Sci. J.* **2010,** *81*, 393-401.

14. Merchant, M. E.; Roche, C. M.; Thibodeaux, D.; Elsey, R. M. Identification of alternative pathway serum complement activity in the blood of the American alligator (Alligator mississippiensis). *Comp. Biochem. Physiol. B* **2005,** *141*, 281-288.

15. Salton, M. R. J. Properties of lysozyme and its action on microorganisms *Bacteriol. Rev.* **1957,** *21*, 82-100.

16. Thammasirirak, S.; Ponkham, P.; Preecharram, S.; Khanchanuan, R.; Phonyothee, P.; Daduang, S.; Srisomsap, C.; Araki, T.; Svasti, J. Purification, characterization and comparison of reptile lysozymes. *Comp. Biochem. Physiol. C* **2006,** *143*, 209-217.

17. Supawadee Pata, S. D., Jisnuson Svasti and Sompong Thammasirirak Isolation of lysozyme like protein from crocodile leukocyte extract (Crocodylus siamensis) *KMITL Sci. Technol. J.* **2007,** *7*, 70-85.

18. Seelen, M. A. J.; Roos, A.; Daha, M. R. Role of complement in innate and autoimmunity. *J. Nephrol.* **2005,** *18*, 642-653.

19. Richard Coico, G. S., and Eli Benjamini *Immunology, A Short Course*. Wiley-Liss Publications: Hoboken, NJ, 2003.

20. Montali, R. J. Comparative pathology of inflammation in the higher vertebrates (reptiles, birds and mammals). *J. Comp. Pathol.* **1988,** *99*, 1-26.

21. Sypek, J. P.; Borysenko, M.; Findlay, S. R. Anti-immunoglobulin induced histamine-release from naturally abundant basophils in the snapping turtle, Chelydra serpentina *Dev. Comp. Immunol.* **1984,** *8*, 359-366.

22. Gilman A, G. L., Hardman JG, Limbird LE *Goodman & Gilman's the pharmacological basis of therapeutics*. McGraw-Hill: New York, 2001.

23. Ganz, T. Defensins: Antimicrobial peptides of innate immunity. *Nature Rev. Immunol.* **2003,** *3*, 710-720.

24. Yount, N. Y.; Bayer, A. S.; Xiong, Y. Q.; Yeaman, M. R. Advances in antimicrobial peptide immunobiology. *Biopolymers* **2006,** *84*, 435-458.

25. Lai, R.; Liu, H.; Lee, W. H.; Zhang, Y. An anionic antimicrobial peptide from toad Bombina maxima. *Biochem. Biophys. Res. Commun.* **2002,** *295*, 796-799.

26. Silphaduang, U.; Hincke, M. T.; Nys, Y.; Mine, Y. Antimicrobial proteins in chicken reproductive system. *Biochem. Biophys. Res. Commun.* **2006,** *340*, 648-655.

27. Selsted, M. E.; Novotny, M. J.; Morris, W. L.; Tang, Y. Q.; Smith, W.; Cullor, J. S. Indolicidin, a novel bactericidal tridecapeptide amide from neutrophils. *J. Biol. Chem.* **1992,** *267*, 4292-4295.

28. Frank, R. W.; Gennaro, R.; Schneider, K.; Przybylski, M.; Romeo, D. Amino acid sequences of two proline-rich bactenecins - Antimicrobial peptides of bovine neutrophils. *J. Biol. Chem.* **1990,** *265*, 18871-18874.

29. Gudmundsson, G. H.; Magnusson, K. P.; Chowdhary, B. P.; Johansson, M.; Andersson, L.; Boman, H. G. Structure of the gene for porcine peptide antibiotic PR-39, a cathelin gene family member: comparative mapping of the locus for the human peptide antibiotic FALL-39. *Proc. Natl. Acad. Sci. U. S. A.* **1995,** *92*, 7085-7089.

30. Kacprzyk, L.; Rydengard, V.; Morgelin, M.; Davoudi, M.; Pasupuleti, M.; Malmsten, M.; Schmidtchen, A. Antimicrobial activity of histidine-rich peptides is dependent on acidic conditions. *Biochim. Biophys. Acta - Biomembranes* **2007,** *1768*, 2667-2680.

31. Lehrer, R. I.; Ganz, T. Defensins of vertebrate animals. *Curr. Opin. Immunol.* **2002,** *14*, 96-102.

32. Ganz, T. Defensins: antimicrobial peptides of vertebrates. *C.R. Biol.* **2004,** *327*, 539-549.

33. Stegemann, C.; Kolobov, A.; Leonova, Y. F.; Knappe, D.; Shamova, O.; Ovchinnikova, T. V.; Kokryakov, V. N.; Hoffmann, R. Isolation, purification and de novo sequencing of TBD-1, the first beta-defensin from leukocytes of reptiles. *Proteomics* **2009,** *9*, 1364-1373.

34. Gennaro, R.; Zanetti, M. Structural features and biological activities of the cathelicidin-derived antimicrobial peptides. *Biopolymers* **2000,** *55*, 31-49.

35. Lopez-Garcia, B.; Lee, P. H. A.; Yamasaki, K.; Gallo, R. L. Anti-fungal activity of cathelicidins and their potential role in Candida albicans skin infection. *J. Invest. Dermatol.* **2005,** *125*, 108-115.

36. Wang, Y. P.; Hong, J.; Liu, X. H.; Yang, H. L.; Liu, R.; Wu, J.; Wang, A. L.; Lin, D. H.; Lai, R. Snake Cathelicidin from Bungarus fasciatus Is a Potent Peptide Antibiotics. *Plos One* **2008,** *3*.

37. Tomasinsig, L.; Zanetti, M. The cathelicidins - Structure, function and evolution. *Curr. Protein Pept. Sci.* **2005,** *6*, 23-34.

38. Xiao, Y. J.; Cai, Y. B.; Bommineni, Y. R.; Fernando, S. C.; Prakash, O.; Gilliland, S. E.; Zhang, G. L. Identification and functional characterization of three chicken cathelicidins with potent antimicrobial activity. *J. Biol. Chem.* **2006,** *281*, 2858-2867.

39. Storici, P.; Tossi, A.; Lenarcic, B.; Romeo, D. Purification and structural characterization of bovine cathelicidins, precursors of antimicrobial peptides. *Eur. J. Biochem.* **1996,** *238*, 769-776.

40. Zhao, H.; Gan, T. X.; Liu, X. D.; Jin, Y.; Lee, W. H.; Shen, J. H.; Zhang, Y. Identification and characterization of novel reptile cathelicidins from elapid snakes. *Peptides* **2008,** *29*, 1685-1691.

41. Zanetti, M. Cathelicidins, multifunctional peptides of the innate immunity. *J. Leukocyte Biol.* **2004,** *75*, 39-48.

42. Zasloff, M. Antimicrobial peptides of multicellular organisms. *Nature* **2002,** *415*, 389-395.

43. Matsuzaki, K. Why and how are peptide-lipid interactions utilized for self-defense? Magainins and tachyplesins as archetypes. *Biochim. Biophys. Acta-Biomembranes* **1999,** *1462*, 1-10.

44. Shai, Y. Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by alpha-helical antimicrobial and cell non-selective membrane-lytic peptides. *Biochim. Biophys. Acta-Biomembranes* **1999,** *1462*, 55-70.

45. Westerhoff, H. V.; Juretic, D.; Hendler, R. W.; Zasloff, M. Magainins and the disruption of membrane-linked free-energy transduction *Proc. Natl. Acad. Sci. U. S. A.* **1989,** *86*, 6597-6601.

46. Yang, L.; Weiss, T. M.; Lehrer, R. I.; Huang, H. W. Crystallization of antimicrobial pores in membranes: Magainin and protegrin. *Biophys. J.* **2000,** *79*, 2002-2009.

47. Bierbaum, G.; Sahl, H. G. Induction of autolysis of staphylococci by the basic peptide antibiotics Pep 5 and nisin and their influence on the activity of autolytic enzymes *Arch. Microbiol.* **1985,** *141*, 249-254.

48. Kragol, G.; Lovas, S.; Varadi, G.; Condie, B. A.; Hoffmann, R.; Otvos, L. The antibacterial peptide pyrrhocoricin inhibits the ATPase actions of DnaK and prevents chaperone-assisted protein folding. *Biochemistry* **2001,** *40*, 3016-3026.

49. Hancock, R. E. W. Peptide antibiotics. *Lancet* **1997,** *349*, 418-422.

50. Jacob, L.; Zasloff, M., Potential therapeutic applications of magainins and other antimicrobial agents of animal origin. In *Antimicrobial Peptides*, Marsh, J.; Goode, J. A., Eds. 1994; Vol. 186, pp 197-216.

51. Hancock, R. E. W.; Lehrer, R. Cationic peptides: a new source of antibiotics. *Trends Biotechnol.* **1998,** *16*, 82-88.

52. Darveau, R. P.; Cunningham, M. D.; Seachord, C. L.; Cassianoclough, L.; Cosand, W. L.; Blake, J.; Watkins, C. S. Beta-lactam antibiotics potentiate magainin 2 antimicrobial activity in vitro and in vivo. *Antimicrob. Agents Chemother.* **1991,** *35*, 1153-1159.

53. Lazarev, V. N.; Govorun, V. M. Antimicrobial peptides and their use in medicine. *Appl. Biochem. Microbiol.* **2010,** *46*, 803-814.

54. Stockert, R. J.; Morell, A. G.; Scheinbe.Ih Mammalian hepatic lectin *Science* **1974,** *186*, 365-366.

55. Drickamer, K.; Taylor, M. E. Biology of Animal Lectins. *Annu. Rev. Cell Biol.* **1993,** *9*, 237-264.

56. Drickamer, K. 2 Distinct classes of carbohydrate-recognition domains in animal lectins *J. Biol. Chem.* **1988,** *263*, 9557-9560.

57. Wang, J. L.; Laing, J. G.; Anderson, R. L. Lectins in the cell nucleus. *Glycobiology* **1991,** *1*, 243-252.

58. Kilpatrick, D. C. Animal lectins: a historical introduction and overview. *Biochim. Biophys. Acta - General Subjects* **2002,** *1572*, 187-197.

59. Debray, H. M., J. 1991; Vol. 4, p 51-96.

60. Aversa, F.; Tabilio, A.; Terenzi, A.; Velardi, A.; Falzetti, F.; Giannoni, C.; Iacucci, R.; Zei, T.; Martelli, M. P.; Gambelunghe, C.; Rossetti, M.; Caputo, P.; Latini, P.; Aristei, C.; Raymondi, C.; Reisner, Y.; Martelli, M. F. Successful engraftment of T-cell-depleted haploidentical 3-loci incompatible transplants in leukemia patients by addition of recombinant human granulocyte-colony-stimulating factor-mobilized peripheral-blood progenitor cells to bone-marrow inoculum *Blood* **1994,** *84*, 3948-3955.

61. Kilpatrick, D. C. G., C. *Adv. Lectin Res.* **1992,** *5*, 51.

62. Coico, R.; Sunshine, G. *Immunology : a short course*. Wiley-Blackwell: Hoboken, N.J., 2009.

63. Domon, B.; Aebersold, R. Review - Mass spectrometry and protein analysis. *Science* **2006,** *312*, 212-217.

64. Ganten, D. R., Klaus *Encyclopedic reference of genomics and proteomics in molecular medicine*. Springer: Berlin Heidelberg New York, 2006; Vol. 1.

65. Yarmush, M. L.; Jayaraman, A. Advances in proteomic technologies. *Annu. Rev. Biomed. Eng.* **2002,** *4*, 349-373.

66. Tyers, M.; Mann, M. From genomics to proteomics. *Nature* **2003,** *422*, 193-197.

67. Norin, M.; Sundström, M. Structural proteomics: developments in structure-to-function predictions. *Trends Biotechnol.* **2002,** *20*, 79-84.

68. Kocher, T.; Superti-Furga, G. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat. Methods* **2007,** *4*, 807-815.

69. Edman, P. Method for determination of the amino acid sequence in peptides *Acta Chem. Scand.* **1950,** *4*, 283-293.

70. Fenselau, C. Beyond Gene Sequencing: Analysis of Protein Structure with Mass Spectrometry. *Annu. Rev. Biophys. Biophys. Chem.* **1991,** *20*, 205-220.

71. Yates, J. R. Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* **1998,** *33*, 1-19.

72. Görg, A.; Weiss, W.; Dunn, M. J. Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **2004,** *4*, 3665-3685.

73. Zheng, S. P.; Schneider, K. A.; Barder, T. J.; Lubman, D. M. Two-dimensional liquid chromatography protein expression mapping for differential proteomic analysis of normal and O157 : H7 Escherichia coli. *BioTechniques* **2003,** *35*, 1202.

74. Debouck, C.; Goodfellow, P. N. DNA microarrays in drug discovery and development. *Nat. Genet.* **1999,** *21*, 48-50.

75. Bunney, W. E.; Bunney, B. G.; Vawter, M. P.; Tomita, H.; Li, J.; Evans, S. J.; Choudary, P. V.; Myers, R. M.; Jones, E. G.; Watson, S. J.; Akil, H. Microarray Technology: A Review of New Strategies to Discover Candidate Vulnerability Genes in Psychiatric Disorders. *Am. J. Psychiat.* **2003,** *160*, 657-666.

76. Luo, Z.; Geschwind, D. H. Microarray applications in neuroscience. *Neurobiol. Dis.* **2001,** *8*, 183-193.

77. Romijn, E. P.; Krijgsveld, J.; Heck, A. J. R. Recent liquid chromatographic-(tandem) mass spectrometric applications in proteomics. *J. Chromatogr. A* **2003,** *1000*, 589-608.

78. Adkins, J. N.; Varnum, S. M.; Auberry, K. J.; Moore, R. J.; Angell, N. H.; Smith, R. D.; Springer, D. L.; Pounds, J. G. Toward a human blood serum proteome - Analysis by multidimensional separation coupled with mass spectrometry. *Mol. Cell Proteomics* **2002,** *1*, 947-955.

79. Anderson, N. L.; Anderson, N. G. The human plasma proteome - History, character, and diagnostic prospects. *Mol. Cell Proteomics* **2002,** *1*, 845-867.

80. Jiang, L.; He, L.; Fountoulakis, M. Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. *J. Chromatogr. A* **2004,** *1023*, 317-320.

81. Corthals, G. L.; Wasinger, V. C.; Hochstrasser, D. F.; Sanchez, J. C. The dynamic range of protein expression: A challenge for proteomic research. *Electrophoresis* **2000,** *21*, 1104-1115.

82. Shen, Y. F.; Smith, R. D. Proteomics based on high-efficiency capillary separations. *Electrophoresis* **2002,** *23*, 3106-3124.

83. Mann, M.; Hendrickson, R. C.; Pandey, A. Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **2001,** *70*, 437-473.

84. Gianazza, E.; Arnaud, P. Chromatography of plasma proteins on immobilized Cibacron Blue F3-GA - Mechanism of the molecular interaction. *Biochem. J.* **1982,** *203*, 637-641.

85. Bjorck, L.; Kronvall, G. Purification and some properties of streptococcal protein G, protein-A novel IgG-binding reagent. *J. Immunol.* **1984,** *133*, 969-974.

86. Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006,** *1*, 2856-2860.

87. Rosenfeld, J.; Capdevielle, J.; Guillemot, J. C.; Ferrara, P. In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Anal. Biochem.* **1992,** *203*, 173-179.

88. Medzihradszky, K. F., In-solution digestion of proteins for mass spectrometry. In *Methods Enzymol.*, Burlingame, A. L., Ed. Academic Press: 2005; Vol. Volume 405, pp 50-65.

89. Wang, S.; Regnier, F. E. Proteolysis of whole cell extracts with immobilized enzyme columns as part of multidimensional chromatography. *J. Chromatogr. A* **2001,** *913*, 429-436.

90. Lee, J.; Soper, S. A.; Murray, K. K. Development of an efficient on-chip digestion system for protein analysis using MALDI-TOF MS. *Analyst* **2009,** *134*, 2426-2433.

91. KR, S. K. W., Enzymatic digestion of proteins in solution and in SDS polyacrylamide gels. In *The Protein Protocols Handbook*, 1996.

92. Klammer, A. A.; MacCoss, M. J. Effects of Modified Digestion Schemes on the Identification of Proteins from Complex Mixtures. *J. Proteome Res.* **2006,** *5*, 695-700.

93. Peterson, D. S.; Rohr, T.; Svec, F.; Frechet, J. M. J. Dual-function microanalytical device by in situ photolithographic grafting of porous polymer monolith: Integrating solid-phase extraction and enzymatic digestion for peptide mass mapping. *Anal. Chem.* **2003,** *75*, 5328-5335.

94. Peterson, D. S.; Rohr, T.; Svec, F.; Fréchet, J. M. J. Enzymatic Microreactor-on-a-Chip: Protein Mapping Using Trypsin Immobilized on Porous Polymer Monoliths Molded in Channels of Microfluidic Devices. *Anal. Chem.* **2002,** *74*, 4081-4088.

95. Sanders, G. H. W.; Manz, A. Chip-based microsystems for genomic and proteomic analysis. *TrAC, Trends Anal. Chem.* **2000,** *19*, 364-378.

96. Liu, J. Y.; Lin, S.; Qi, D. W.; Deng, C. H.; Yang, P. Y.; Zhang, X. M. On-chip enzymatic microreactor using trypsin-immobilized superparamagnetic nanoparticles for highly efficient proteolysis. *J. Chromatogr. A* **2007,** *1176*, 169-177.

97. Deutsch, E. W.; Lam, H.; Aebersold, R. Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. *Phycol. Genomics* **2008,** *33*, 18-25.

98. Liebler, D. C. *Introduction to Proteomics: Tools for the New Biology*. Humana Press Inc: Totowa, NJ, 2002.

99. Aitken, A. e. a. *Protein sequencing: A practical approach*. IRL Press: Oxford, 1989.

100. Twyman, R. M. *Principles of proteomics*. BIOS Scientific New York, NY, 2004.

101. Jacobson, G. R.; Schaffer, M. H.; Stark, G. R.; Vanaman, T. C. Specific chemical cleavage in high-yield at amino peptide-bonds of cysteine and cysteine residues *J. Biol. Chem.* **1973,** *248*, 6583-6591.

102. Bornstei.P; Balian, G. Specific nonenzymatic cleavage of bovine ribonuclease with hydroxylamine *J. Biol. Chem.* **1970,** *245*, 4854-4856.

103. Drapeau, G. R., Protease from Staphyloccus aureus. In *Methods Enzymol.*, Laszlo, L., Ed. Academic Press: 1976; Vol. Volume 45, pp 469-475.

104. Drapeau, G. R., Cleavage at glutamic acid with staphylococcal protease. In *Methods Enzymol.*, Hirs, C. H. W.; Serge, N. T., Eds. Academic Press: 1977; Vol. Volume 47, pp 189-191.

105. Villa, S.; De Fazio, G.; Canosi, U. Cyanogen bromide cleavage at methionine residues of polypeptides containing disulfide bonds. *Anal. Biochem.* **1989,** *177*, 161-164.

106. Westermeier, R.; Naven, T. *Proteomics in practice : a laboratory manual of proteome analysis*. Wiley-VCH: Weinheim, 2002.

107. O'Farrell, P. H. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **1975,** *250*, 4007-4021.

108. Gorg, A.; Obermaier, C.; Boguth, G.; Harder, A.; Scheibe, B.; Wildgruber, R.; Weiss, W. The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **2000,** *21*, 1037-1053.

109. Ye, M. L.; Jiang, X. G.; Feng, S.; Tian, R. J.; Zou, H. F. Advances in chromatographic techniques and methods in shotgun proteome analysis. *Trends Anal. Chem.* **2007,** *26*, 80-84.

110. Kastner, M., *Journal of chromatography library* Elsevier: Amsterdam, The Netherlands, 2000; Vol. 61.

111. Dong, M. W. *Modern HPLC for practicing scientists*. Wiley-Interscience: Hoboken, N.J., 2006.

112. Swartz, M. UPLC ™ : An Introduction and Review. *J. Liq. Chromatogr. Relat. Technol.* **2005,** *28*, 1253-1263.

113. de Villiers, A.; Lestremau, F.; Szucs, R.; Gelebart, S.; David, F.; Sandra, P. Evaluation of ultra performance liquid chromatography - Part I. Possibilities and limitations. *J. Chromatogr. A* **2006,** *1127*, 60-69.

114. Swartz, M. E. Ultra performance liquid chromatography (UPLC): An introduction. *Lc Gc North America* **2005**, 8-14.

115. de Villiers, A.; Lestremau, F.; Szucs, R.; Gélébart, S.; David, F.; Sandra, P. Evaluation of ultra performance liquid chromatography: Part I. Possibilities and limitations. *Journal of Chromatography A* **2006,** *1127*, 60-69.

116. Alpert, A. J. Hydrophilic-interaction chromatography for the separation of peptides, nucleic-acids and other polar compounds *J. Chromatogr.* **1990,** *499*, 177-196.

117. Lee, H. J.; Lee, E. Y.; Kwon, M. S.; Paik, Y. K. Biomarker discovery from the plasma proteome using multidimensional fractionation proteomics. *Curr. Opin. Chem. Biol.* **2006,** *10*, 42-49.

118. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989,** *246*, 64-71.

119. Hillenkamp, F.; Karas, M. Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption ionization. *Methods Enzymol.* **1990,** *193*, 280-295.

120. Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. Matrix-assisted laser desorption ionization mass spectrometry of biopolymers. *Anal. Chem.* **1991,** *63*, A1193-A1202.

121. Karas, M.; Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **1988,** *60*, 2299-2301.

122. Dass, C. *Fundamentals of contemporary mass spectrometry*. Wiley-Interscience: Hoboken, N.J., 2007.

123. Biemann, K.; Scoble, H. A. Characterization by tandem mass-spectrometry of structural modifications in proteins *Science* **1987,** *237*, 992-998.

124. Chait, B. T. Mass spectrometry: Bottom-up or top-down? *Science* **2006,** *314*, 65-66.

125. Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003,** *422*, 198-207.

126. Reid, G. E.; McLuckey, S. A. 'Top down' protein characterization via tandem mass spectrometry. *J. Mass Spectrom.* **2002,** *37*, 663-675.

127. Han, X.; Aslanian, A.; Yates Iii, J. R. Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol.* **2008,** *12*, 483-490.

128. VerBerkmoes, N. C.; Bundy, J. L.; Hauser, L.; Asano, K. G.; Razumovskaya, J.; Larimer, F.; Hettich, R. L.; Stephenson, J. L. Integrating "top-down" and "bottom-up" mass spectrometric approaches for proteomic analysis of Shewanella oneidensis. *J. Proteome Res.* **2002,** *1*, 239-252.

129. Wu, C. C.; Yates, J. R. The application of mass spectrometry to membrane proteomics. *Nature Biotechnol.* **2003,** *21*, 262-267.

130. Santoni, V.; Molloy, M.; Rabilloud, T. Membrane proteins and proteomics: Un amour impossible? *Electrophoresis* **2000,** *21*, 1054-1070.

131. Wu, C. C.; MacCoss, M. J. Shotgun proteomics: Tools for the analysis of complex biological systems. *Curr. Opin. Mol. Ther.* **2002,** *4*, 242-250.

132. Wehr, T. Top-down versus bottom-up approaches in proteomics. *Lc Gc North America* **2006,** *24*, 1004.

133. Uetrecht, C.; Rose, R. J.; van Duijn, E.; Lorenzen, K.; Heck, A. J. R. Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* **2010,** *39*, 1633-1655.

134. McLean, J. A.; Ruotolo, B. T.; Gillig, K. J.; Russell, D. H. Ion mobility-mass spectrometry: a new paradigm for proteomics. *Int. J. Mass spectrom.* **2005,** *240*, 301-315.

135. Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; Hill, H. H. Ion mobility-mass spectrometry. *J. Mass Spectrom.* **2008,** *43*, 1-22.

136. Clemmer, D. E.; Jarrold, M. F. Ion mobility measurements and their applications to clusters and biomolecules. *J. Mass Spectrom.* **1997,** *32*, 577-592.

137. Liu, X. Y.; Valentine, S. J.; Plasencia, M. D.; Trimpin, S.; Naylor, S.; Clemmer, D. E. Mapping the human plasma proteome by SCX-LC-IMS-MS. *J. Am. Soc. Mass. Spectrom.* **2007,** *18*, 1249-1264.

138. Hoaglund-Hyzer, C. S.; Lee, Y. J.; Counterman, A. E.; Clemmer, D. E. Coupling ion mobility separations, collisional activation techniques, and multiple stages of MS for analysis of complex peptide mixtures. *Anal. Chem.* **2002,** *74*, 992-1006.

139. Wyttenbach, T.; Kemper, P. R.; Bowers, M. T. Design of a new electrospray ion mobility mass spectrometer. *Int. J. Mass spectrom.* **2001,** *212*, 13-23.

140. Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass spectrom.* **2007,** *261*, 1-12.

141. Mo, W. J.; Karger, B. L. Analytical aspects of mass spectrometry and proteomics. *Curr. Opin. Chem. Biol.* **2002,** *6*, 666-675.

142. Holland, J. F.; Enke, C. G.; Allison, J.; Stults, J. T.; Pinkston, J. D.; Newcome, B.; Watson, J. T. Mass-spectrometry on the chromatographic time scale - realistic expectations *Anal. Chem.* **1983,** *55*, A997-A1004.

143. Woods, A. S.; Ugarov, M.; Egan, T.; Koomen, J.; Gillig, K. J.; Fuhrer, K.; Gonin, M.; Schultz, J. A. Lipid/peptide/nucleotide separation with MALDI-ion mobility-TOF MS. *Anal. Chem.* **2004,** *76*, 2187-2195.

144. Koomen, J. M.; Ruotolo, B. T.; Gillig, K. J.; McLean, J. A.; Russell, D. H.; Kang, M. J.; Dunbar, K. R.; Fuhrer, K.; Gonin, M.; Schultz, J. A. Oligonucleotide analysis with MALDI-ion-mobility-TOFMS. *Anal. Bioanal.Chem.* **2002,** *373*, 612-617.

145. Ruotolo, B. T.; Verbeck, G. F.; Thomson, L. M.; Gillig, K. J.; Russell, D. H. Observation of conserved solution-phase secondary structure in gas-phase tryptic peptides. *J. Am. Chem. Soc.* **2002,** *124*, 4214-4215.

146. Dwivedi, P.; Wu, P.; Klopsch, S. J.; Puzon, G. J.; Xun, L.; Hill, H. H. Metabolic profiling by ion mobility mass spectrometry (IMMS). *Metabolomics* **2008,** *4*, 63-80.

147. Jin, L.; Barran, P. E.; Deakin, J. A.; Lyon, M.; Uhrin, D. Conformation of glycosaminoglycans by ion mobility mass spectrometry and molecular modelling. *Phys. Chem. Chem. Phys.* **2005,** *7*, 3464-3471.

148. Kumar, C.; Mann, M. Bioinformatics analysis of mass spectrometry-based proteomics data sets. *FEBS Lett.* **2009,** *583*, 1703-1712.

149. Cristoni, S.; Bernardi, L. R. Bioinformatics in mass spectrometry data analysis for proteomics studies. *Expert Rev. Proteomic* **2004,** *1*, 469-483.

150. Gevaert, K.; Vandekerckhove, J. Protein identification methods in proteomics. *Electrophoresis* **2000,** *21*, 1145-1154.

151. Cottrell, J. S.; Sutton, C. W., The identification of electrophoretically separated proteins by peptide mass fingerprinting. In 1996; Vol. 61, pp 67-82.

152. Beavis, R. C.; Fenyö, D. Database searching with mass-spectrometric information. *Trends Biotechnol.* **2000,** *18*, 22-27.

153. Lam, H.; Deutsch, E. W.; Eddes, J. S.; Eng, J. K.; King, N.; Stein, S. E.; Aebersold, R. Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* **2007,** *7*, 655-667.

154. Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999,** *20*, 3551-3567.

155. Eng, J. K.; Fischer, B.; Grossmann, J.; MacCoss, M. J. A fast SEQUEST cross correlation algorithm. *J. Proteome Res.* **2008,** *7*, 4598-4602.

156. Craig, R.; Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **2004,** *20*, 1466-1467.

157. Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X. Y.; Shi, W. Y.; Bryant, S. H. Open mass spectrometry search algorithm. *J. Proteome Res.* **2004,** *3*, 958-964.

158. Field, H. I.; Fenyo, D.; Beavis, R. C. RADARS, a bioinformatics solution that automates proteome mass spectral analysis, optimises protein identification, and archives data in a relational database. *Proteomics* **2002,** *2*, 36-47.

159. Zhang, N.; Aebersold, R.; Schwilkowski, B. ProbID: A probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data. *Proteomics* **2002,** *2*, 1406-1412.

160. Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **2002,** *74*, 5383-5392.

161. Magnin, J.; Masselot, A.; Menzel, C.; Colinge, J. OLAV-PMF: A novel scoring scheme for high-throughput peptide mass fingerprinting. *J. Proteome Res.* **2004,** *3*, 55-60.

162. Kapp, E. A.; Schutz, F.; Connolly, L. M.; Chakel, J. A.; Meza, J. E.; Miller, C. A.; Fenyo, D.; Eng, J. K.; Adkins, J. N.; Omenn, G. S.; Simpson, R. J. An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: Sensitivity and specificity analysis. *Proteomics* **2005,** *5*, 3475-3490.

163. Hughes, C.; Ma, B.; Lajoie, G. A. De Novo Sequencing Methods in Proteomics. *Proteome Bioinf.* **2010**, 105-121.

164. Standing, K. G. Peptide and protein de novo sequencing by mass spectrometry. *Curr. Opin. Struct. Biol.* **2003,** *13*, 595-601.

165. Bertsch, A.; Leinenbach, A.; Pervukhin, A.; Lubeck, M.; Hartmer, R.; Baessmann, C.; Elnakady, Y. A.; Muller, R.; Bocker, S.; Huber, C. G.; Kohlbacher, O. De novo peptide sequencing by tandem MS using complementary CID and electron transfer dissociation. *Electrophoresis* **2009,** *30*, 3736-3747.

166. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101*, 9528-9533.

167. Ma, B.; Zhang, K. Z.; Hendrie, C.; Liang, C. Z.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003,** *17*, 2337-2342.

168. Johnson, R. S.; Taylor, J. A. Searching sequence databases via de novo peptide sequencing by tandem mass spectrometry. *Mol. Biotechnol.* **2002,** *22*, 301-315.

169. Frank, A.; Pevzner, P. PepNovo: De novo peptide sequencing via probabilistic network modeling. *Anal. Chem.* **2005,** *77*, 964-973.

170. Dancik, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. De novo peptide sequencing via tandem mass spectrometry. *J. Comput. Biol.* **1999,** *6*, 327-342.

171. Mann, M. A shortcut to interesting human genes: peptide sequence tags, expressed-sequence tags and computers. *Trends Biochem. Sci.* **1996,** *21*, 494-495.

172. Papayannopoulos, I. A. The Interpretation of Coliision-Induced Dissociation Tandem Mass-Spectra of Peptides *Mass Spectrom. Rev.* **1995,** *14*, 49-73.

173. McGinnis, S.; Madden, T. L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* **2004,** *32*, W20-W25.

174. Shevchenko, A.; Sunyaev, S.; Loboda, A.; Shevehenko, A.; Bork, P.; Ens, W.; Standing, K. G. Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time of flight mass spectrometry and BLAST homology searching. *Anal. Chem.* **2001,** *73*, 1917-1926.

175. Waridel, P.; Frank, A.; Thomas, H.; Surendranath, V.; Sunyaev, S.; Pevzner, P.; Shevchenko, A. Sequence similarity-driven proteomics in organisms with unknown genomes by LC-MS/MS and automated de novo sequencing. *Proteomics* **2007,** *7*, 2318-2329.

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

177. Altschul, S. F.; Gish, W., Local alignment statistics. In *Computer Methods for Macromolecular Sequence Analysis*, 1996; Vol. 266, pp 460-480.

178. Olson, G. A.; Hessler, J. R.; Faith, R. E. Techniques for blood collection and intravascular infusion of reptiles *Lab. Anim. Sci.* **1975,** *25*, 783-786.

179. Zippel, K. C.; Lillywhite, H. B.; Mladinich, C. R. J. Anatomy of the crocodilian spinal vein. *J. Morphol.* **2003,** *258*, 327-335.

180. Tomer, K. B.; Moseley, M. A.; Deterding, L. J.; Parker, C. E. Capillary liquid-chromatography mass spectrometry *Mass Spectrom. Rev.* **1994,** *13*, 431-457.

181. Temesi, D.; Law, B. The effect of LC eluent composition on MS responses using electrospray ionization. *Lc Gc North America* **1999,** *17*, 626-632.

182. Shvartsburg, A. A.; Smith, R. D. Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Anal. Chem.* **2008,** *80*, 9689-9699.

183. Lanzillo, J. J.; Stevens, J.; Fanburg, B. L. A comparison of commonly used discontinuous and continuous buffer systems for electrophoresis in sodium dodecyl sulfate-containing polyacrylamide gels. *Electrophoresis* **1980,** *1*, 180-186.

184. Svasti, J.; Panijpan, B. SDS-polyacrylamide gel electrophoresis. A simple explanation of why it works. *J. Chem. Educ.* **1977,** *54*, 560-562.

185. Garfin, D. E. *2-D electrophoresis for proteomics : a methods and product manual*. Bio-Rad: Hercules, California, 2001.

186. Righetti, P. G. Recent developments in electrophoretic methods. *J. Chromatogr. A* **1990,** *516*, 3-22.

187. Mortz, E.; Krogh, T. N.; Vorum, H.; Gorg, A. Improved silver staining protocols for high sensitivity protein identification using matrix-assisted laser desorption/ionization-time of flight analysis. *Proteomics* **2001,** *1*, 1359-1363.

188. Link, A. J. *Methods in Moelcular Biology: 2-D proteome analysis protocols*. Humana Press: Totowa, N.J., 1999.

189. Steinberg, T. H.; Top, K. P. O.; Berggren, K. N.; Kemper, C.; Jones, L.; Diwu, Z. J.; Haugland, R. P.; Patton, W. F. Rapid and simple single nanogram detection of glycoproteins in polyacrylamide gels and on electroblots. *Proteomics* **2001,** *1*, 841-855.

190. Suckau, D.; Resemann, A.; Schuerenberg, M.; Hufnagel, P.; Franzen, J.; Holle, A. A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Anal. Bioanal.Chem.* **2003,** *376*, 952-965.

191. Schafer, R. ultrafleXtreme: Redefining MALDI-TOF-TOF Mass Spectrometry Performance. *Lc Gc Europe* **2009**, 26-27.

192. Bahr, U.; Karas, M.; Hillenkamp, F. Analysis of biopolymers by matrix-assisted laser-desorption ionization (MALDI) mass-spectrometry *Fresenius J. Anal. Chem.* **1994,** *348*, 783-791.

193. Roepstorff, P.; Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **1984,** *11*, 601-601.

194. Biemann, K. Contributions of mass spectrometry to peptide and protein-structure *Biomed. Environ. Mass Spectrom.* **1988,** *16*, 99-111.

195. Wysocki, V. H.; Resing, K. A.; Zhang, Q. F.; Cheng, G. L. Mass spectrometry of peptides and proteins. *Methods* **2005,** *35*, 211-222.

196. Pappin, D. J. C.; Hojrup, P.; Bleasby, A. J. Rapid identification of proteins by peptide-mass fingerprinting *Curr. Biol.* **1993,** *3*, 327-332.

197. Berndt, P.; Hobohm, U.; Langen, H. Reliable automatic protein identification from matrix-assisted laser desorption/ionization mass spectrometric peptide fingerprints. *Electrophoresis* **1999,** *20*, 3521-3526.

198. Bruker, Protein Identification and Detailed Primary Structure Analysis with BioTools and RapiDeNovo. In *Application Note*, Bruker Daltonics.

199. Habermann, B.; Oegema, J.; Sunyaev, S.; Shevchenko, A. The Power and the Limitations of Cross-Species Protein Identification by Mass Spectrometry-driven Sequence Similarity Searches. *Mol. Cell Proteomics* **2004,** *3*, 238-249.

200. Morgenstern, B. DIALIGN: multiple DNA and protein sequence alignment at BiBiServ. *Nucleic Acids Res.* **2004,** *32*, W33-W36.

201. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. Clustal W and clustal X version 2.0. *Bioinf.* **2007,** *23*, 2947-2948.

202. Maley, F.; Trimble, R. B.; Tarentino, A. L.; Plummer, T. H. Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Anal. Biochem.* **1989,** *180*, 195-204.

203. Dwek, R. A.; Edge, C. J.; Harvey, D. J.; Wormald, M. R.; Parekh, R. B. Analysis of glycoprotien-associated oligosaccharides *Annu. Rev. Biochem.* **1993,** *62*, 65-100.

204. Plummer, T. H.; Tarentino, A. L. Purification of the oligosaccharide-cleaving enzymes of Flavobacterium meningosepticum. *Glycobiology* **1991,** *1*, 257-263.

205. Edge, A. S. B. Deglycosylation of glycoproteins with trifluoromethanesulphonic acid: elucidation of molecular structure and function. *Biochem. J.* **2003,** *376*, 339-350.

206. Lehrer, R. I.; Rosenman, M.; Harwig, S.; Jackson, R.; Eisenhauer, P. Ultrasensitive assays for endogenous antimicrobial polypeptides *Journal of Immunological Methods* **1991,** *137*, 167-173.

207. Harwig, S. S. L.; Ganz, T.; Lehrer, R. I., Nuetrophil defensins - purification, characterization, and antimicrobial testing In *Bacterial Pathogenesis, Pt B*, 1994; Vol. 236, pp 160-172.

208. Merchant, M. E.; Pallansch, M.; Paulman, R. L.; Wells, J. B.; Nalca, A.; Ptak, R. Antiviral activity of serum from the American alligator (Alligator mississippiensis). *Antiviral Res* **2005,** *66*, 35-38.

209. Merchant, M. E.; Mills, K.; Leger, N.; Jenkins, E.; Vliet, K. A.; McDaniel, N. Comparisons of innate immune activity of all known living crocodylian species. *Comp Biochem Physiol B, Biochem Mol Biol* **2006,** *143*, 133-137.

210. Mayeaux, M. H.; Winston, G. W. Antibiotic effects on cytochromes P450 content and mixed-function oxygenase (MFO) activities in the American alligator, Alligator mississippiensis. *J. Vet. Pharmacol. Ther.* **1998,** *21*, 274-281.

211. Brown, D. R.; Nogueira, M. F.; Schoeb, T. R.; Vliet, K. A.; Bennett, R. A.; Pye, G. W.; Jacobson, E. R. Pathology of experimental mycoplasmosis in American alligators. *J. Wildl. Dis.* **2001,** *37*, 671-679.

212. Debyser, I. W. J.; Zwart, P. Review of the most important diseases in crocodylia which possibly interfere with human health *Vlaams Diergeneeskundig Tijdschrift* **1991,** *60*, 164-169.

213. Merchant, M.; Thibodeaux, D.; Loubser, K.; Elsey, R. M. Amoebacidal effects of serum from the American alligator (alligator mississippiensis). *J Parasitol* **2004,** *90*, 1480-1483.

214. Merchant, M. E.; Roche, C.; Elsey, R. M.; Prudhomme, J. Antibacterial properties of serum from the American alligator (Alligator mississippiensis). *Comp Biochem Physiol B, Biochem Mol Biol* **2003,** *136*, 505-513.

215. Merchant, M. E.; Leger, N.; Jerkins, E.; Mills, K.; Pallansch, M. B.; Paulman, R. L.; Ptak, R. G. Broad spectrum antimicrobial activity of leukocyte extracts from the American alligator (Alligator mississipiensis). *Vet Immunol Immunopathol* **2006,** *110*, 221-228.

216. Zeya, H. I.; Spitznag.Jk Cationic protein-bearing granules of polymorphonuclear leukocytes - Separation from enzyme-rich granules. *Science* **1969,** *163*, 1069.

217. Hancock, R. E. W.; Lehrer, R. Cationic peptides: a new source of antibiotics. *Trends Biotechnol* **1998,** *16*, 82-88.

218. Bradshaw, J. P. Cationic antimicrobial peptides - Issues for potential clinical use. *Biodrugs* **2003,** *17*, 233-240.

219. Scott, M. G.; Hancock, R. E. W. Cationic antimicrobial peptides and their multifunctional role in the immune system. *Crit. Rev. Immunol.* **2000,** *20*, 407-431.

220. Hancock, R. E. W.; Patrzykat, A. Clinical development of cationic antimicrobial peptides: from natural to novel antibiotics. *Curr. Drug Targets Infect. Disord.* **2002,** *2*, 79-83.

221. Aebersold, R.; Goodlett, D. R. Mass spectrometry in proteomics. *Chem. Rev.* **2001,** *101*, 269-295.

222. Yates, J. R. Mass spectral analysis in proteomics. *Annu Rev Biophys Biomol Struct* **2004,** *33*, 297-316.

223. Tomazella, G. G.; da Silva, I.; Laure, H. J.; Rosa, J. C.; Chammas, R.; Wiker, H. G.; de Souza, G. A.; Greene, L. J. Proteomic analysis of total cellular proteins of human neutrophils. *Proteome Sci.* **2009,** *7*, 32-40.

224. Ambatipudi, K.; Old, J.; Guilhaus, M.; Raftery, M.; Hinds, L.; Deane, E. Proteomic analysis of the neutrophil proteins of the tammar wallaby (Macropus eugenii). *Comp. Biochem. Physiol. D: Genomics Proteomics* **2006,** *1*, 283-291.

225. Adkins, J. N.; Varnum, S. M.; Auberry, K. J.; Moore, R. J.; Angell, N. H.; Smith, R. D.; Springer, D. L.; Pounds, J. G. Toward a human blood serum proteome - Analysis by multidimensional separation coupled with mass spectrometry. *Mol Cell Proteomics* **2002,** *1*, 947-955.

226. Marcus, K.; Immler, D.; Sternberger, J.; Meyer, H. E. Identification of platelet proteins separated by two-dimensional gel electrophoresis and analyzed by matrix assisted laser desorption/ioniztion-time of flight-mass spectrometry and detection of tyrosine-phosphorylated proteins. **2000**, 2622-2636.

227. Thadikkaran, L.; Siegenthaler, M. A.; Crettaz, D.; Queloz, P. A.; Schneider, P.; Tissot, J. D. In *Recent advances in blood-related proteomics*, 2005; 2005; pp 3019-3034.

228. Anderson, N. L.; Polanski, M.; Pieper, R.; Gatlin, T.; Tirumalai, R. S.; Conrads, T. P.; Veenstra, T. D.; Adkins, J. N.; Pounds, J. G.; Fagan, R.; Lobley, A. The human plasma proteome - A nonredundant list developed by combination of four separate sources. *Mol. Cell Proteomics* **2004,** *3*, 311-326.

229. Kakhniashvili, D. G.; Bulla, L. A.; Goodman, S. R. The human erythrocyte proteome - Analysis by ion trap mass spectrometry. *Mol. Cell Proteomics* **2004,** *3*, 501-509.

230. King C. Chan, D. A. L., Denise Hise, Carl F. Schaefer, Zhen Xiao, George M. Janini, Kenneth H. Buetow, Haleem J. Issaq, Timothy D. Veenstra and Thomas P. Conrads Analysis of the human serum proteome *Clin. Proteomics* **2004,** *1*, 101-225.

231. Jacobson, E. R. *Infectious diseases and pathology of reptiles: color atlas and text*. CRC Press: Boca Raton, 2007.

232. Wells, J. M.; McLuckey, S. A. *Collision-induced dissociation (CID) of peptides and proteins*. 2005; Vol. 402, p 148-185.

233. A. Shevchenko, I. C., M. Wilm and M. Mann *De Novo Peptide Sequencing by Nanoelectrospray Tandem Mass Spectrometry Using Triple Quadrupole and Quadrupole/Time-of-Flight Instruments*. Humana Press: 2000; Vol. 146.

234. Bringans, S.; Eriksen, S.; Kendrick, T.; Gopalakrishnakone, P.; Livk, A.; Lock, R.; Lipscombe, R. Proteomic analysis of the venom of Heterometrus longimanus (Asian black scorpion). *Proteomics* **2008,** *8*, 1081-1096.

235. Budnik, B. A.; Olsen, J. V.; Egorov, T. A.; Anisimova, V. E.; Galkina, T. G.; Musolyamov, A. K.; Grishin, E. V.; Zubarev, R. A. De novo sequencing of antimicrobial peptides isolated from the venom glands of the wolf spider Lycosa singoriensis. *J. Mass. Spectrom.* **2004,** *39*, 193-201.

236. Tannu, N.; Hemby, S. De novo protein sequence analysis of Macaca mulatta. *BMC Genomics* **2007,** *8*, 270.

237. Samgina, T. Y.; Artemenko, K. A.; Gorshkov, V. A.; Ogourtsov, S. V.; Zubarev, R. A.; Lebedev, A. T. De novo sequencing of peptides secreted by the skin glands of the Caucasian Green Frog Rana ridibunda. *Rapid Commun. Mass Spectrom.* **2008,** *22*, 3517-3525.

238. Gharahdaghi, F.; Weinberg, C. R.; Meagher, D. A.; Imai, B. S.; Mische, S. M. Mass spectrometric identification of proteins from silver-stained polyacrylamide gel: A method for the removal of silver ions to enhance sensitivity. *Electrophoresis* **1999,** *20*, 601-605.

239. Speicher, K., O. Kolbas, S. Harper, and D. Speicher Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies. *J. Biomol. Tech.* **2000,** *11*, 74–86.

240. Sechi, S.; Chait, B. T. Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Anal. Chem.* **1998,** *70*, 5150-5158.

241. Liu, J.; Bell, A. W.; Bergeron, J. J. M.; Yanofsky, C. M.; Carrillo, B.; Beaudrie, C. E. H.; Kearney, R. E. Methods for peptide identification by spectral comparison. *Proteome Sci.* **2007,** *5*, 3-14.

242. Chalmers, M. J.; Mackay, C. L.; Hendrickson, C. L.; Wittke, S.; Walden, M.; Mischak, H.; Fliser, D.; Just, I.; Marshall, A. G. Combined top-down and bottom-up mass spectrometric approach to characterization of biomarkers for renal disease. *Anal. Chem.* **2005,** *77*, 7163-7171.

243. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990,** *215*, 403-410.

244. The UniProt, C. The Universal Protein Resource (UniProt) in 2010. *Nucl. Acids Res. 38*, D142-148.

245. Hedges, S. B.; Poling, L. L. A molecular phylogeny of reptiles. *Science* **1999,** *283*, 998-1001.

246. Varshavsky, A. V. The early history of the ubiquitin field. *Protein Sci.* **2006,** *15*, 647-654.

247. Ciechanover, A. The ubiquitin-proteasome pathway: on protein death and cell life. *EMBO J.* **1998,** *17*, 7151-7160.

248. Beck, G.; Habicht, G. S. Immunity and the invertebrates - The fabulously complex immune systems of humans and other mammals evolved over hundreds of millions of years-in sometimes surprising ways. *Sci. Am.* **1996,** *275*, 60.

249. Sahu, A.; Lambris, J. D. Structure and biology of complement protein C3, a connecting link between innate and acquired immunity. *Immunol. Rev.* **2001,** *180*, 35-48.

250. Parseghian, M. H.; Luhrs, K. A. In *Beyond the walls of the nucleus: the role of histones in cellular signaling and innate immunity*, 2006; Natl Research Council Canada-N R C Research Press: 2006; pp 589-604.

251. Kabsch, W.; Vandekerckhove, J. Structure and function of actin *Annu. Rev. Biophys. Biomol. Struct.* **1992,** *21*, 49-76.

252. Cooper, G. M. *The Cell: A Molecular Appraoch*. Sinauer Associates: Sunderland, 2000.

253. Schmidt, J. M.; Zhang, J. W.; Lee, H. S.; Stromer, M. H.; Robson, R. M. Interaction of talin with actin: Sensitive modulation of filament crosslinking activity. *Arch. Biochem. Biophys.* **1999,** *366*, 139-150.

254. Burridge, K.; Feramisco, J. R. Non-muscle alpha-actinins are calcium-sensitive actin-binding proteins *Nature* **1981,** *294*, 565-567.

255. Bonnemann, C. G.; Thompson, T. G.; van der Ven, P. F. M.; Goebel, H. H.; Warlo, I.; Vollmers, B.; Reimann, J.; Herms, J.; Gautel, M.; Takada, F.; Beggs, A. H.; Furst, D. O.; Kunkel, L. M.; Hanefeld, F.; Schroder, R. Filamin C accumulation is a strong but

nonspecific immunohistochemical marker of core formation in muscle. *J. Neurol. Sci.* **2003,** *206*, 71-78.

256. Jones, S. L.; Wang, J.; Turck, C. W.; Brown, E. J. A role for the actin-bundling protein L-plastin in the regulation of leukocyte integrin function. *Proc. Natl. Acad. Sci. U.S.A.* **1998,** *95*, 9331-9336.

257. Lippolis, J. D.; Reinhardt, T. A. Proteornic survey of bovine neutrophils. *Vet. Immunol. Immunopathol.* **2005,** *103*, 53-65.

258. Gerke, V.; Moss, S. E. Annexins: From structure to function. *Physiol. Rev.* **2002,** *82*, 331-371.

259. Goldmann, W. H.; Ingber, D. E. Intact vinculin protein is required for control of cell shape, cell mechanics, and rac-dependent lamellipodia formation. *Biochem. Biophys. Res. Commun.* **2002,** *290*, 749-755.

260. Mor-Vaknin, N.; Punturieri, A.; Sitwala, K.; Markovitz, D. M. Vimentin is secreted by activated macrophages. *Nat. Cell Biol.* **2003,** *5*, 59-63.

261. Boguski, M. S.; McCormick, F. Proteins regulating Ras and its relatives *Nature* **1993,** *366*, 643-654.

262. Pingoud, A.; Jeltsch, A. Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* **2001,** *29*, 3705-3727.

263. Poelstra, K.; Bakker, W. W.; Klok, P. A.; Hardonk, M. J.; Meijer, D. K. F. A physiologic function for alkaline phosphatase: Endotoxin detoxification. *Lab. Invest.* **1997,** *76*, 319-327.

264. Silver, L. L.; Bostian, K. A. Discovery and development of new antibiotics - The problem of antibiotic-reistance *Antimicrob. Agents Chemother.* **1993,** *37*, 377-383.

265. Alanis, A. J. Resistance to antibiotics: Are we in the post-antibiotic era? *Arch. Med. Res.* **2005,** *36*, 697-705.

266. Mark E. Merchant, N. L., Erin Jerkins, Kaili Mills, Melanie B. Pallansch, Robin L. Paulman, Roger G. Ptak Broad spectrum antimicrobial activity of leukocyte extracts from the American alligator (Alligator mississippiensis). *Vet. Immunol. Immunopathol.* **2006,** *110*, 221-228.

267. Vaara, M. Agents that increase the permeability of the outer-membrane *Microbiol. Rev.* **1992,** *56*, 395-411.

268. Rawal, B. D.; Owen, W. R. Combined Action of Sulfamethoxazole, Trimethoprim, and Ethylenediaminetetraacetic Acid on Pseudomonas-Aeruginosa C. *Appl. Microbiol.* **1971,** *21*, 367-368.

269. Weiser, R.; Wimpenny, J.; Asscher, A. W. Synergistic Effect of Edetic-Acid/Antibiotic Combinations on Pseudomonas Aeruginosa. *Lancet* **1969,** *2*, 619-620.

270. Nagai, T.; Oita, S. Anti-Helicobacter pylori activity of EDTA. *J. Gen. Appl. Microbiol.* **2004,** *50*, 115-118.

271. Ko, K. Y.; Mendoncam, A. F.; Ismail, H.; Ahn, D. U. Ethylenediaminetetraacetate and lysozyme improves antimicrobial activities of ovotransferrin against Escherichia coli O157:H7. *Poult. Sci.* **2009,** *88*, 406-414.

272. Marx, V. Watching peptide drugs grow up. *Chem. Eng. News* **2005,** *83*, 17-24.

273. Rozansky, R.; Bachrach, U.; Grossowicz, N. Studies on the antibacterial action of spermine *J. Gen. Microbiol.* **1954,** *10*, 11-16.

274. Grossowicz, N.; Razin, S.; Rozansky, R. Factors influencing the antibacterial action of spermine and spermidine on staphylococcus-aureus *J. Gen. Microbiol.* **1955,** *13*, 436-441.

275. Fair, W. R.; Wehner, N. Antibacterial action of spermine-effect on urinary tract pathogens *Appl. Microbiol.* **1971,** *21*, 6-8.

276. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976,** *72*, 248-254.

277. Shafer, W. M., Designer Assays for Antimicrobial Peptides. In *Antibacterial peptide protocols*, Shafer, W. M., Ed. Humana Press: 1997; pp 169-216.

278. Lehrer, R. I.; Rosenman, M.; Harwig, S. S. S. L.; Jackson, R.; Eisenhauer, P. Ultrasensitive assays for endogenous antimicrobial polypeptides. *J. Immunol. Methods* **1991,** *137*, 167-173.

279. Gordon, Y. J.; Romanowski, E. G.; McDermott, A. M. A review of antimicrobial peptides and their therapeutic potential as anti-infective drugs. *Curr. Eye Res.* **2005,** *30*, 505-515.

280. Garcia-Olmedo, F.; Molina, A.; Alamillo, J. M.; Rodriguez-Palenzuela, P. Plant defense peptides. *Biopolymers* **1998,** *47*, 479-491.

281. Amiche, M.; Seon, A. A.; Wroblewski, H.; Nicolas, P. Isolation of dermatoxin from frog skin, an antibacterial peptide encoded by a novel member of the dermaseptin genes family. *Eur. J. Biochem.* **2000,** *267*, 4583-4592.

282. Carlsson, A.; Nystrom, T.; de Cock, H.; Bennich, H. Attacin - an insect immune protein - binds LPS and triggers the specific inhibition of bacterial outer-membrane protein synthesis. *Microbiology-Sgm* **1998,** *144*, 2179-2188.

283. Levy, O. Antimicrobial proteins and peptides of blood: templates for novel antimicrobial agents. *Blood* **2000,** *96*, 2664-2672.

284. Martin, E.; Ganz, T.; Lehrer, R. I. Defensins and other endogenous peptide antibiotics of vertebrates *J. Leukoc. Biol.* **1995,** *58*, 128-136.

285. Ganz, T. Immunology - Defensins and host defense. *Science* **1999,** *286*, 420-421.

286. Selsted, M. E.; Tang, Y. Q.; Morris, W. L.; McGuire, P. A.; Novotny, M. J.; Smith, W.; Henschen, A. H.; Cullor, J. S. Purification, primary structures, and antibacterial activities of beta-defensins, a new family of antimicrobial peptides from bovine neutrophils *J. Biol. Chem.* **1993,** *268*, 6641-6648.

287. Herve-Grepinet, V.; Rehault-Godbert, S.; Labas, V.; Magallon, T.; Derache, C.; Lavergne, M.; Gautron, J.; Lalmanach, A. C.; Nys, Y. Purification and Characterization of Avian beta-Defensin 11, an Antimicrobial Peptide of the Hen Egg. *Antimicrob. Agents Chemother.* **2010,** *54*, 4401-4408.

288. Bensch, K. W.; Raida, M.; Magert, H. J.; Schulzknappe, P.; Forssmann, W. G. HBD-1 - A novel beta-defensin from human plasma. *FEBS Lett.* **1995,** *368*, 331-335.

289. Pata, S.; Yaraksa, N.; Daduang, S.; Temsiripong, Y.; Svasti, J.; Araki, T.; Thammasirirak, S. Characterization of the novel antibacterial peptide Leucrocin from crocodile (Crocodylus siamensis) white blood cell extracts. *Dev. Comp. Immunol.* **2011,** *35*, 545-553.

290. Preecharram, S.; Daduang, S.; Bunyatratchata, W.; Araki, T.; Thammasirirak, S. Antibacterial activity from Siamese crocodile (Crocodylus siamensis) serum. *African Journal of Biotechnology* **2008,** *7*, 3121-3128.

291. Darville, L. N. F.; Merchant, M. E.; Hasan, A.; Murray, K. K. Proteome analysis of the leukocytes from the American alligator (Alligator mississippiensis) using mass spectrometry. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* **2010,** *5*, 308-316.

292. Wang, Z.; Wang, G. S. APD: the Antimicrobial Peptide Database. *Nucleic Acids Res.* **2004,** *32*, D590-D592.

293. Torrent, M.; Andreu, D.; Nogues, V. M.; Boix, E. Connecting Peptide Physicochemical and Antimicrobial Properties by a Rational Prediction Model. *Plos One* **2011,** *6*.

294. Boman, H. G. Antibacterial peptides: basic facts and emerging concepts. *J. Intern. Med.* **2003,** *254*, 197-215.

295. Lu, X. Y.; Kurago, Z.; Brogden, K. A. Effects of polymicrobial communities on host immunity and response. *FEMS Microbiol. Lett.* **2006,** *265*, 141-150.

296. Chen, Y. X.; Guarnieri, M. T.; Vasil, A. I.; Vasil, M. L.; Mant, C. T.; Hodges, R. S. Role of peptide hydrophobicity in the mechanism of action of alpha-helical antimicrobial peptides. *Antimicrob. Agents Chemother.* **2007,** *51*, 1398-1406.

297. Drickamer, K.; Taylor, M. E. Biology of animal lectins. *Annu. Rev. Cell Biol.* **1993,** *9*, 237-264.

298. Lis, H.; Sharon, N. Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* **1998,** *98*, 637-674.

299. Dahms, N. M.; Hancock, M. K. P-type lectins. *BBA-Gen. Subjects.* **2002,** *1572*, 317-340.

300. Lasky, L. A. Selectin-carbohydrate interactions and the initiation of the inflammatory response *Annu. Rev. Biochem.* **1995,** *64*, 113-139.

301. Sharon, N. Lectins: past, present and future. *Biochem. Soc. Trans.* **2008,** *36*, 1457-1460.

302. Dommett, R. M.; Klein, N.; Turner, M. W. Mannose-binding lectin in innate immunity: past, present and future. *Tissue Antigens* **2006,** *68*, 193-209.

303. Kilpatrick, D. C. Animal lectins: a historical introduction and overview. *BBA-Gen. Subjects.* **2002,** *1572*, 187-197.

304. Koch, A.; Melbye, M.; Sorensen, P.; Homoe, P.; Madsen, H. O.; Molbak, K.; Hansen, C. H.; Andersen, L. H.; Hahn, G. W.; Garred, P. Acute respiratory tract infections and mannose-binding lectin insufficiency during early childhood. *JAMA-J. AM. Med. Assoc.* **2001,** *285*, 1316-1321.

305. Tabona, P.; Mellor, A.; Summerfield, J. A. Mannose-binding proteins involved in first-line host-defense - Evidence from transgenic mice. *Immunology* **1995,** *85*, 153-159.

306. Charles Janeway, P. T., Mark Walport, Mark Shlomchik *Immunobiology*. Garland Science: New York, 2005.

307. Tsuji, S.; Uehori, J.; Matsumoto, M.; Suzuki, Y.; Matsuhisa, A.; Toyoshima, K.; Seya, T. Human intelectin is a novel soluble lectin that recognizes galactofuranose in carbohydrate chains of bacterial cell wall. *J. Biol. Chem.* **2001,** *276*, 23456-23463.

308. Lee, J. K.; Schnee, J.; Pang, M.; Wolfert, M.; Baum, L. G.; Moremen, K. W.; Pierce, M. Human homologs of the Xenopus oocyte cortical granule lectin XL35. *Glycobiology* **2001,** *11*, 65-73.

309. Komiya, T.; Tanigawa, Y.; Hirohashi, S. Cloning of the novel gene intelectin, which is expressed in intestinal paneth cells in mice. *Biochem. Biophys. Res. Commun.* **1998,** *251*, 759-762.

310. Lee, J. K.; Baum, L. G.; Moremen, K.; Pierce, M. The X-lectins: A new family with homology to the Xenopus laevis oocyte lectin XL-35. *Glycoconjugate J.* **2004,** *21*, 443-450.

311. Russell, S.; Young, K. M.; Smith, M.; Hayes, M. A.; Lumsden, J. S. Identification, cloning and tissue localization of a rainbow trout (Oncorhynchus mykiss) intelectin-like protein that binds bacteria and chitin. *Fish Shellfish Immunol.* **2008,** *25*, 91-105.

312. Abe, Y.; Tokuda, M.; Ishimoto, R.; Azumi, K.; Yokosawa, H. A unique primary structure, cDNA cloning and function of a galactose-specific lectin from ascidian plasma. *Eur. J. Biochem.* **1999,** *261*, 33-39.

313. Wells, V.; Mallucci, L. Identification of an autocrine negative growth-factor - Mouse beta-galactoside-binding protein is a cytostatic factor and cell-growth regulator *Cell* **1991,** *64*, 91-97.

314. Perillo, N. L.; Pace, K. E.; Seilhamer, J. J.; Baum, L. G. Apoptosis of T-cells mediated by galectin-1 *Nature* **1995,** *378*, 736-739.

315. Ohannesian, D. W.; Lotan, D.; Thomas, P.; Jessup, J. M.; Fukuda, M.; Gabius, H. J.; Lotan, R. Carcinoembryonic antigen and other glycoconjugates act as ligands for galectin-3 in human colon-carcinoma cells *Cancer Res.* **1995,** *55*, 2191-2199.

316. Fenton, N. B.; Arreguin, L. B.; Mendez, C. F.; Arreguin, E. R. Purification and characterization of liver lectins from a lizard, Sceloporus spinosus. *Prep. Biochem. Biotechnol.* **2004,** *34*, 153-168.

317. Guimaraes-Gomes, V.; Oliveira-Carvalho, A. L.; Junqueira-De-Azevedo, I. D. M.; Dutra, D. L. S.; Pujol-Luz, M.; Castro, H. C.; Ho, P. L.; Zingali, R. B. Cloning, characterization, and structural analysis of a C-type lectin from Bothrops insularis (BiL) venom. *Arch. Biochem. Biophysics.* **2004,** *432*, 1-11.

318. Merchant, M. E.; Roche, C.; Elsey, R. M.; Prudhomme, J. Antibacterial properties of serum from the American alligator (Alligator mississippiensis). *Comp. Biochem. Physiol B. Biochem. Mol. Biol.* **2003,** *136*, 505-513.

319. Fujii, Y.; Kawsar, S. M. A.; Matsumoto, R.; Yasumitsu, H.; Ishizaki, N.; Dogasaki, C.; Hosono, M.; Nitta, K.; Hamako, J.; Taei, M.; Ozeki, Y. A d-galactose-binding lectin purified from coronate moon turban, Turbo (Lunella) coreensis, with a unique amino acid sequence and the ability to recognize lacto-series glycosphingolipids. *Comp. Biochem. Physiol B. Biochem. Mol. Biol.* **2011,** *158*, 30-7.

320. Ashraf, G. M.; Rizvi, S.; Naqvi, S.; Suhail, N.; Bilal, N.; Hasan, S.; Tabish, M.; Banu, N. Purification, characterization, structural analysis and protein chemistry of a buffalo heart galectin-1. *Amino Acids* **2010,** *39*, 1321-1332.

321. Ewart, K. V.; Johnson, S. C.; Ross, N. W. Identification of a pathogen-binding lectin in salmon serum. *Comp. Biochem. Physiol. C* **1999,** *123*, 9-15.

322. Zhu, L.; Song, L. S.; Xu, W.; Qian, P. Y. Identification of a C-type lectin from the bay scallop Argopecten irradians. *Mol. Biol. Rep.* **2009,** *36*, 1167-1173.

323. Wang, N.; Whang, I.; Lee, J. A novel C-type lectin from abalone, Haliotis discus discus, agglutinates Vibrio alginolyticus. *Dev. Comp. Immunol.* **2008,** *32*, 1034-1040.

324. Kang, Y. S.; Kim, Y. M.; Park, K. I.; Cho, S. K.; Choi, K. S.; Cho, M. Analysis of EST and lectin expressions in hemocytes of Manila clams (Ruditapes philippinarum) (Bivalvia : Mollusca) infected with Perkinsus olseni. *Dev. Comp. Immunol.* **2006,** *30*, 1119-1131.

325. Yates, J. R. Mass spectral analysis in proteomics. *Annu. Rev. Biophys. Biomol. Struct.* **2004,** *33*, 297-316.

326. Rameshwaram, N. R.; Karanam, N. K.; Scharf, C.; Volker, U.; Nadimpalli, S. K. Complete primary structure of a newly characterized galactose-specific lectin from the seeds of Dolichos lablab. *Glycoconjugate J.* **2009,** *26*, 161-172.

327. Soler, M. H.; Stoeva, S.; Schwamborn, C.; Wilhelm, S.; Stiefel, T.; Voelter, W. Complete amino acid sequence of the A chain of mistletoe lectin I. *Febs Letters* **1996,** *399*, 153-157.

328. Agrawal, P.; Kumar, S.; Das, H. R. Mass spectrometric characterization of isoform variants of peanut (Arachis hypogaea) stem lectin (SL-I). *J. Proteomics* **2010,** *73*, 1573-1586.

329. Martinez-Cruz, M.; Zenteno, E.; Cordoba, F. Purification and characterization of a galactose-specific lectin from corn (Zea mays) coleoptyle. *BBA-Gen. Subjects.* **2001,** *1568*, 37-44.

330. Hughes, C.; Ma, B.; Lajoie, G. A. *De novo* sequencing methods in proteomics. *Proteome Bioinform.* **2010**, 105-121.

331. Wells, J. M.; McLuckey, S. A. Collision-induced dissociation (CID) of peptides and proteins. *Biol. Mass. Spectrom.* **2005,** *402*, 148-185.

332. Darville, L. N. F.; Merchant, M. E.; Hasan, A.; Murray, K. K. Proteome analysis of the leukocytes from the American alligator (Alligator mississippiensis) using mass spectrometry. *Comp. Biochem. Physiol. D:* **2010,** *5*, 308-316.

333. Rasband, W. S. ImageJ. http://imagej.nih.gov/ij/

334. Speicher KD, K. O., Harper S, Speicher DW Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies. *J. Biomol. Tech.* **2000,** *11*, 74-86.

335. Scheler, C.; Lamer, S.; Pan, Z. M.; Li, X. P.; Salnikow, J.; Jungblut, P. Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization mass spectrometry (MALDI-MS). *Electrophoresis* **1998,** *19*, 918-927.

336. Jeno, P.; Mini, T.; Moes, S.; Hintermann, E.; Horst, M. Internal sequencs from proteins digested in polyacrylamide gels. *Anal. Biochem.* **1995,** *224*, 75-82.

337. Komatsu, S., Edman Sequencing of Proteins from 2D Gels In *Plant Proteomics Methods and Protocols*, Thiellement, H.; Zivy, M.; Damerval, C.; Méchin, V., Eds. Humana Press: Totowa, NJ, 2007; pp 211-217.

338. Lu, J. H.; Le, Y. Ficolins and the fibrinogen-like domain. *Immunobiology* **1998,** *199*, 190-199.

339. Tsuji, S.; Yamashita, M.; Nishiyama, A.; Shinohara, T.; Zhongwei, U.; Myrvik, Q. N.; Hoffman, D. R.; Henriksen, R. A.; Shibata, Y. Differential structure and activity between human and mouse intelectin-1: Human intelectin-1 is a disulfide-linked trimer, whereas mouse homologue is a monomer. *Glycobiology* **2007,** *17*, 1045-1051.

340. Matsushita, M.; Endo, Y.; Taira, S.; Sato, Y.; Fujita, T.; Ichikawa, N.; Nakata, M.; Mizuochi, T. A novel human serum lectin with collagen- and fibrinogen-like domains that functions as an opsonin. *J. Biol. Chem.* **1996,** *271*, 2448-2454.

341. Lu, J. H.; Teh, C.; Kishore, U.; Reid, K. B. M. Collectins and ficolins: sugar pattern recognition molecules of the mammalian innate immune system. *BBA-Gen. Subjects.* **2002,** *1572*, 387-400.

342. Matsushita, M. Ficolins: Complement-activating lectins involved in innate immunity. *J. Innate Immun.* **2010,** *2*, 24-32.

343. Sundararajan, N.; Mao, D. Q.; Chan, S.; Koo, T. W.; Su, X.; Sun, L.; Zhang, J. W.; Sung, K. B.; Yamakawa, M.; Gafken, P. R.; Randolph, T.; McLerran, D.; Feng, Z. D.; Berlin, A. A.; Roth, M. B. Ultrasensitive detection and characterization of posttranslational modifications using surface-enhanced Raman spectroscopy. *Anal. Chem.* **2006,** *78*, 3543-3550.

344. Dwek, R. A. Glycobiology: Toward understanding the function of sugars. *Chem. Rev.* **1996,** *96*, 683-720.

345. Lazar, I. M.; Lazar, A. C.; Cortes, D. F.; Kabulski, J. L. Recent advances in the MS analysis of glycoproteins: Theoretical considerations. *Electrophoresis* **2011,** *32*, 3-13.

346. Medzihradszky, K. F., In-solution digestion of proteins for mass spectrometry. In *Mass Spectrometry: Modified Proteins and Glycoconjugates*, 2005; Vol. 405, pp 50-65.

347. Williams, K. L. S. a. K. R. *The Protein Protocols Handbook*. Humana Press Inc.: Totowa, Vol. Enzymatic Digstion of Proteins in Solution and in SDS Polyacrylamide Gels.

348. Edge, A. S. B.; Faltynek, C. R.; Hof, L.; Reichert, L. E.; Weber, P. Deglycosylation of glycoproteins by trifluoromethanesulfonic acid. *Anal. Biochem.* **1981,** *118*, 131-137.

349. Tarentino, A. L.; Plummer, T. H., Enzymatic deglycosylation of asparagine-linked glycans - Purification, properties, and specificity of oligosaccharide-cleaving enzymes from flavobacterium-meningosepticum. In *Guide to Techniques in Glycobiology*, 1994; Vol. 230, pp 44-57.

350. Patel, T.; Bruce, J.; Merry, A.; Bigge, C.; Wormald, M.; Parekh, R.; Jaques, A. Use of hydrazine to release in intact and unreduced form both N- and O-linked oligosaccharides from glycoproteins. *Biochemistry* **1993,** *32*, 679-693.

351. Lloyd, K. O.; Burchell, J.; Kudryashov, V.; Yin, B. W. T.; TaylorPapadimitriou, J. Comparison of O-linked carbohydrate chains in MUC-1 mucin from normal breast epithelial cell lines and breast carcinoma cell lines - Demonstration of simpler and fewer glycan chains in tumor cells. *J. Biol. Chem.* **1996,** *271*, 33325-33334.

352. Koutsioulis, D.; Landry, D.; Guthrie, E. P. Novel endo-alpha-N-acetylgalactosaminidases with broader substrate specificity. *Glycobiology* **2008,** *18*, 799-805.

353. Fic, E.; Kedracka-Krok, S.; Jankowska, U.; Pirog, A.; Dziedzicka-Wasylewska, M. Comparison of protein precipitation methods for various rat brain structures prior to proteomic analysis. *Electrophoresis* **2010,** *31*, 3573-3579.

354. Fryksdale, B. G.; Jedrzejewski, P. T.; Wong, D. L.; Gaertner, A. L.; Miller, B. S. Impact of deglycosylation methods on two-dimensional gel electrophoresis and matrix assisted laser desorption/ionization-time of flight-mass spectrometry for proteomic analysis. *Electrophoresis* **2002,** *23*, 2184-2193.

355. Dziegielewska, K. M.; Brown, W. M.; Casey, S. J.; Christie, D. L.; Foreman, R. C.; Hill, R. M.; Saunders, N. R. The complete cDNA and amino acid sequence of bovine fetuin. Its homology with alpha 2HS glycoprotein and relation to other members of the cystatin superfamily. *J. Biol. Chem.* **1990,** *265*, 4354-4357.

356. Gey, M.; Unger, K. A strategy for chromatographic and structural analysis of monosaccharide species from glycoproteins. *Fresn. J. Anal. Chem.* **1996,** *356*, 488-494.

357. Oren, Z.; Shai, Y. Mode of action of linear amphipathic alpha-helical antimicrobial peptides. *Biopolymers* **1998,** *47*, 451-463.

358. Podolak, E. Modified AFM captures antimicrobial peptides in action. *BioTechniques* **2010**.

359. Fantner, G. E.; Barbero, R. J.; Gray, D. S.; Belcher, A. M. Kinetics of antimicrobial peptide activity measured on individual bacterial cells using high-speed atomic force microscopy. *Nature Nanotechnol.* **2010,** *5*, 280-285.

APPENDIX A. PROTEINS MATCHED FROM *DE NOVO* SEQUENCING

The peptide sequences in the table were generated from MS/MS data using *de novo* sequencing. *De novo* peptide sequences were subjected to database searching using BLAST to find proteins that contain sequences matching with the determined *de novo* sequences. The results were used to identify biological components that comprise the alligator leukocyte proteome (Chapter 3, Section 3.3).

Table A.1. Proteins identified at single peptide level using BLAST. *De novo* sequenced peptides obtained from gel digest were searched using BLAST and proteins were matched based on sequence similarity and evolutionary relationship using the SwissProt and MSDB database. The top three E-values are reported.

Band Number	*m/z*	Charge	*De novo* Sequence	Protein	Accession Number	Protein Length	Organism	E-Value
B4	1060.6	+2	EINLSPDSTSAVVSGLMVATK					
			EINLSPDSTSAVVSGLMVATK	Fibronectin	P11722.3	1256	Gallus gallus	2e-11 (1st hit)
			EINLSPDSTSAVV+GLMVATK	Fibronectin	Q91289.1	1328	Pleurodeles waltl	9e-11 (2nd hit)
			EINLSPDSTS +VSGLMVATK	Fibronectin	P04937.2	2477	Rattus norvegicus	7e-10 (3rd hit)
B4	536.3	+2	YEVSVYALK					
			YEVSVYALK	Fibronectin	P07589.4	2478	Bos taurus	0.33 (1st hit)
			YEVSVYALK	Fibronectin	P11722.3	1256	Gallus gallus	0.33 (2nd hit)
			YEVSVYALK	Fibronectin	P11276.3	2477	Mus musculus	0.33 (3rd hit)
			YEVSVYALK	Fibronectin	P04937.2	2477	Rattus norvegicus	0.33 (4th hit)
			YEVSVYALK	Fibronectin	Q91289.1	1328	Pleurodeles waltl	0.33 (5th hit)

Table A.1. cont'd.

			YEVSVYALK	Fibronectin	Q28275.2	522	Canis lupus familiaris	0.33 (6th hit)
			YEVSVYALK	Fibronectin	Q28377.2	522	Equus caballus	0.33 (7th hit)
			YEVSVYALK	Fibronectin	P02751.4	2386	Homo sapiens	0.33 (8th hit)
			YEV+VYALK	Fibronectin	Q91740.1	2481	Xenopus laevis	1.9 (9th hit)
			Y+VSVYA	Collagen alpha-1(XII) chain	Q60847.3	3120	Mus musculus	117 (10th hit)
			Y VS+YALK	RNA 2-thiouridine synthesizing protein B	NP_240337.2	95	Buchnera aphidicola	117 (11th hit)
			YEVSVY	Neuronal acetylcholine receptor subunit β-4	NP_434693.1	495	Rattus norvegicus	117 (12th hit)
			YEVSVY	Neuronal acetylcholine receptor subunit β-4	NP_000741.1	498	Homo sapiens	117 (13th hit)
			YEVSVY	Neuronal acetylcholine receptor subunit β-4	NP_683746.1	495	Mus musculus	117 (14th hit)
B4	701.3	+2	YQINQQWER					
			YQINQQWER	Fibronectin	P07589.4	2478	Bos taurus	0.013 (1st hit)
			YQINQQWER	Fibronectin	P11276.3	2477	Mus musculus	0.013 (2nd hit)
			YQINQQWER	Fibronectin	Q91740.1	2481	Xenopus laevis	0.013 (3rd hit)
			YQINQQWER	Fibronectin	P04937.2	2477	Rattus norvegicus	0.013 (4th hit)
			YQINQQWER	Fibronectin	P02751.4	2386	Homo sapiens	0.013 (5th hit)
			INQQWE	Protein lava lamp	NP_525064.1	2779	Drosophila melanogaster	27 (6th hit)
			YQINQQ	DNA topoisomerase 2-associated protein pat1	NP_595976.2	754	Schizosaccharomyces pombe	65 (7th hit)
			YQINQQ	Dynamin-like protein C	XP_645576.2	904	Dictyostelium discoideum	65 (8th hit)
			YQINQQ	Ornithine	YP_672340.1	326	Escherichia coli 536	65

Table A.1. cont'd.

				carbamoyltransferase				(9th hit)
			QI QQW R	Protein Cpn60	NP_849007.1	526	Cyanidioschyzon merolae	65 (10th hit)
			YQINQQ	Thymidylate kinase	NP_077850.1	230	Ureaplasma parvum	65 (11th hit)
			YQINQQ	Ornithine carbamoyltransferase	NP_757196.2	334	Escherichia coli O6	65 (12th hit)
			YQINQQ	Exodeoxyribonuclease V gamma chain	ZP_07124340.1	1122	Escherichia coli K-12	65 (13th hit)
B1	613.9	+2	LIALLEVLSQK					
			LIALLEVLSQK	Filamin-C	NP_001449.3	2725	Homo sapiens	0.019 (1st hit)
			LIALLEVLSQK	Filamin-C	NP_001074654.1	2726	Mus musculus	0.019 (2nd hit)
			LIALLEVLSQK	Filamin-B	Q80X90.2	2602	Mus musculus	0.019 (3rd hit)
			LIALLEVLSQK	Filamin-A	CAT00728.1	2647	Mus musculus	0.019 (4th hit)
			LIALLEVLSQK	Filamin-A	NP_001104026.1	2647	Homo sapiens	0.019 (5th hit)
			LIALLEVLSQK	Filamin-B	NP_001448.2	2602	Homo sapiens	0.019 (6th hit)
			+IALL+VLSQ	Autophagy-related protein 6	A5DIV5.2	461	Pichia guilliermondii	5.1 (7th hit)
			L+ALLEVLS QK	Nesprin-1	NP_892006.2	8797	Homo sapiens	5.1 (8th hit)
			LI+LLEVLS	Bullous pemphigoid antigen 1	AAK83382.1	1678	Mus musculus	9.2 (9th hit)
B1	487.3	+2	VYGPGVEPR					
			+YGPGVEP	50S ribosomal protein L25 1	YP_076110.1	205	Symbiobacterium thermophilum	15 (1st hit)
			+YGPGVEP	50S ribosomal protein L25 2	YP_076123.1	194	Symbiobacterium thermophilum	15 (2nd hit)
			VYGPGVE	Filamin-C	NP_001449.3	2725	Homo sapiens	36 (3rd hit)
			VYGPGVE	Filamin-C	NP_001074654.1	2726	Mus musculus	36 (4th hit)
			VYGPGVE	Actin-binding protein 120	P13466.1	857	Dictyostelium discoideum	36 (5th hit)
			GPGVEPR	Apolipoprotein N-acyltransferase	B9J8B4.1	217	Thermosipho africanus TCF52B	49 (6th hit)

Table A.1. cont'd.

			VYGP VEP	50S ribosomal protein L25	B7IFM6.1	217	Thermosipho africanus TCF52B	49 (7th hit)
B2	736.4	+2	AQQVSQGLDLLTAK					
			AQQVSQGLDLLTAK	Vinculin	P12003.4	1135	Gallus gallus	5e-05 (1st hit)
			AQQVSQGLD+LTAK	Vinculin	P85972.1	1066	Rattus norvegicus	7e-04 (2nd hit)
			AQQVSQGLD+LTAK	Vinculin	P26234.4	1135	Sus scrofa	7e-04 (3rd hit)
			AQQVSQGLD+LTAK	Vinculin	Q64727.4	1066	Mus musculus	7e-04 (4th hit)
			AQQVSQGLD+LTAK	Vinculin	P18206.4	1134	Homo sapiens	7e-04 (5th hit)
			QV+QGLD L AK	Macrophage colony-stimulating factor 1 receptor	Q9I8N6.1	977	Danio rerio	3.8 (6th hit)
B2	739.4	+2	MLGQMTDQVADLR					
			MLGQMTDQVADLR	Vinculin	P85972.1	1066	Rattus norvegicus	2e-05 (1st hit)
			MLGQMTDQVADLR	Vinculin	P26234.4	1135	Sus scrofa	2e-05 (2nd hit)
			MLGQMTDQVADLR	Vinculin	Q64727.4	1066	Mus musculus	2e-05 (3rd hit)
			MLGQMTDQVADLR	Vinculin	P18206.4	1134	Homo sapiens	2e-05 (4th hit)
			LGQMTDQ+ADLR	Vinculin	P12003.4	1135	Gallus gallus	0.006 (5th hit)
			MLGQM +QV	Uncharacterized protein SE_2353	Q8CQQ7.1	300	Staphylococcus epidermidis ATCC 12228	6.8 (6th hit)
			L QMT D +ADLR	Pre-mRNA-processing ATP-dependent RNA helicase	Q2HAD8.1	1064	Chaetomium globosum	6.8 (7th hit)
B2	655.9	+2	TVTAMDVVYALK					
			TVTAMDVVYALK	Histone H4	P83865.2	103	Litopenaeus vannamei	6e-04 (1st hit)
			TVTAMDVVYALK	Histone H4	P82888.2	103	Olisthodiscus luteus	6e-04 (2nd hit)
			TVTAMDVVYALK	Histone H4	P02310.2	103	Tetrahymena pyriformis	6e-04 (3rd hit)
			TVTAMDVVYALK	Histone H4	Q6WV74.3	103	Mytilus chilensis	6e-04 (4th hit)
			TVTAMDVVYALK	Histone H4	Q8MTV8.3	103	Aplysia californica	6e-04 (5th hit)

Table A.1. cont'd.

TVTAMDVVYALK	Histone H4	P50566.2	103	Chlamydomonas reinhardtii	6e-04 (6th hit)
TVTAMDVVYALK	Histone H4	P91890.3	103	Trichogramma cacoeciae	6e-04 (7th hit)
TVTAMDVVYALK	Histone H4	P91882.3	103	Diadromus pulchellus	6e-04 (8th hit)
TVTAMDVVYALK	Histone H4	P70081.3	103	Gallus gallus	6e-04 (9th hit)
TVTAMDVVYALK	Histone H4	Q43083.3	103	Pyrenomonas salina	6e-04 (10th hit)
TVTAMDVVYALK	Histone H4	P91849.3	103	Apis mellifera	6e-04 (11th hit)
TVTAMDVVYALK	Histone H4	P62786.2	103	Triticum aestivum	6e-04 (12th hit)
TVTAMDVVYALK	Histone H4	P59259.2	103	Arabidopsis thaliana	6e-04 (13th hit)
TVTAMDVVYALK	Histone H4	P08436.2	103	Volvox carteri	6e-04 (14th hit)
TVTAMDVVYALK	Histone H4	Q6WV72.3	103	Mytilus trossulus	6e-04 (15th hit)
TVTAMDVVYALK	Histone H4	P62776.2	103	Holothuria tubulosa	6e-04 (1thth hit)
TVTAMDVVYALK	Histone H4	P69151.2	103	Tetrahymena pyriformis	6e-04 (17th hit)
TVTAMDVVYALK	Histone H4	P35057.2	103	Solanum lycopersicum	6e-04 (18th hit)
TVTAMDVVYALK	Histone H4	P35059.2	103	Acropora Formosa	6e-04 (19th hit)
TVTAMDVVYALK	Histone H4	Q27443.2	103	Ascaris suum	6e-04 (20th hit)
TVTAMDVVYALK	Histone H4	Q9U7D0.3	108	Mastigamoeba balamuthi	6e-04 (21st hit)
TVTAMDVVYALK	Histone H4	Q41811.3	103	Zea mays	6e-04 (22nd hit)
TVTAMDVVYALK	Histone H4	P04915.2	103	Physarum polycephalum	6e-04 (23rd hit)
TVTAMDVVYALK	Histone H4	P84040.2	103	Drosophila melanogaster	6e-04 (24th hit)
TVTAMDVVYALK	Histone H4	P40287.1	118	Entamoeba histolytica	6e-04

Table A.1. cont'd.

								(25st hit)
			TVTAMDVVYALK	Histone H4	Q76NW2.1	108	Dictyostelium discoideum	6e-04 (26th hit)
			TVT+MDVVYALK	Histone H4	P27996.2	103	Solaster stimpsoni	0.003 (27th hit)
			TVTA+DVVYALK	Histone H4	P80737.3	103	Blepharisma japonicum	0.011 (28th hit)
			TVTA+DVVYALK	Histone H4	P62790.2	104	Sterkiella nova	0.011 (29th hit)
			TVTA+DVVYALK	Histone H4	Q27765.3	103	Styela plicata	0.011 (30th hit)
			TVTA+DVVYALK	Histone H4	P90516.1	89	Blepharisma japonicum	0.011 (31st hit)
			TVTA+DVVYALK	Histone H4	P62792.2	103	Phanerochaete chrysosporium	0.011 (32nd hit)
			TVTA+DVVYALK	Histone H4	P80739.3	107	Moneuplotes crassus	0.011 (33rd hit)
B2	628.7	+2	LITKAVSASK					
			LITKAVSASK	Histone H1.5	P43276.2	223	Mus musculus	0.49 (1st hit)
			LITKAVSASK	Histone H1.03	P08285.2	224	Gallus gallus	0.49 (2nd hit)
			LITKAVSASK	Histone H1.10	P08286.3	220	Gallus gallus	0.49 (3rd hit)
			LITKAVSASK	Histone H1.11	P08287.2	225	Gallus gallus	0.49 (4th hit)
			LITKAVSASK	Histone H1.01	P08284.2	219	Gallus gallus	0.49 (5th hit)
			LITKAVSASK	Histone H1	P09987.2	218	Gallus gallus	0.49 (6th hit)
			LITKAVSASK	Histone H1.11R	P08288.2	219	Gallus gallus	0.49 (7th hit)
			LITKAV+ASK	Histone H1	P16403.2	213	Homo sapiens	2.9 (8th hit)
			LITKAV+ASK	Histone H1.3	P02251.1	213	Cryctolagus cuniculus	2.9 (9th hit)
			LITKAV+ASK	Histone H1.4	P10412.2	219	Homo sapiens	2.9 (10th hit)
			LITKAV+ASK	Histone H1.3	P16402.2	221	Homo sapiens	2.9 (11th hit)

Table A.1. cont'd.

			LITKAV+ASK	Histone H1.1	P02253.1	104	Bos taurus	2.9 (12th hit)
			LITKAV+ASK	Histone H1.2	P15865.3	219	Rattus norvegicus	2.9 (13th hit)
			LITKAV+ASK	Histone H1.2	P15864.2	212	Mus musculus	2.9 (14th hit)
			LITKAV+ASK	Histone H1.4	P02252.1	73	Oryctolagus cuniculus	2.9 (15th hit)
			LITKAV+ASK	Histone H1.4	P43274.2	219	Mus musculus	2.9 (16th hit)
			LITKAV+ASK	Histone H1	P09426.3	218	Anas platyrhynchos	2.9 (17th hit)
			LITKAV+ASK	Histone H1.5	P16401.3	226	Homo sapiens	2.9 (18th hit)
			LITKAV+ASK	Histone H1	P43277.2	221	Mus musculus	2.9 (19th hit)
			LI KAVSASK	Histone H1	P84408.3	55	Salmo salar	9.3 (20th hit)
			LI KAVSASK	Histone H1	P06350.2	207	Oncorhynchus	9.3 (21st hit)
B6	833.9	+2	FSGSGSGTDFTFTISR					
			FSGSGSGTDFTFTIS	Ig K chain V-I region Lay	P01605.1	108	Homo sapiens	5e-06 (1st hit)
			FSGSGSGTDFTFTIS	Ig K chain V-I region WAT	P80362.1	108	Homo sapiens	5e-06 (2nd hit)
			FSGSGSGT+FTFTIS	Ig K chain V-I region OU	P01606.1	108	Homo sapiens	9e-06 (3rd hit)
			FSGSGSGTDFT TISR	Ig K chain V-III region HIC	P18136.1	129	Homo sapiens	2e-05 (4th hit)
			FSGSGSGTDFT TISR	Ig K chain V-III region Ti	P01622.1	109	Homo sapiens	2e-05 (5th hit)
			FSGSGSGTDFT TISR	Ig K chain V-III region SIE	P01620.1	109	Homo sapiens	2e-05 (6th hit)
			FSGSGSGTDFT TISR	Ig Kchain V-III region HAH	P18135.1	129	Homo sapiens	2e-05 (7th hit)
			FSGSGSGTDFT TISR	Ig K chain V-III region WOL	P01623.1	109	Homo sapiens	2e-05 (8th hit)
			FSGSGSGTDFT TISR	Ig Kchain V-III region GOL	P04206.1	109	Homo sapiens	2e-05 (9th hit)
B6	751.4	+2	VFGGGTKLTVLGQPK					

Table A.1. cont'd.

			VFGGGTKLTVL QPK	Ig Λ chain V-III region LOI	P80748.1	111	Homo sapiens	7e-05 (1st hit)
			VFGGGT+LT+LGQPK	IΛ-like polypeptide 1	P20764.2	209	Mus musculus	6e-04 (2nd hit)
			VFGGGTKLTVLG	Ig Λ-1 chain V regions MOPC	P01724.1	129	Mus musculus	0.006 (3rd hit)
			VFGGGTKLTVLG	Ig Λ chain V-I region BL2	P06316.1	130	Homo sapiens	0.006 (4th hit)
			VFGGGTKLTVLG	Ig Λ chain V-V region DEL	P01719.1	108	Homo sapiens	0.006 (5th hit)
			VFGGGTKLTVLG	Ig Λ chain V-VI region EB4	P06319.1	131	Homo sapiens	0.006 (6th hit)
			VFGGGTKLTVLG	Ig Λchain V-II region VIL	P01711.1	111	Homo sapiens	0.006 (7th hit)
			VFGGGTKLTVLG	Ig Λ-1 chain V region H2020	P01726.1	129	Mus musculus	0.006 (8th hit)
			VFGGGTKLTVLG	Ig Λ-1 chain V region S43	P01727.1	129	Mus musculus	0.006 (9th hit)
			VFGGGTKLTVLG	Ig Λ-1 chain V region S178	P01725.1	110	Mus musculus	0.006 (10th hit)
			VFGGGTKLTVLG	Ig Λ chain V-VI region SUT	P06317.1	111	Homo sapiens	0.006 (11th hit)
B2	510.8	+2	V(I/L)ASFGEAVK					
			VLASFGEAVK	Hemoglobin, β	P02131.1	146	Caiman crocodilus	0.27 (1st hit)
			VLASFGEAVK	Hemoglobin, β	P02130.1	146	Alligator mississippiensis	0.27 (2nd hit)
			VL+SFGEAVK	Hemoglobin, ε	P02128.2	147	Gallus gallus	1.6 (3rd hit)
			VL+SFGEAVK	Hemoglobin, ρ	P02127.2	147	Gallus gallus	1.6 (4th hit)
			VL SFGEAVK	Hemoglobin , ε	Q28496.3	147	Microcebus murinus	2.1 (5th hit)
			VLASFGEAV	Hemoglobin, β	P02129.1	146	Crocodylus niloticus	2.1 (6th hit)
			VL SFGEAVK	Hemoglobin, ε	P19759.2	147	Otolemur crassicaudatus	2.1 (7th hit)
			VL SFGEAVK	Hemoglobin, β	P13274.1	146	Chrysemys picta bellii	2.1 (8th hit)

Table A.1. cont'd.

VL SFGEAVK	Hemoglobin, β	P83133.	146	Dipsochelys dussumieri	2.1 (9th hit)
VL SFGEAVK	Hemoglobin, β	P02125.1	146	Ciconia ciconia	2.1 (10th hit)
VL SFGEAVK	Hemoglobin, β	P14524.1	146	Turdus merula	2.1 (11th hit)
VL SFGEAVK	Hemoglobin, β	P08851.1	146	Accipiter gentilis	2.1 (12th hit)
VL SFGEAVK	Hemoglobin, β	P02116.1	146	Ara ararauna	2.1 (13th hit)
VL SFGEAVK	Hemoglobin, β	P07411.1	146	Vultur gryphus	2.1 (14th hit)
VL SFGEAVK	Hemoglobin, β	P15165.1	146	Apus apus	2.1 (15th hit)
VL SFGEAVK	Hemoglobin, β	P08261.1	146	Larus ridibundus	2.1 (16th hit)
VL SFGEAVK	Hemoglobin, β	P21668.1	146	Psittacula krameri	2.1 (17th hit)
VL SFGEAVK	Hemoglobin, ε	Q95238.3	147	Propithecus verreauxi	2.1 (18th hit)
VL SFGEAVK	Hemoglobin, β	P02120.1	146	Anseranas semipalmata	2.1 (19th hit)
VL SFGEAVK	Hemoglobin, β	P10782.1	146	Phalacrocorax carbo	2.1 (20th hit)
VL SFGEAVK	Hemoglobin, ε	P08223.2	146	Eulemur fulvus fulvus	2.1 (21st hit)
VL SFGEAVK	Hemoglobin, β	P07406.1	146	Passer montanus	2.1 (22nd hit)
VL SFGEAVK	Hemoglobin, β	P68061.1	146	Aegypius monachus	2.1 (23rd hit)
VL SFGEAVK	Hemoglobin, β	P02122.1	146	Aquila chrysaetos	2.1 (24th hit)
VL SFGEAVK	Hemoglobin, ε	Q28338.3	147	Cheirogaleus medius	2.1 (25th hit)
VL SFGEAVK	Hemoglobin, β	P02109.3	147	Didelphis virginiana	2.1 (26th hit)
VL SFGEAVK	Hemoglobin, β	P02121.1	146	Phoenicopterus ruber ruber	2.1 (27th hit)

Table A.1. cont'd.

			VL SFGEAVK	Hemoglobin, β	Q98905.3	147	Geochelone carbonaria	2.1 (28th hit)
			VL SFGEAVK	Hemoglobin, β	P82113.1	146	Stercorarius maccormicki	2.1 (29th hit)
B2	500.8	+2	LSSPISGDPK					
			PISGDPK	Microtubule-actin cross-linking factor 1	Q9UPN3.3	5430	Homo sapiens	54 (1st hit)
			PISGDPK	Microtubule-actin cross-linking factor 1, isoform 4	Q96PK2.2	5938	Homo sapiens	54 (2nd hit)
			PISGDPK	Microtubule-actin cross-linking factor 1	Q9QXZ0.1	5327	Mus musculus	54 (3rd hit)
			SP+SGDPK	TonB-dependent heme receptor A	P44523.1	744	Haemophilus influenza	129 (4th hit)
			L SPIS DP	M-phase inducer phosphatase 1-B	P30309.1	550	Xenopus laevis	175 (5th hit)
			L SPIS DP	M-phase inducer phosphatase 1-A	P30308.1	550	Xenopus laevis	175 (6th hit)
			LSSP I+G+PK	Hemoglobin subunit β-3	P02136.2	147	Rana catesbeiana	175 (7th hit)
			L SPIS DP	M-phase inducer phosphatase 3	P30311.1	572	Xenopus laevis	175 (8th hit)
			L SPIS DP	M-phase inducer phosphatase 2	P30310.1	599	Xenopus laevis	175 (9th hit)
B6	522.8	+2	IMSIVDPNR					
			IMSIVDPNR	α-actinin-1	Q3B7N2.1	892	Bos taurus	0.1 (1st hit)
			IMSIVDPNR	α-actinin-1	Q2PFV7.1	892	Macaca fascicularis	0.1 (2nd hit)
			IMSIVDPNR	α-actinin-1	Q7TPR4.1	892	Mus musculus	0.1 (3rd hit)
			IMSIVDPNR	α-actinin-1	P05094.3	893	Gallus gallus	0.1 (4th hit)
			IMSIVDPNR	α-actinin-1	P12814.2	892	Homo sapiens	0.1 (5th hit)
			IMSIVDPNR	α-actinin-1	Q9Z1P2.1	892	Rattus norvegicus	0.1 (6th hit)
			IMS+VDPN	α-actinin-4	A5D7D1.1|	911	Bos taurus	6.2 (7th hit)
			IMS+VDPN	α -actinin-4	Q9QXQ0.2	911	Rattus norvegicus	6.2 (8th hit)
			IMS+VDPN	α-actinin-4	P57780.1	912	Mus musculus	6.2 (9th hit)

Table A.1. cont'd.

			IMS+VDPN	α-actinin-4	Q5RCS6.1	911	Pongo abelii	15 (10th hit)
			IMS+VDPN	α-actinin-2	P20111.1	897	Gallus gallus	15 (11th hit)
			IMS+VDPN	α-actinin-4	Q43707.2	911	Homo sapiens	15 (12th hit)
			IMS+VDPN	α-actinin-4	Q90734.1	904	Gallus gallus	15 (13th hit)
B5	506.2	+3	QEYDESGPSIVHR					
			QEYDESGPSIVHR	γ-Actin	P20359.2	375	Emericella nidulans	3e-05 (1st hit)
			QEYDESGPSIVHR	Actin	P53455.2	375	Ajellomyces capsulatus G186AR	3e-05 (2nd hit)
			QEYDESGPSIVHR	POTE ankyrin domain family member E	Q6S8J3.3	1075	Homo sapiens	3e-05 (3rd hit)
			QEYDESGPSIVHR	Actin	Q0PGG4.1	375	Bos grunniens	3e-05 (4th hit)
			QEYDESGPSIVHR	β-Actin	P84856.1	361	Chlorocebus pygerythrus	3e-05 (5th hit)
			QEYDESGPSIVHR	Chimeric POTE-actin protein	A5A3E0.2	1075	Homo sapiens	3e-05 (6th hit)
			QEYDESGPSIVHR	γ-Actin	Q5JAK2.1	375	Rana lessonae	3e-05 (7th hit)
			QEYDESGPSIVHR	Actin, cytoplasmic 2	Q6P378.1	375	Xenopus tropicalis	3e-05 (8th hit)
			QEYDESGPSIVHR	K-actin	Q9BYX7.1	375	Homo sapiens	3e-05 (9th hit)
			QEYDESGPSIVHR	Actin	Q2U7A3.1	375	Aspergillus oryzae	3e-05 (10th hit)
			QEYDESGPSIVHR	β-Actin	P84336.1	375	Camelus dromedarius	3e-05 (11th hit)
			QEYDESGPSIVHR	β-Actin	Q6NVA9.1	375	Xenopus tropicalis	3e-05 (12th hit)
			QEYDESGPSIVHR	β-Actin	Q7ZVI7.2	375	Danio rerio	3e-05 (13th hit)
			QEYDESGPSIVHR	Actin, cytoskeletal 3	Q25379.1	172	Lytechinus pictus	3e-05 (14th hit)
			QEYDESGPSIVHR	Actin, cytoplasmic	Q964E1.1	376	Biomphalaria obstructa	3e-05 (15th hit)

Table A.1. cont'd.

QEYDESGPSIVHR	Actin, cytoplasmic	P12716.1	376	Pisaster ochraceus	3e-05 (16th hit)
QEYDESGPSIVHR	Actin-3	P41340.1	376	Limulus polyphemus	3e-05 (17th hit)
QEYDESGPSIVHR	Actin, cytoplasmic A3	P04829.3	376	Bombyx mori	3e-05 (18th hit)
QEYDESGPSIVHR	Actin-2	P26197.1	377	Absidia glauca	3e-05 (19th hit)
QEYDESGPSIVHR	Actin-2	P92176.1	376	Lumbricus	3e-05 (20th hit)
QEYDESGPSIVHR	Actin	P91754.1	372	Lumbricus rubellus	3e-05 (21st hit)
QEYDESGPSIVHR	Actin, cytoskeletal 3A	P53474.1	376	Strongylocentrotus purpuratus	3e-05 (22nd hit)
QEYDESGPSIVHR	Actin, cytoskeletal 3B	P18499.1	376	Strongylocentrotus purpuratus	3e-05 (23rd hit)
QEYDESGPSIVHR	Actin-3	P41113.1	376	Podocoryna carnea	3e-05 (24th hit)
QEYDESGPSIVHR	Actin, cytoskeletal 2A	Q07903.1	376	Strongylocentrotus purpuratus	3e-05 (25th hit)
QEYDESGPSIVHR	Actin, cytoplasmic A3a	Q25010.1	376	Helicoverpa armigera	3e-05 (26th hit)
QEYDESGPSIVHR	Actin, cytoplasmic	Q964E3.1	376	Biomphalaria alexandrina	3e-05 (27th hit)
QEYDESGPSIVHR	γ-Actin	P63256.1	375	Anser anser anser	3e-05 (28th hit)
QEYDESGPSIVHR	Actin	P18602.1	327	Artemia species	3e-05 (29th hit)
QEYDESGPSIVHR	Actin, cytoplasmic	P92179.2	376	Biomphalaria glabrata	3e-05 (30th hit)
QEYDESGPSIVHR	β-Actin	P68142.1	375	Takifugu rubripes	3e-05 (31st hit)
QEYDESGPSIVHR	Actin	P53461.1	376	Halocynthia roretzi	3e-05 (32nd hit)
QEYDESGPSIVHR	Actin, cytoplasmic	Q93131.1	375	Branchiostoma floridae	3e-05 (33rd hit)
QEYDESGPSIVHR	Actin, cytoskeletal 1	P53465.1	376	Lytechinus pictus	3e-05 (34th hit)
QEYDESGPSIVHR	Actin	P10365.1	375	Thermomyces lanuginosus	3e-05

Table A.1. cont'd.

					(35th hit)
QEYDESGPSIVHR	Actin-1	O18499.1	376	Saccoglossus kowalevskii	3e-05 (36th hit)
QEYDESGPSIVHR	β-Actin	Q91ZK5.1	375	Sigmodon hispidus	3e-05 (37th hit)
QEYDESGPSIVHR	Actin, cytoplasmic type 8	P53506.1	376	Xenopus laevis	3e-05 (38th hit)
QEYDESGPSIVHR	Actin-1	P92182.1	376	Lumbricus terrestris	3e-05 (39th hit)
QEYDESGPSIVHR	γ-Actin	Q8JJB8.1	375	Triakis scyllium	3e-05 (40th hit)
QEYDESGPSIVHR	Actin-15A	P10990.1	376	Strongylocentrotus franciscanus	3e-05 (41st hit)
QEYDESGPSIVHR	γ-Actin	Q9UVW9.1	375	Acremonium chrysogenum	3e-05 (42nd hit)
QEYDESGPSIVHR	Actin, cytoskeletal 2	P53466.1	376	Lytechinus pictus	3e-05 (43rd hit)
QEYDESGPSIVHR	Actin-1/2	P41112.1	376	Podocoryna carnea	3e-05 (44th hit)
QEYDESGPSIVHR	Actin, cytoskeletal	P53464.1	376	Heliocidaris tuberculata	3e-05 (45th hit)
QEYDESGPSIVHR	Actin, adductor muscle	Q26065.1	376	Placopecten magellanicus	3e-05 (46th hit)
QEYDESGPSIVHR	Actin, cytoskeletal 1B	P53473.1	376	Strongylocentrotus purpuratus	3e-05 (47th hit)
QEYDESGPSIVHR	β-Actin	P48975.1	375	Cricetulus griseus	3e-05 (48th hit)
QEYDESGPSIVHR	Actin-15B	P69004.1	376	Strongylocentrotus franciscanu	3e-05 (49th hit)
QEYDESGPSIVHR	Actin, muscle	P17304.1	376	Aplysia californica	3e-05 (50th hit)
QEYDESGPSIVHR	β-Actin	P60711.1	375	Rattus norvegicus	3e-05 (51st hit)
QEYDESGPSIVHR	Actin, muscle-type	Q25472.1	378	Molgula oculata	3e-05 (52nd hit)
QEYDESGPSIVHR	Actin	Q8X119.1	375	Exophiala dermatitidis	3e-05 (53rd hit)
QEYDESGPSIVHR	Actin-42A	P02572.3	376	Drosophila melanogaster	3e-05 (54th hit)

Table A.1. cont'd.

QEYDESGPSIVHR	β-Actin	P79818.2	375	Oryzias latipes	3e-05 (55th hit)
QEYDESGPSIVHR	β-Actin	P83750.1	375	Cyprinus carpio	3e-05 (56th hit)
QEYDESGPSIVHR	Actin CyI, cytoplasmic	P69002.1	376	Heliocidaris erythrogramma	3e-05 (57th hit)
QEYDESGPSIVHR	Actin-1	P30163.1	376	Onchocerca volvulus	3e-05 (58th hit)
QEYDESGPSIVHR	Actin-2	P30162.1	376	Onchocerca volvulus	3e-05 (59th hit)
QEYDESGPSIVHR	Actin, cytoplasmic	Q93129.1	375	Branchiostoma belcheri	3e-05 (60th hit)
QEYDESGPSIVHR	Actin, non-muscle	P17126.1	376	Hydra vulgaris	3e-05 (61st hit)
QEYDESGPSIVHR	β-Actin	O93400.1	375	Xenopus laevis	3e-05 (62nd hit)
QEYDESGPSIVHR	Actin-2	O18500.1	376	Saccoglossus kowalevskii	3e-05 (63rd hit)
QEYDESGPSIVHR	Actin, clone 403	P18603.1	376	Artemia species	3e-05 (64th hit)
QEYDESGPSIVHR	β-Actin	P15475.1	376	Xenopus borealis	3e-05 (65th hit)
QEYDESGPSIVHR	Actin, cytoskeletal 1A	P53472.1	376	Strongylocentrotus purpuratus	3e-05 (66th hit)
QEYDESGPSIVHR	β-Actin	O42161.1	375	Salmo salar	3e-05 (67th hit)
QEYDESGPSIVHR	Actin, cytoplasmic	Q964E0.1	376	Biomphalaria tenagophila	3e-05 (68th hit)
QEYDESGPSIVHR	Actin	O13419.1	375	Botryotinia fuckeliana	3e-05 (69th hit)
QEYDESGPSIVHR	β-Actin	P53486.1	375	Takifugu rubripes	3e-05 (70th hit)
QEYDESGPSIVHR	Actin-5	P41339.1	376	Limulus polyphemus	3e-05 (71st hit)
QEYDESGPSIVHR	γ-Actin	Q9URS0.1	375	Penicillium chrysogenum	3e-05 (72nd hit)
QEYDESGPSIVHR	Actin, cytoskeletal	P53463.1	376	Heliocidaris erythrogramma	3e-05 (73rd hit)
QEYDESGPSIVHR	Actin	P50138.1	375	Puccinia graminis	3e-05 (74th hit)

Table A.1. cont'd.

QEYDESGPSIVHR	Actin-1/3	P10983.2	376	Caenorhabditis elegans	3e-05 (75th hit)
QEYDESGPSIVHR	Actin, cytoplasmic type 5	P53478.1	376	Gallus gallus	3e-05 (76th hit)
QEYDESGPSIVHR	β-Actin B	P53485.1	375	Takifugu rubripes	3e-05 (77th hit)
QEYDESGPSIVHR	Actin-3	Q03342.1	309	Echinococcus granulosus	3e-05 (78th hit)
QEYDESGPSIVHR	Actin, muscle	P12717.1	376	Pisaster ochraceus	3e-05 (79th hit)
QEYDESGPSIVHR	β-Actin	P29751.1	375	Oryctolagus cuniculus	3e-05 (80th hit)
QEYDESGPSIVHR	Actin-2	P10984.3	376	Caenorhabditis elegans	3e-05 (81st hit)
QEYDESGPSIVHR	Actin, muscle	Q25381.1	172	Lytechinus pictus	3e-05 (82nd hit)
QEYDESGPSIVHR	Actin-6	P53459.1	373	Diphyllobothrium dendriticum	3e-05 (83rd hit)
QEYDESGPSIVHR	Actin-11	P41341.1	376	Limulus polyphemus	3e-05 (84th hit)
QEYDESGPSIVHR	Actin, cytoplasmic	Q964E2.1	376	Biomphalaria pfeifferi	3e-05 (85th hit)
QEYDESGPSIVHR	Actin	Q24733.1	33	Dictyocaulus viviparus	3e-05 (86th hit)
QEYDESGPSIVHR	Actin, cytoplasmic type 5	P53505.1	376	Xenopus laevis	3e-05 (87th hit)
QEYDESGPSIVHR	Actin, cytoplasmic	Q00215.1	375	Styela plicata	3e-05 (88th hit)
QEYDESGPSIVHR	Actin	O17320.1	376	Crassostrea gigas	3e-05 (89th hit)
QEYDESGPSIVHR	Actin-4	P10986.2	376	Caenorhabditis elegans	3e-05 (90th hit)
QEYDESGPSIVHR	Actin-1	P35432.1	375	Echinococcus granulosus	3e-05 (91st hit)
QEYDESGPSIVHR	Actin, cytoplasmic	O17503.1	375	Branchiostoma lanceolatum	3e-05 (92nd hit)
QEYDESGPSIVHR	Actin, cytoplasmic	Q964D9.1	376	Helisoma trivolvis	3e-05 (93rd hit)

Table A.1. cont'd.

			QEYDESGPSIVHR	Actin, muscle	Q00214.1	379	Styela plicata	3e-05 (94th hit)
			QEYDESGPSIVHR	Actin-5C	P10987.4	376	Drosophila	3e-05 (95th hit)
			QEYDESGP+IVHR	Actin-1	P04751.1	377	Xenopus laevis	2e-04 (96th hit)
			QEYDE+GPSIVHR	α-Actin-1	P53479.1	377	Cyprinus carpio	2e-04 (97th hit)
			QEYDE+GPSIVHR	α-cardiac actin	P04751.1	377	Xenopus laevis	2e-04 (98th hit)
			QEYDE+GPSIVHR	Actin, α cardiac	P53480.	377	Takifugu rubripes	2e-04 (99th hit)
B1	683.9	+2	STDYGILQINSR					
			STDYGILQINSR	Lysozyme C	P84492.1	130	Chelonia mydas	6e-04 (1st hit)
			STDYGILQINSR	Lysozyme C	Q7LZI3.1	129	Tragopan satyra	6e-04 (2nd hit)
			STDYGILQINSR	Lysozyme C	Q7LZQ1.3	131	Pelodiscus sinensis	6e-04 (3rd hit)
			STDYGILQINSR	Lysozyme C	P24364.1	129	Lophura leucomelanos	6e-04 (4th hit)
			STDYGILQINSR	Lysozyme C	P81711.1	129	Syrmaticus soemmerringii	6e-04 (5th hit)
			STDYGILQINSR	Lysozyme C	P00702.2	147	Phasianus colchicus colchicus	6e-04 (6th hit)
			STDYGILQINSR	Lysozyme C	P00703.2	147	Meleagris gallopavo	6e-04 (7th hit)
			STDYGILQINSR	Lysozyme C	Q7LZP9.1	129	Lophophorus impejanus	6e-04 (8th hit)
			STDYGILQINSR	Lysozyme C	P49663.1	130	Phasianus versicolor	6e-04 (9th hit)
			STDYGILQINSR	Lysozyme C	P00698.1	147	Gallus gallus	6e-04 (10th hit)

Table A.1. cont'd.

			STDYGILQINSR	Lysozyme C	P00705.1	147	Anas platyrhynchos	6e-04 (11th hit)
			STDYGILQINSR	Lysozyme C	P19849.1	129	Pavo cristatus	6e-04 (12th hit)
			STDYGILQINSR	Lysozyme C	P00701.2	147	Coturnix japonica	6e-04 (13th hit)
			STDYGILQINSR	Lysozyme C	Q7LZQ3.1	129	Crax fasciolata	6e-04 (14th hit)
			STDYGILQINSR	Lysozyme C	P37156.1	125	Tachyglossus aculeatus aculeatus	6e-04 (15th hit)
			STDYGILQINSR	Lysozyme C	P24533.1	129	Syrmaticus reevesii	6e-04 (16th hit)
			STDYGILQINSR	Lysozyme C	Q7LZT2.1	129	Tragopan temminckii	6e-04 (17th hit)
			STDYGILQINSR	Lysozyme C	P00707.1	129	Ortalis vetula	6e-04 (18th hit)
			STDYGILQINSR	Lysozyme C	P22910.1	129	Chrysolophus amherstiae	6e-04 (19th hit)
			STDYG+LQINSR	Lysozyme C	P00704.2	129	Numida meleagris	0.003 (20th hit)
			STDYG+LQINSR	Lysozyme C	P00699.1	129	Callipepla californica	0.003 (21st hit)
			STDYG+LQINSR	Lysozyme C	P00700.2	129	Colinus virginianus	0.003 (22nd hit)
			STDYGIL+INSR	Lysozyme C	P00706.1	129	Anas platyrhynchos	0.004 (23rd hit)
			STDYGIL+INSR	Lysozyme C	Q7LZQ2.1	129	Aix sponsa	0.004 (24th hit)
B1	503.3	+2	WDAWDALK					
			WDAW+ALK	Acyl-CoA-binding protein	Q9PRL8.1	86	Gallus gallus	1.7 (1st hit)
			WDAW+AL	Peroxisomal 3,2-trans-enoyl-CoA isomerase	Q9WUR2.2	391	Mus musculus	13 (2nd hit)
			WDAW+AL	Peroxisomal 3,2-trans-enoyl-CoA isomerase	Q5XIC0.1	391	Rattus norvegicus	13 (3rd hit)

Table A.1. cont'd.

			WDAW+AL	Peroxisomal 3,2-trans-enoyl-CoA isomerase	O75521.4	394	Homo sapiens	13 (4th hit)
			WDAW+ LK	Acyl-CoA-binding protein	P12026.2	87	Sus scrofa	18 (5th hit)
			WDAW+ LK	Acyl-CoA-binding protein	Q8WN94.3	87	Oryctolagus cuniculus	18 (6th hit)
			WDAW+ LK	Acyl-CoA-binding protein	P07108.2	87	Homo sapiens	18 (7th hit)
			WDAW+ LK	Acyl-CoA-binding protein	P07107.2	87	Bos taurus	18 (8th hit)
B4	509.8	+2	IGTMLPMQK					
			I TMLPM QK	Peptidoglycan hydrolase flgJ	Q9I4P4.1	400	Pseudomonas aeruginosa	15 (1st hit)
			MLPMQK	Myeloid protein 1	P08940.2	326	Gallus gallus	36 (2nd hit)
			IG MLPM K	Arginine biosynthesis bifunctional protein argJ	Q71Z77.1	398	Listeria monocytogenes str. 4b F2365	49 (3rd hit)
			IG MLPM K	Arginine biosynthesis bifunctional protein argJ	Q8Y6U2.1	398	Listeria monocytogenes	49 (4th hit)
			G+MLPM+K	Replicase large subunit	P18339.2	1609	Tobacco mild green mosaic virus	49 (5th hit)
B1	565.3	+2	LVTDVQEAVR					
			LVTDVQEAVR	Proactivator polypeptide (Containing Saposin-A, B, C & D)	O13035.1	518	Gallus gallus	0.084 (1st hit)
			LVTDVQEA+R	Hypothetical protein	YP_001608779.1	525	Bartonella tribocorum CIP 105476	1.2 (2nd hit)
			LVT DVQEAV	Aliphatic sulfonates import ATP-binding protein ssuB	Q21XJ9.1	278	Rhodoferax ferrireducens	17 (3rd hit)
			LVT DVQEAV	Aliphatic sulfonates import ATP-binding protein ssuB	Q2SVN0.1	335	Burkholderia thailandensis E264	17 (4th hit)
			LVT DVQEAV	Aliphatic sulfonates import ATP-binding protein ssuB	Q39GW5.1	319	Burkholderia species 383	17 (5th hit)
			LVT DVQEAV	Aliphatic sulfonates import ATP-binding protein ssuB	Q0BFQ0.1	319	Burkholderia ambifaria AMMD	17 (6th hit)
			LVT DVQEAV	Aliphatic sulfonates	Q3JSR6.1	335	Burkholderia pseudomallei 1710b	17

Table A.1. cont'd.

				import ATP-binding protein ssuB				(7th hit)
			LVT DVQEAV	Aliphatic sulfonates import ATP-binding protein ssuB	Q62K56.2	335	Burkholderia mallei	17 (8th hit)
			LVT DVQEAV	Aliphatic sulfonates import ATP-binding protein ssuB	Q1BWL4.1	319	Burkholderia cenocepacia AU 1054	17 (9th hit)
			LVT DVQEAV	Aliphatic sulfonates import ATP-binding protein ssuB	Q63TW1.1	327	Burkholderia pseudomallei	17 (10th hit)
B2	646.9	+2	ISMPDFDLNLK					
			ISMPDFDLNLK	Neuroblast differentiation-associated protein AHNAK	Q09666.2	5890	Homo sapiens	0.001 (1st hit)
			ISMPDFD++	DNA polymerase III subunit alpha	Q1RKF9.1	1172	Rickettsia bellii RML369-C	5.1 (2nd hit)
			ISMPDFD++	DNA polymerase III subunit alpha	Q68VX1.1	1180	Rickettsia typhi	5.1 (3rd hit)
			ISMPDFD++	DNA polymerase III subunit alpha	Q92GB2.1	1181	Rickettsia conorii	5.1 (4th hit)
			ISMPDFD++	DNA polymerase III subunit alpha	Q4UK40.1	1207	Rickettsia felis	5.1 (5th hit)
			ISMPDFD++	DNA polymerase III subunit alpha	O05974.2	1182	Rickettsia prowazekii	5.1 (6th hit)
			ISMPDFD++	DNA polymerase III subunit alpha	O51526.2	1147	Borrelia burgdorferi	5.1 (7th hit)
			ISMPDFD++	DNA polymerase III subunit alpha	O83675.1	1170	Treponema pallidum	5.1 (8th hit)
			ISMPDFD++	DNA polymerase III subunit alpha	Q9RX08.1	1335	Deinococcus radiodurans	5.1 (9th hit)
			ISMPDFD++	DNA polymerase III subunit alpha	P57332.1	1161	Buchnera aphidicola	5.1 (10th hit)
			ISMPDFD	DNA polymerase III subunit alpha	Q9HXZ1.1	1173	Pseudomonas aeruginosa	6.8 (11th hit)
B2	772.0	+2	AVASAAAALVLK					
			AVASAAAALVLK	Talin-1	P26039.1	2541	Mus musculus	0.026 (1st hit)
			AVASAAAALVLK	Talin-1	P54939.2	2541	Gallus gallus	0.026 (2nd hit)

Table A.1. cont'd.

			AVASAAAALVLK	Talin-1	Q9Y490.3	2541	Homo sapiens	0.026 (3rd hit)
			AVA+AAA LVLK	Talin-2	Q9Y4G6.4	2542	Homo sapiens	3.8 (4th hit)
			AVA+AAA LVLK	Talin-2	Q71LX4.2	2375	Mus musculus	3.8 (5th hit)
			AVA AAALVLK	Phycocyanin-645 α-1 chain	P23815.1	70	Chroomonas species	6.8 (6th hit)
B2	525.8	+2	LGTFLENEK					
			LGTFLE+E	Glutamyl-tRNA reductase	Q2FTL0.1	424	Methanospirillum hungatei JF-1	36 (1st hit)
			LGTFLEN	Histone acetyltransferase p300	Q09472.2	2414	Homo sapiens	36 (2nd hit)
			L TFLEN K	Probable nucleoporin C890.06	Q9URX8.3	1315	Schizosaccharomyces pombe	87 (3rd hit)
			TFL+NEK	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	B0BBM4.1	354	Chlamydia trachomatis L2b/UCH-1/proctitis	117 (4th hit)
			TFL+NEK	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	B0B7F9.1	354	Chlamydia trachomatis	117 (5th hit)
			TFL+NEK	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	Q254I9.1	359	Chlamydophila felis	117 (6th hit)
			TFL+NEK	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	Q5L612.1	359	Chlamydophila abortus	117 (7th hit)
			TFL+NEK	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	Q823E0.1	359	Chlamydophila caviae	117 (8th hit)
			TFL+NEK	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	Q9PKF1.1	354	Chlamydia muridarum	117 (9th hit)
			TFL+NEK	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	O84245.1	354	Chlamydia trachomatis	117 (10th hit)

Table A.1. cont'd.

B4	821.9	+2	V(I/L)QQQADDAEER					
			V+QQQADDAEER	Putative tropomyosin α-3 chain-like protein	A6NL28.2	223	Homo sapiens	0.006 (1st hit)
			V+QQQADDAEER	Tropomyosin α-3 chain	Q63610.2	248	Rattus norvegicus	0.006 (2nd hit)
			QQQADDAE+R	Tropomyosin α-4 chain	P09495.3	248	Rattus norvegicus	0.20 (3rd hit)
			QQQADDAE+R	Tropomyosin α-4 chain	Q6IRU2.3	248	Mus musculus	0.20 (4th hit)
			QQQAD+AE+R	Tropomyosin α-4 chain	P67937.3	248	Sus scrofa	1.2 (5th hit)
			QQQAD+AE+R	Tropomyosin α-4 chain	P02561.2	248	Equus caballus	1.2 (6th hit)
B6	587.3	+2	IPPKPPARAAR					
			PP+PPARAA	Protein bassoon	Q9UPA5.4	3926	Homo sapiens	17 (1st hit)
			PPKPPAR	Frizzled-8	O93274.1	581	Xenopus laevis	30 (2nd hit)
			IPPKPP R	Sodium/hydrogen exchanger 2	Q9UBY0.1	812	Homo sapiens	30 (3rd hit)
			PPKPPAR	Structure-specific endonuclease subunit slx4	Q7SFJ3.1	1013	Neurospora crassa	30 (4th hit)
			PP+PPARA	Protein bassoon	O88737.3	3942	Mus musculus	97 (5th hit)
			PPKPPA A	DNA polymerase subunit γ-1	P54098.1	1239	Homo sapiens	97 (6th hit)
			PP+PPARA	Protein bassoon	O88778.3	3938	Rattus norvegicus	97 (7th hit)
			P PPAR+AR	Stress response protein NST1	Q5KHY3.1	1353	Filobasidiella neoformans	97 (8th hit)
B5	506.8	+2	IPSLPSGVDK					
			PSLPSGVD	Stromelysin-1	P08254.2	477	Homo sapiens	12 (1st hit)
			PSLPSG+D	Stromelysin-1	Q28397.1	477	Equus caballus	54 (2nd hit)
			IP LPSGV+	Amino-acid acetyltransferase	A1C6J2.1	718	Aspergillus clavatus	72 (3rd hit)
B6	583.8	+2	AAGEIIAIPRR					
			AAGEIIAIPRR	Envelope glycoprotein gp160	Q89607.1	852	HIV-2 B_EHO	0.011 (1st hit)
			AA IIAIPRR	Envelope glycoprotein gp160	P15831.2	859	Human immunodeficiency virus type 2	17 (2nd hit)

Table A.1. cont'd.

			EIIAIP+R	Porphobilinogen deaminase	A7GBW3.1	290	Clostridium botulinum F str. Langeland	22 (3rd hit)
			IIAIPRR	Envelope glycoprotein gp160	Q76638.1	857	HIV-2 B_UC1	22 (4th hit)
B1	831.4	+2	NSWGTSWGEDGYFR					
			NSWGTSWGEDGYFR	Dipeptidyl-peptidase 1 (Cathepsin C)	Q60HG6.1	463	Macaca fascicularis	5e-07 (1st hit)
			NSWGTSWGE+GYFR	Dipeptidyl-peptidase 1 (Cathepsin C)	Q3ZCJ8.1	463	Bos taurus	3e-06 (2nd hit)
			NSWGT WGEDGYFR	Dipeptidyl-peptidase 1 (Cathepsin C)	Q5RB02.1	463	Pongo abelii	5e-06 (3rd hit)
B5	684.9	+2	FRTTMLQDSIR					
			FRTTMLQDSI	Origin recognition complex subunit 2	Q75PQ8.1	576	Rattus norvegicus	0.014 (1st hit)
			FRTTMLQDSI	Origin recognition complex subunit 2	Q60862.1	586	Mus musculus	0.014 (2nd hit)
			FRTTMLQDSI	Origin recognition complex subunit 2	Q13416.2	577	Homo sapiens	0.014 (3rd hit)
			FRTTMLQD I	Origin recognition complex subunit 2	A6QNM3.2	577	Bos taurus	0.20 (4th hit)
			FRT++LQDS	Origin recognition complex subunit 2	Q91628.1	558	Xenopus laevis	22 (5th hit)
B6	586.8	+2	LPPEQGTTSSR					
			LPPEQGTTSSR	Cadherin EGF LAG seven-pass G-type receptor 2	XP_002686199.1	2920	Bos Taurus (flamingo homolog, Drosophila)	0.36 (1st hit)
			LPPEQGTTSSR	Cadherin EGF LAG seven-pass G-type receptor 2	NP_001179860.1	2920	Bos Taurus	0.36 (2nd hit)
			LPPE+GTTS	26S proteasome regulatory subunit 7, putative	XP_002778882.1	543	Perkinsus marinus ATCC 50983	173 (3rd hit)
			LPPE+GTT S	Hypothetical protein LELG_03374	XP_001525446.1	716	Lodderomyces elongisporus NRRL YB-4239	173 (4th hit)
			PPE+GTTS R	Thiamine ABC transporter, permease protein	YP_002827746.1	540	Rhizobium sp. NGR234	232 (5th hit)
B6	593.3	+2	AILYNYWDK					
			AILYNYWDK	Complement component c3	CAC69535.1	401	Crocodylus niloticus	0.50 (1st hit)
			ILYNYWD	Collagen-like cell surface-anchored protein SclD	YP_002745264.1	324	Streptococcus equi subspecies Zooepidemicus	24 (2nd hit)

Table A.1. cont'd.

			ILYNYWD	Collagen-like cell surface-anchored protein SclD	YP_002745683.1	313	Streptococcus equi subspecies equi 4047	24 (3rd hit)
			ILYNYWD	Collagen-like cell surface-anchored protein SclD	YP_002122604.1	336	Streptococcus equi subspecies zooepidemicus MGCS10565	24 (4th hit)
			ILYNYWD	Collagen-like protein D	ABA41494.1	258	Streptococcus equi subspecies Equi	24 (5th hit)
			A+LY YWDK	Hypothetical protein Btr_0321	YP_001608779.1	711	Bartonella tribocorum CIP 105476	41 (6th hit)
			AILYNYW	Complement component c3	NP_990736.1	1652	Gallus gallus	41 (7th hit)
B6	506.8	+2	NEALIALLR					
			NEALIALLR	Plastin-2	Q6P698.1	624	Danio rerio	0.79 (1st hit)
			NEALIALLR	Plastin-2	P13796.5	627	Homo sapiens	0.79 (2nd hit)
			NEALIALLR	Plastin-2	Q61233.4	627	Mus musculus	0.79 (3rd hit)
			NEALIALL	Plastin-1	A6H742.1	630	Bos taurus	8.3 (4th hit)
			NEALIALL	Plastin-1	Q14651.2	629	Homo sapiens	8.3 (5th hit)
			NEALIALL	Plastin-1	Q3V0K9.1	630	Mus musculus	8.3 (6th hit)
			NEALIALL	Plastin-1	P19179.1	630	Gallus gallus	8.3 (7th hit)
			NEALIAL LR	Metal-dependent hydrolase	Q6KZZ9.1	220	Picrophilus torridus	20 (8th hit)

Table A.2. Proteins identified at single peptide level using BLAST. *De novo* sequenced peptides obtained from gel digest were searched using BLAST with a limited taxonomy, containing birds (taxid:8782), crocodiles (taxid:8493), turtles (taxid:8459), tuataras (taxid:8508) and squamates (taxid:8509) using the SwissProt database. The top three E-values are reported.

Band Number	*m/z*	Charge	*De Novo* Sequence	Protein	Accession Number	Protein Length	Organism	E-Value
B4	1060.6	+2	EINLSPDSTSAVVSGLMVATK					
			EINLSPDSTSAVVSGLMVATK	Fibronectin	**P11722**	1256	Gallus gallus	2e-13 (1st hit)
			NLSP+S SA	Endoribonuclease Dicer	**Q25BN1**	1921	Gallus gallus	2.2 (2nd hit)
			NL P TS VSGL	Protein sidekick-1	**Q8AV58**	2169	Gallus gallus	4.0 (3rd hit)
B4	536.3	+2	YEVSVYALK					
			YEVSVYALK	Fibronectin	**P11722**	1256	Gallus gallus	0.003 (1st hit)
			Y+++VYA+K	Collagen α-1(XII)	**P13944**	3124	Gallus gallus	3.7 (2nd hit)
			Y VSVYA	Tyrosine-protein kinase BTK	**Q8JH64**	657	Gallus gallus	12 (3rd hit)
B4	701.3	+2	YQINQQWER					
			YQINQ ER	Solute carrier family 41 member 2	**Q5ZHX6**	573	Gallus gallus	2.1 (1st hit)
			INQQW	Thrombin-like enzyme	**Q072L7**	258	Lachesis stenophrys	2.8 (2nd hit)
			INQQW	Serine α-fibrinogenase	**Q8JH85**	258	Macrovipera lebetina	2.8 (3rd hit)
			INQQW	Serine β-fibrinogenase	**Q8JH62**	257	Macrovipera lebetina	2.8 (4th hit)
			QQWER	Zyxin	**Q04584**	542	Gallus gallus	2.8 (5th hit)
			YQ + QQWE	Structural maintenance of chromosomes protein 2	**Q90988**	1189	Gallus gallus	5.0 (6th hit)
B1	613.9	+2	LIALLEVLSQK					
			IA LLEVLS	Adenylate cyclase type 9	**Q9DGG6**	1334	Gallus gallus	2.3 (1st hit)
			LLEVL+Q	Exportin-7	**Q5ZLT0**	1087	Gallus gallus	4.1 (2nd hit)
			LI +LE +LSQK	E3 ubiquitin-protein ligase MIB2	**Q5ZIJ9**	954	Gallus gallus	5.5 (3rd hit)

Table A.2. cont'd.

			AL EV+SQK	Cell division cycle-associated protein 1	Q76I90	469	Gallus gallus	5.5 (4th hit)
B1	487.3	+2	VYGPGVEPR					
			VYGP VE	116 kDa U5 small nuclear ribonucleoprotein component	Q5F3X4	972	Gallus gallus	6.7 (1st hit)
			VYGPG	Homeobox protein Hox-C8	Q9YH13	242	Gallus gallus	29 (2nd hit)
			YGP VE	Protein sidekick-2	Q8AV57	2177	Gallus gallus	39 (3rd hit)
B2	736.4	+2	AQQVSQGLDLLTAK					
			AQQVSQGLDLLTAK	Vinculin	P12003	1135	Gallus gallus	5e-07 (1st hit)
			L+LLTAK	NADH-ubiquinone oxidoreductase chain 3	O47874	115	Alligator mississippiensis	5.4 (2nd hit)
			LDLLT K	Probable ATP-dependent RNA helicase DDX10	Q5ZJF6	875	Gallus gallus	9.8 (3rd hit)
B2	739.4	+2	MLGQMTDQVADLR					
			LGQMTDQ+ADLR	Vinculin	P12003	1135	Gallus gallus	6e-05 (1st hit)
			Q+TDQVA LR	COUP transcription factor 2	Q90733	410	Gallus gallus	2.3 (2nd hit)
			MLGQM	Importin-13	Q5ZIC8	958	Gallus gallus	5.5 (3rd hit)
B2	655.9	+2	TVTAMDVVYALK					
			TVTAMDVVYALK	Histone H4	P70081	103	Gallus gallus	5e-06 (1st hit)
			TVTAMDVVYALK	Histone H4	P62800	103	Cairina moschata	5e-06 (2nd hit)
			TVT +D VYALK	Rho guanine nucleotide exchange factor 6	Q5ZLR6	764	Gallus gallus	1.7 (3rd hit)
			T+T +DVVY	UPF0636 protein C4orf41 homolog	Q5ZI89	1132	Gallus gallus	5.5 (4th hit)
B2	628.7	+2	LITKAVSASK					
			LITKAVSASK	Histone H1.03	P08285	224	Gallus gallus	0.005 (1st hit)
			LITKAVSASK	Histone H1.10	P08286	220	Gallus gallus	0.005 (2nd hit)
			LITKAVSASK	Histone H1.11L	P08287	225	Gallus gallus	0.005 (3rd hit)

Table A.2. cont'd.

			LITKAVSASK	Histone H1.01	**P08284**	219	Gallus gallus	0.005 (4th hit)
			LITKAVSASK	Histone H1	**P09987**	218	Gallus gallus	0.005 (5th hit)
			LITKAVSASK	Histone H1.11R	**P08288**	219	Gallus gallus	0.005 (6th hit)
			LITKAV+ASK	Histone H1	**P09426**	218	Anas platyrhynchos	0.028 (7th hit)
			L+TKAV A	Neurochondrin	**Q5ZIG0**	702	Gallus gallus	18 (8th hit)
B6	833.9	+2	FSGSGSGTDFTFTISR					
			FSG GSG +FT	Large fibroblast proteoglycan	**Q90953**	3562	Gallus gallus	1.7 (1st hit)
			S SG GT FTF+ R	Docking protein 3	**A3R064**	426	Gallus gallus	5.4 (2nd hit)
			SG+GSG+ TF	Protein PRRC1	**Q5F3I0**	442	Gallus gallus	13 (3rd hit)
B6	751.4	+2	VFGGGTKLTVLGQPK					
			TVLGQP	Transmembrane protein 184C	**Q5ZMP3**	445	Gallus gallus	5.4 (1st hit)
			GT LTVL Q K	Integrator complex subunit 2	**Q5ZKU4**	1192	Gallus gallus	13 (2nd hit)
			VFG +KLT	Protein HIRA	**P79987**	1019	Gallus gallus	18 (3rd hit)
B2	510.8	+2	VLASFGEAVK					
			VLASFGEAVK	Hemoglobin subunit β	**P02131**	146	Caiman crocodilus	0.003 (1st hit)
			VLASFGEAVK	Hemoglobin subunit β	**P02130**	146	Alligator mississippiensis	0.003 (2nd hit)
			VL+SFGEAVK	Hemoglobin subunit epsilon	**P02128**	147	Gallus gallus	0.015 (3rd hit)
			VL+SFGEAVK	Hemoglobin ρ chain	**P02127**	147	Gallus gallus	0.015 (4th hit)
			VLASFGEAV	Hemoglobin β chain	**P02129**	146	Crocodylus niloticus	0.021 (5th hit)
			VL SFGEAVK	Hemoglobin β chain	**P13274**	146	Chrysemys picta bellii	0.021 (6th hit)
			VL SFGEAVK	Hemoglobin β chain	**P83133**	146	Dipsochelys dussumieri	0.021 (7th hit)
			VL SFGEAVK	Hemoglobin β chain	**P02125**	146	Ciconia ciconia	0.021 (8th hit)

Table A.2. cont'd.

			VL SFGEAVK	Hemoglobin β chain	**P14524**	146	Turdus merula	0.021 (9th hit)
			VL SFGEAVK	Hemoglobin β chain	**P08851**	146	Accipiter gentilis	0.021 (10th hit)
			VL SFGEAVK	Hemoglobin β chain	**P02116**	146	Ara ararauna	0.021 (11th hit)
			VL SFGEAVK	Hemoglobin β chain	**P07411**	146	Vultur gryphus	0.021 (12th hit)
			VL SFGEAVK	Hemoglobin β chain	**P15165**	146	Apus apus	0.021 (13th hit)
			VL SFGEAVK	Hemoglobin β/β' chain	**P08261**	146	Larus ridibundus	0.021 (14th hit)
			VL SFGEAVK	Hemoglobin β chain	**P21668**	146	Psittacula krameri	0.021 (15th hit)
			VL SFGEAVK	Hemoglobin β chain	**P02120**	146	Anseranas semipalmata	0.021 (16th hit)
			VL SFGEAVK	Hemoglobin β chain	**P10782**	146	Phalacrocorax carbo	0.021 (17th hit)
			VL SFGEAVK	Hemoglobin β chain	**P07406**	146	Passer montanus	0.021 (18th hit)
			VL SFGEAVK	Hemoglobin β chain	**P68061**	146	Aegypius monachus	0.021 (19th hit)
			VL SFGEAVK	Hemoglobin β chain	**P02122**	146	Aquila chrysaetos	0.021 (20th hit)
			VL SFGEAVK	Hemoglobin β chain	**P02121**	146	Phoenicopters ruber ruber	0.021 (21st hit)
			VL SFGEAVK	Hemoglobin β chain	**Q98905**	147	Geochelone carbonaria	0.021 (22nd hit)
			VL SFGEAVK	Hemoglobin β chain	**P82113**	146	Stercorarius maccormicki	0.021 (23rd hit)
B2	500.8	+2	LSSPISGDPK					
			LSSP ISG+P	Hemoglobin subunit β	**P07036**	146	Chloephaga melanoptera	2.3 (1st hit)
			PISGDP	Myosin-binding protein C	**Q90688**	1272	Gallus gallus	4.1 (2nd hit)
			LSSP I G+PK	Hemoglobin subunit ε	**P14261**	147	Cairina moschata	13 (3rd hit)
			LSSP I G+PK	Hemoglobin ρ chain	**P02127**	147	Gallus gallus	13

Table A.2. cont'd.

								(4th hit)
			LSSP I+G+P	Hemoglobin β/β' chain	P08261	146	Larus ridibundus	13 (5th hit)
			LSSP I+G+P	Hemoglobin β-1 chain	P18993	146	Varanus exanthematicus albigularis	13 (6th hit)
B6	522.8	+2	IMSIVDPNR					
			IMSIVDPNR	α-actinin-1	P05094	893	Gallus gallus	0.001 (1st hit)
			IMS+VDPN	α-actinin-2	P20111	897	Gallus gallus	0.15 (2nd hit)
			IMS+VDPN	α-actinin-4	Q90734	904	Gallus gallus	0.15 (3rd hit)
			SIVDP NR	PRELI domain-containing protein 1, mitochondrial	Q90673	215	Gallus gallus	5.0 (4th hit)
B5	506.2	+3	QEYDESGPSIVHR					
			QEYDESGPSIVHR	γ-actin	Q5ZMQ2	375	Gallus gallus	3e-07 (1st hit)
			QEYDESGPSIVHR	γ-actin	P63256	375	Anser anser	3e-07 (2nd hit)
			QEYDESGPSIVHR	β-actin	P60706	375	Gallus gallus	3e-07 (3rd hit)
			QEYDESGPSIVHR	Actin, cytoplasmic type 5	P53478	376	Gallus gallus	3e-07 (4th hit)
			QEYDE+GPSIVHR	Actin, alpha cardiac muscle 1	P68034	377	Gallus gallus	2e-06 (5th hit)
			QEYDE+GPSIVHR	α-actin	P08023	377	Gallus gallus	2e-06 (6th hit)
			QEYDE+GPSIVHR	α-actin-1	P68139	377	Gallus gallus	2e-06 (7th hit)
			QEYDE+GPSIVHR	α-actin-1	Q90X97	377	Atractaspis microlepidota	2e-06 (8th hit)
			EYDE+GPSIVHR	Γ-2-actin	P63270	376	Gallus gallus	2e-05 (9th hit)
B1	683.9	+2	STDYGILQINSR					
			STDYGILQINSR	Lysozyme C	P84492	130	Chelonia mydas	5e-06 (1st hit)
			STDYGILQINSR	Lysozyme C	Q7LZI3	129	Tragopan satyra	5e-06 (2nd hit)

Table A.2. cont'd.

STDYGILQINSR	Lysozyme C	**Q7LZQ1**	131	Pelodiscus sinensis	5e-06 (3rd hit)
STDYGILQINSR	Lysozyme C	**P24364**	129	Lophura leucomelanos	5e-06 (4th hit)
STDYGILQINSR	Lysozyme C	**P81711**	129	Syrmaticus soemmerringii	5e-06 (5th hit)
STDYGILQINSR	Lysozyme C	**P00702**	147	Phasianus colchicus colchicus	5e-06 (6th hit)
STDYGILQINSR	Lysozyme C	**P00703**	147	Meleagris gallopavo	5e-06 (7th hit)
STDYGILQINSR	Lysozyme C	**Q7LZP9**	129	Lophophorus impejanus	5e-06 (8th hit)
STDYGILQINSR	Lysozyme C	**P49663**	130	Phasianus versicolor	5e-06 (9th hit)
STDYGILQINSR	Lysozyme C	**P00698**	147	Gallus gallus	5e-06 (10th hit)
STDYGILQINSR	Lysozyme C-1	**P00705**	147	Anas platyrhynchos	5e-06 (11th hit)
STDYGILQINSR	Lysozyme C	**P19849**	129	Pavo cristatus	5e-06 (12th hit)
STDYGILQINSR	Lysozyme C	**P00701**	147	Coturnix japonica	5e-06 (13th hit)
STDYGILQINSR	Lysozyme C	**Q7LZQ3**	129	Crax fasciolata	5e-06 (14th hit)
STDYGILQINSR	Lysozyme C	**P24533**	129	Syrmaticus reevesii	5e-06 (15th hit)
STDYGILQINSR	Lysozyme C	**Q7LZT2**	129	Tragopan temminckii	5e-06 (16th hit)
STDYGILQINSR	Lysozyme C	**P00707**	129	Ortalis vetula	5e-06 (17th hit)
STDYGILQINSR	Lysozyme C	**P22910**	129	Chrysolophus amherstiae	5e-06 (18th hit)
STDYG+LQINSR	Lysozyme C	**P00704**	129	Numida meleagris	3e-05 (19th hit)

Table A.2. cont'd.

			STDYG+LQINSR	Lysozyme C	**P00699**	129	Callipepla californica	3e-05 (20th hit)
			STDYG+LQINSR	Lysozyme C	**P00700**	129	Colinus virginianus	3e-05 (21st hit)
			STDYGIL+INSR	Lysozyme C-3	**P00706**	129	Anas platyrhynchos	4e-05 (22nd hit)
			STDYGIL+INSR	Lysozyme C	**Q7LZQ2**	129	Aix sponsa	4e-05 (23rd hit)
B1	503.3	+2	WDAWDALK					
			WDAW+ALK	Acyl-coenzyme A binding protein	**Q9PRL8**	86	Gallus gallus	0.017 (1st hit)
			WDAW AL	Acyl-CoA-binding domain-containing protein 5	**Q5ZHQ6**	492	Gallus gallus	0.76 (2nd hit)
			W+ WDA	β-crystallin B3	**P55165**	211	Gallus gallus	11 (3rd hit)
			WDAW	β-crystallin A3	**P10042**	215	Gallus gallus	11 (4th hit)
			W+AW+ LK	Acyl-CoA-binding protein	**P45882**	103	Anas platyrhynchos	11 (5th hit)
B4	509.8	+2	IGTMLPMQK					
			MLPMQK	Myeloid protein-1	**P08940**	326	Gallus gallus	0.35 (1st hit)
			TMLPM	Brachyury protein	**P79777**	433	Gallus gallus	3.7 (2nd hit)
			ML MQK	Eukaryotic translation initiation factor 3 subunit	**Q5F428**	564	Gallus gallus	12 (3rd hit)
			TM LPMQ	Paired box protein Pax-6	**P47237**	216	Gallus gallus	12 (4th hit)
			TM LPMQ	Paired box protein Pax-6	**P47238**	416	Coturnix japonica	12 (5th hit)
B1	565.3	+2	LVTDVQEAVR					
			LVTDVQEAVR	Saposin-A	**O13035**	518	Gallus gallus	8e-04 (1st hit)
			VTDV EA	Cadherin-4	**P24503**	913	Gallus gallus	13 (2nd hit)
			LVT+VQ EA	Prolactin-2	**P55752**	199	Gallus gallus	13 (3rd hit)

Table A.2. cont'd.

			LVT+VQ EA	Prolactin-1	**P55751**	199	Alligator mississippiensis	13 (4th hit)
			VTD VQEA	Cystatin	**P08935**	111	Bitis arietans	13 (5th hit)
			LVT+VQ EA	Prolactin-2	**P55754**	199	Crocodylus novaeguineae	13 (6th hit)
			LVT+VQ EA	Prolactin-1	**P55753**	199	Crocodylus novaeguineae	13 (7th hit)
			VTDVQ+	Fibronectin	**P11722**	1256	Gallus gallus	18 (8th hit)
B2	646.9	+2	ISMPDFDLNLK					
			SMPDF	Myosin-Ig	**Q5ZMC2**	1007	Gallus gallus	7.4 (1st hit)
			DFDL+L	Zinc finger FYVE domain-containing protein 27	**Q5ZL36**	406	Gallus gallus	18 (2nd hit)
			M D+DL LK	Uncharacterized protein C17orf85 homolog	**Q5ZM19**	604	Gallus gallus	24 (3rd hit)
			IS+PDF	Collagen α-3(VI) chain	**P15989**	3137	Gallus gallus	24 (4th hit)
B2	772.0	+2	AVASAAAALVLK					
			AVASAAAALVLK	Talin-1	**P54939**	2541	Gallus gallus	2e-04 (1st hit)
			VASAAAA	Insulin-like growth factor 2 mRNA-binding protein 3	**Q5ZLP8**	584	Gallus gallus	5.5 (2nd hit)
			V +AA+ALVL	NADH-ubiquinone oxidoreductase chain 2	**O21398**	346	Struthio camelus	7.3 (3rd hit)
B2	525.8	+2	LGTFLENEK					
			LGTFL N EK	α-type phospholipase A2 inhibitor anMIP	**A1XRN2**	166	Atropoides nummifer	2.1 (1st hit)
			LGTFL N	α-phospholipase A2 inhibitor clone 09	**B1A4R4**	166	Lachesis muta muta	6.7 (2nd hit)
			LGTFL N	α-phospholipase A2 inhibitor	**B1A4R0**	166	Lachesis muta muta	6.7 (3rd hit)
			LGTFL N	Phospholipase A2 myotoxin inhibitor protein	**Q8AYA2**	166	Bothrops moojeni	6.7 (4th hit)
			LGTF LEN	Cannabinoid receptor 1	**P56971**	473	Taeniopygia guttata	8.9 (5th hit)

Table A.2. cont'd.

B4	821.9	+2	VIQQQADDAEER					
			VIQ QQ+D+ EE	Tenascin	**P10039**	1808	Gallus gallus	0.29 (1st hit)
			I+Q+ADD EE	Transcription factor SOX-11	**P48435**	396	Gallus gallus	0.70 (2nd hit)
			QQ DDAEER	Myosin heavy chain	**P29616**	1102	Gallus gallus	0.70 (3rd hit)
			+Q +AD DAEER	Myosin-3	**P02565**	1940	Gallus gallus	1.7 (4th hit)
			+Q +AD DAEER	Myosin heavy chain	**P13538**	1939	Gallus gallus	1.7 (5th hit)
B6	587.3	+2	IPPKPPARAAR					
			+PPKPPA	G2/mitotic-specific cyclin-B2	**P29332**	399	Gallus gallus	1.7 (1st hit)
			PPKP P RAA	Venom protease inhibitor 2	**B7S4N9**	88	Oxyuranus scutellatus	5.5 (2nd hit)
			PP+PPAR	Eukaryotic translation initiation factor 3 subunit H	**Q5ZLE6**	348	Gallus gallus	5.5 (3rd hit)
			PPKP A+AA	Transcription factor SOX-11	**P48435**	396	Gallus gallus	5.5 (4th hit)
			PP PPAR AR	Histone H3-like centromeric protein A	**Q6XXM1**	131	Gallus gallus	7.4 (5th hit)
			PARAAR	SCO-spondin	**Q2PC93**	5255	Gallus gallus	7.4 (6th hit)
B5	506.8	+2	IPSLPSGVDK					
			IP LPSGV	Coatomer subunit Δ	**Q5ZL57**	510	Gallus gallus	13 (1st hit)
			LPS VDK	116 kDa U5 small nuclear ribonucleoprotein component	**Q5F3X4**	972	Gallus gallus	18 (2nd hit)
			LPSG+D	Neural cell adhesion molecule 1	**P13590**	1091	Gallus gallus	32 (3rd hit)
			IP+LP+G	Hepatocyte nuclear factor 1-α	**Q90867**	634	Gallus gallus	32 (4th hit)
B6	583.8	+2	AAGEIIAIPRR					
			AGEI+ IP	Jumonji domain-containing protein	**Q5ZHV5**	425	Gallus gallus	24 (1st hit)

Table A.2. cont'd.

IAIPR	4 Lysine-specific demethylase 3A	**Q5ZIX8**	1325	Gallus gallus	24 (2nd hit)
IIAIP	Cytochrome c oxidase polypeptide I	**079548**	533	Gallus gallus	24 (3rd hit)
IIAIP	Cytochrome c oxidase subunit 1	**Q94WR7**	516	Buteo buteo	24 (4th hit)
IIAIP	Cytochrome c oxidase subunit 1	**O21399**	516	Struthio camelus	24 (5th hit)
IIAIP	Cytochrome c oxidase subunit 1	**O03546**	337	Rhea americana	24 (6th hit)
IIAIP	Cytochrome c oxidase subunit 1	**O03539**	337	Nothoprocta perdicaria	24 (7th hit)
IIAIP	Cytochrome c oxidase subunit 1	**O03521**	337	Casuarius bennetti	24 (8th hit)
IIAIP	Cytochrome c oxidase subunit 1	**O03515**	337	Apteryx australis	24 (9th hit)
IIAIP	Cytochrome c oxidase subunit 1	**P18943**	515	Gallus gallus	24 (10th hit)
IIAIP	Cytochrome c oxidase subunit 1	**O03524**	337	Dromaius novaehollandiae	24 (11th hit)
IIAIP	Cytochrome c oxidase subunit 1	**P24984**	516	Coturnix japonica	24 (12th hit)
IIAIP	Cytochrome c oxidase subunit 1	**O03554**	337	Tinamus major	24 (13th hit)
IIAIP	Cytochrome c oxidase subunit 1	**O79672**	514	Pelomedusa subrufa	24 (14th hit)
AGEII	MBD1-containing chromatin-associated factor 1	**Q5ZIE8**	1085	Gallus gallus	43 (15th hit)
I IPRR	Tenascin	**P10039**	1808	Gallus gallus	43 (16th hit)
EIIA RR	Protein sidekick-1	**Q8AV58**	472	Gallus gallus	58 (17th hit)
AAG IIAI	Alcohol dehydrogenase 1	**P49645**	375	Gallus gallus	58

Table A.2. cont'd.

								(18th hit)
			AAG IIAI	Alcohol dehydrogenase 1	**P19631**	375	Coturnix japonica	58 (19th hit)
			GEI IPR	Lysozyme C	**Q91159**	145	Opisthocomus hoazin	58 (20th hit)
			GEI IPR	DNA damage-binding protein 1	**Q805F9**	1140	Gallus gallus	58 (21st hit)
B1	831.4	+2	NSWGTSWGEDGYFR					
			NSWGT WG GY	Cathepsin K	**Q90686**	334	Gallus gallus	0.008 (1st hit)
			NSWG WG+ GY	Cathepsin L1	**P09648**	218	Gallus gallus	0.12 (2nd hit)
			NSW T WG G+F	Cathepsin B	**P43233**	340	Gallus gallus	0.29 (3rd hit)
B5	684.9	+2	FRTTMLQDSIR					
			TML+DS+	Talin-1	**P54939**	2541	Gallus gallus	7.3 (1st hit)
			MLQD+I	Hsp90 co-chaperone Cdc37	**O57476**	393	Gallus gallus	7.3 (2nd hit)
			FR+TM L+DS	78 kDa glucose-regulated protein	**Q90593**	652	Gallus gallus	7.3 (3rd hit)
			MLQDS	5,6-dihydroxyindole-2-carboxylic acid oxidase	**O57405**	535	Gallus gallus	13 (4th hit)
			LQ+SIR	Coiled-coil domain-containing protein 132	**Q5ZKV9**	949	Gallus gallus	18 (5th hit)
			TTML+ +I	Filensin	**Q06637**	657	Gallus gallus	18 (6th hit)
B6	586.8	+2	LPPEQGTTSSR					
			LPPEQG	Basic protease inhibitor	**P00993**	110	Caretta caretta	3.0 (1st hit)
			LPPE G+ SS	Muscle-specific regulatory factor 4	**Q01795**	242	Gallus gallus	13 (2nd hit)
			LPPEQ	Abhydrolase domain-containing protein FAM108C1	**Q5ZJX1**	310	Gallus gallus	18 (3rd hit)

Table A.2. cont'd.

			LPPEQ	α-actinin-2	**P20111**	897	Gallus gallus	18 (4th hit)
			PEQG TS	Dihydropyrimidinase-related protein 2	**Q90635**	572	Gallus gallus	18 5th hit)
			LPPEQ	α-actinin-4	**Q90734**	904	Gallus gallus	18 6th hit)
B6	593.3	+2	AILYNYWDK					
			AILYNY	Complement C3	**Q01833**	1651	Naja naja	0.85 (1st hit)
			LYNYW	Ectonucleoside triphosphate diphosphohydrolase	**P79784**	495	Gallus gallus	1.1 (2nd hit)
			LY+YW	Fatty acyl-CoA reductase 1	**Q5ZM72**	515	Gallus gallus	12 (3rd hit)
			LY NYW+	Monocarboxylate transporter 9	**Q5ZJU0**	507	Gallus gallus	12 (4th hit)
B6	506.8	+2	NEALIALLR					
			NEALIALL	Plastin-1	**P19179**	630	Gallus gallus	0.081 (1st hit)
			NEA+IA LL	tRNA (cytosine-)-methyltransferase NSUN2	**Q5ZLV4**	796	Gallus gallus	21 (2nd hit)
			NE+LI L R	Ubiquitin carboxyl-terminal hydrolase 28	**Q5ZID5**	1047	Gallus gallus	21 (3rd hit)
			EALI+L	CCR4-NOT transcription complex subunit 10	**Q5ZIW2**	744	Gallus gallus	39 (4th hit)

APPENDIX B. LECTIN SEQUENCE

Protein sequencing with MS/MS and *de novo* sequencing yielded almost full coverage of an alligator lectin. The sequence was compared to lectins from other vertebrates using DIALIGN software to investigate sequence homology (Chapter 6, Section 6.4.4).

Figure 1. *Alligator mississippiensis* lectin sequence aligned with *Homo sapiens* and *Mus musculus* intelectin-2.

```
Human           MLSMLRTMTRLCFLLFFSVATSGCSAAAASSLEMLSREFETCAFSFSSLPRSCKEIKERC 60
Mouse           -------MTQLGFLLFIMIATRVCSAAEEN----LDTNRWGNSF-FSSLPRSCKEIKQED 48
Alligator       -------NNQLALKLAATGGSTNXLPALALQN--LLNTWEDTSCCSQTSPGQQSWPRD-- 49
                       .:* : *    .:    .*       *       :   .: * . .  ::

Human           HSAGDGLYFLRTKNGVVYQTFCDMTSGGGGWTLVASVHENDMRGKCTVGDRWSSQQGNKA 120
Mouse           TKAQDGLYFLRTENGVIYQTFCDMTTAGGGWTLVASVHENNLRGRCTVGDRWSSQQGNRA 108
Alligator       -GAQDGLYTLSTADGEIYQTFCDMSTHGGGWTLVASVHENNAHGKCTVGDRWSSQQGNSP 108
                  * **** * * :* :*******:: ************: :*:************* .

Human           DYPEGDGNWANYNTFGSAEAATSDDYKNPGYYDIQAKDLGIWHVPNKSPM--QHWRNSAL 178
Mouse           DYPEGDGNWANYNTFGSAEGATSDDYKNPGYFDIQAENLGIWHVPNNSPL--HTWRNSSL 166
Alligator       LYPEGDGNWANNNIFGSAMGSTSDDYKNPGYYDLQAGDLSVWHVPDRAPLRKEMIESSVL 168
                 ********** * **** .:**********:*:** :*.:****:.:*:  .  ..* *

Human           LRYRTNTGFLQRLGHNLFGIYQKYPVKYRSGKCWNDNGPAIPVVYDFGDAKKTASYYSPY 238
Mouse           LRYRTFTGFLQRLGHNLFGLYQKYPVKYGEGKCWTDNGPAFPVVYDFGDAQKTASYYSPS 226
Alligator       LFYR--TGFLSSEGGNLLRLYEKYPVKYGAGSCKVDNGPAVPIVYDFGSAEKTAAYYSPS 226
                * **  ****.  * **: :*:******  *.*  *****.*:*****.*:***:****

Human           GQREFVAGFVQFRVFNNERAANALCAGIKVTGCNTEHHCIGGGGFFPQGKPRQCGDFSAF 298
Mouse           GRNEFTAGYVQFRVFNNERAASALCAGVRVTGCNTEHHCIGGGGFFPEFDPEECGDFAAF 286
Alligator       GRGEFTAGFVQFRVFNNEKAPMALCSGLKVTGCNTEHHCIGGGGFFPEGNPRQCGDFPAF 286
                *: **.**:*********:*. ***:*::******************: .*.:****.**

Human           DWDGYGTHVKSSCSREITEAAVLLFYR 325
Mouse           DANGYGTHIRYSNSREITEAAVLLFYR 313
Alligator       DWDGYGTHQSWSTSREMIESSVLLFYR 313
                * :*****   * ***: *::******
```

Figure 2. *Alligator mississippiensis* lectin sequence aligned with *Homo sapiens* ficolin-α and ficolin-β.

```
Ficolin-1       -MELSGATMARGLAVLLVLFLHIKNLPAQAADT-CPEVKVVGLEGSDKLTILRGCPGLPG 58
Alligator       NNQLALKLAATGGSTNSLPALALQNLLNTWEDTSCCSQTSPGQQSWPRDGAQDGLYTLST 60
                  :*:    * * :.  :  * ::**     ** * . .  * :.  :     *   *.

Ficolin-1       APGPKGEAGVIGERGERGLPGAPGKAGPVGPKGDRGEKGMRGEKGDAGQSQSCATGPRNC 118
Alligator       ADG-----EIYQTFCDMSTHGGGWTLVASVHENNAHGKCTVGDRWSSQQG-NSPLYPEGD 114
                * *      :     : .  *.  .  .   :.:   *   *:: .: *. ...  *..

Ficolin-1       KDLLDRGYFLSGWHTIYLPDCRPLTVLCDMDTDGGGWTVFQRRMDGSVDFYRDWAAYKQG 178
Alligator       GNWANNNIFGSAMG----STSDDYKNPGYYDLQAGDLSVWH--VPDRAPLRKEMIESSVL 168
                 :  :.. * *.      . .   .     * :.*. :*::  : . . : ::    .

Ficolin-1       FGSQLGEFWLGNDNIHALTAQGSSELRVDLVDFEGNHQFAKYKSFKVADEAEKYKLVLG- 237
Alligator       LFYRTGFLSSEGGNLLRLYEKYPVKYGAGSCKVDNGPAVPIVYDFGSAEKTAAYYSPSGR 228
                :  : * :   ..*:  *  : . :  ..  ..:..  ..   .*  *:::  *    *

Ficolin-1       -AFVGGSAGNSLTGHNNNFFSTKDQDNDVSSSNCAEKFQGAWWYADCHASNLNGLYLMGP 296
Alligator       GEFTAGFVQFRVFNNEKAPMALCSG-LKVTGCNTEHHCIGGGGFFPEGNPRQCGDFPAFD 287
                  *..* .   : .:::  ::  .   .*:..*  .:  *.  :     ..  * :

Ficolin-1       HESYANGINWSAAKGYKYSYKVSEMKVRPA 326
Alligator       WDGYGTHQSWSTSREMIESSVLLFYR---- 313
                 :.*..  .**:::    *  :   :


Ficolin-2       -MELDRAVGVLGAATLLLSFLGMAWALQAAD--TCPEVKMVGLEGSDKLTILRGCPGLPG 57
Alligator       NNQLALKLAATGGSTNSLPALALQNLLNTWEDTSCCSQTSPGQQSWPRDGAQDGLYTLST 60
                  :*   :.. *.:*  *. *.:   *:: :  :* . .  * :.  :     *   *.

Ficolin-2       APGPKGE--AGTNGKRGERGPPGPPGKAGPPGPNGAPGEPQPCLTGPRTCKDLLDRGHFL 115
Alligator       ADGEIYQTFCDMSTHGGGWTLVASVHENNAHG-KCTVGDRWSSQQGNSPLYPEGD-GNWA 118
                * *   :  .. . : *     ..  : .. * : : *:  ..  *  .     * *::

Ficolin-2       SG--WHTIYLPDCRPLTVLCDMDTDGGGWTVFQRRVDGSVDFYRDWATYKQGFGSRLGEF 173
Alligator       NNNIFGSAMGSTSDDYKNPGYYDLQAGDLSVWH--VPDRAPLRKEMIESSVLLFYRTGFL 176
                ..  : :   . .   .     * :.*. :*::  * . . : ::    .  :  * * :

Ficolin-2       WLGNDNIHALTAQGTSELRVDLVDFEDNYQFAKYRSFKVADEAEKYNLVLG--AFVEGSA 231
Alligator       SSEGGNLLRLYEKYPVKYGAGSCKVDNGPAVPIVYDFGSAEKTAAYYSPSGRGEFTAG-F 235
                   ..*:  *  : . :  ..  ..::.  ..   .*  *:::  *    *   *. *

Ficolin-2       GDSLTFHNNQSFSTKDQDNDLNTGNCAVMFQGAWWYKNCHVSNLNGRYLRGTHGSFANGI 291
Alligator       VQFRVFNNEKAPMALCSGLKVTGCNTEHHCIGGGGFFPEGNPRQCGDFPAFDWDGYGTHQ 295
                 :  .*:*:::  :  .. .:.  *     *.  :     ..  * :     ..:..

Ficolin-2       NWKSGKGYNYSYKVSEMKVRPA 313
Alligator       SWSTSREMIESSVLLFYR---- 313
                .*.:.:    *  :   :
```

APPENDIX C. LETTERS OF PERMISSION

ELSEVIER LICENSE
TERMS AND CONDITIONS

Jul 13, 2011

This is a License Agreement between Lancia N.F Darville ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Lancia N.F Darville
Customer address	232 Choppin Hall Baton Rouge, LA 70803
License number	2707130503266
License date	Jul 13, 2011
Licensed content publisher	Elsevier
Licensed content publication	Comparative Biochemistry and Physiology Part D: Genomics and Proteomics
Licensed content title	Proteome analysis of the leukocytes from the American alligator (*Alligator mississippiensis*) using mass spectrometry
Licensed content author	Lancia N.F. Darville, Mark E. Merchant, Azeem Hasan, Kermit K. Murray
Licensed content date	December 2010
Licensed content volume number	5
Licensed content issue number	4
Number of pages	9
Start Page	308
End Page	316
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	Proteomic Analysis of the Blood of Alligator mississippiensis

Expected completion date	Dec 2011
Estimated size (number of pages)	195
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
Total	0.00 USD
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that

you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all

languages.

16. **Website**: The following terms and conditions apply to electronic reserve and author websites:
Electronic reserve: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:
This license was made in connection with a course,
This permission is granted for 1 year only. You may obtain a license for future website posting,
All content posted to the web site must maintain the copyright information line on the bottom of each image,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com , and
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

17. **Author website** for journals with the following additional clauses:

All content posted to the web site must maintain the copyright information line on the bottom of each image, and
he permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx , As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

18. **Author website** for books with the following additional clauses:
Authors are permitted to place a brief summary of their work online only.
A hyper-text must be included to the Elsevier homepage at http://www.elsevier.com

All content posted to the web site must maintain the copyright information line on the bottom of each image. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx. or for books to the Elsevier homepage at http://www.elsevier.com

20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published

commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

21. **Other Conditions**:

v1.6

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK11019482. Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact Rightslink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Aug 02, 2011

This is a License Agreement between Lancia N.F Darville ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Lancia N.F Darville
Customer address	232 Choppin Hall Baton Rouge, LA 70803
License number	2720870182331
License date	Aug 02, 2011
Licensed content publisher	Elsevier
Licensed content publication	Microchemical Journal
Licensed content title	A mass spectrometry approach for the study of deglycosylated proteins
Licensed content author	Lancia N.F. Darville, Mark E. Merchant, Kermit K. Murray
Licensed content date	1 June 2011
Licensed content volume number	n/a
Licensed content issue number	n/a
Number of pages	1
Start Page	
End Page	
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	Proteomic Analysis of the Blood of Alligator mississippiensis
Expected completion date	Dec 2011
Estimated size (number of pages)	195

Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / GBP
Total	0.00 USD
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a

timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Website**: The following terms and conditions apply to electronic reserve and author websites:

Electronic reserve: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:
This license was made in connection with a course,
This permission is granted for 1 year only. You may obtain a license for future website posting,
All content posted to the web site must maintain the copyright information line on the bottom of each image,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com , and
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

17. **Author website** for journals with the following additional clauses:

All content posted to the web site must maintain the copyright information line on the bottom of each image, and
he permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx , As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

18. **Author website** for books with the following additional clauses:
Authors are permitted to place a brief summary of their work online only.
A hyper-text must be included to the Elsevier homepage at http://www.elsevier.com

All content posted to the web site must maintain the copyright information line on the bottom of each image.You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx. or for books to the Elsevier homepage at http://www.elsevier.com

20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should

your thesis be published commercially, please reapply for permission.

21. **Other Conditions**:

v1.6

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK11030596. Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact Rightslink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

APPENDIX D. ANIMAL RESEARCH PLAN

MCNEESE STATE UNIVERSITY
INSTITUTIONAL ANIMAL CARE AND USE COMMITTEE
ANIMAL USE PROTOCOL

All investigators and instructors using species currently covered by federal statute and guidelines must complete this form and submit it to the Director of Research Services and Sponsored Programs (Campus Box 90655) for review by the IACUC a minimum of two weeks prior to initiating the protocol. The investigator will be informed by the Director of the IACUC of the committee's action and recommendations. IACUC evaluation of a proposal may require visiting the research site or animal facility to be used.

Date Received by RSSP ——

Date Received by IACUC 09 Oct 09

Date of IACUC review 20 Oct 09

IACUC Action (X) Accept ( ) Recommend changes ( ) Reject

( ) Site Visit _____ (date)

Additional Comments Attached ( )

IACUC DIRECTOR [signature] (date) 22 Oct 09

---

1. Project Title (or laboratory exercise title): Collection of blood from Crocodilians

2. Investigator(s)/Instructor(s): Mark Merchant

Department: Chemistry Phone: _____

3. Time period of protocol: From Oct 2009 to: Oct 2012

4. Attach a description of the experimental protocol and any supporting material to this form.

5. Name of funding source for this proposal: LEQSF

6. Description of Animals:

Species and Common Name (if any) and number of animals:

American alligator (Alligator mississippiensis)
Broad-snouted caiman (Caiman latirostris)
Yacare caiman (Caiman yacare)
Orinoco crocodile (Crocodylus intermedius)
American crocodile (Crocodylus acutus)
Spectacled Caimain (Caiman crocodylus)
Nile crocodile (Crocodylus niloticus)
Saltwater crocodile (Crocodylus porosus)

March 2006

(more on back)

Freshwater crocodile (Crocodylus johnstonii)
Dwarf Crocodile (Osteolaemos tetra spis)
Cuvier's dwarf caiman (Paleosuchus palpebrosus)
Scheidel's dwarf caiman (Paleosuchus trigonatus)
Black caiman (Melanosuchus niger)
Slender-snouted crocodile (Crocodylus cataphractus)
Phillipine crocodile (Crocodylus mindorensis)
Morelet's crocodile (Crocodylus moreletii)
Siamese crocodile (Crocodylus siamensis)
New Guinea crocodile (Crocodylus novaguineae)
Cuban crocodile (Crocodylus rhombifer)
Mugger crocodile (Crocodylus palustris)
African dwarf croc (Osteolaemus tetraspis)
False Gharial (Tomistoma schegelii)
Gharial (Gavialis gangeticus)

Source of animals (if wild, list current collecting permit number):
wild and captive

Location where animals will be housed: McNeese LERC alligator handling facilities

Location where animals will be used:
Alligator handling facilities

7. Plan for disposition of sick animals (check all that apply):

( ) Euthanize (state method) _____

( ) Clinical Treatment

(X) Other procedures (specify) cervical dislocation followed by pithing

8. Plan for disposition of dead animals (check all that apply):

(X) Save, Freeze ( ) Save, Refrigerate ( ) Necropsy

( ) Other, (Specify) _____

9. Is there a potential for pain, distress or discomfort to the animals?

Yes ✓ No _____

If yes, use the appropriate number (see below) to indicate the maximum level of pain/discomfort/distress to be experienced by the animals.

Pain Level: 1

1 = Negligible, 2 = Pain avoided by appropriate drug use, 3 = Short-term discomfort/distress/pain, 4 = Long-term discomfort/distress/pain. (An explanation must be attached to the protocol for a pain level of 4.

10. Federal law requires a written statement of A. and B. below to be included in protocols for any use of applicable species. Presently excluded from this requirement are rats and mice of the genus Rattus and Mus, birds, and animals used in accepted agricultural practices.

A. "The principal investigator/Instructor has considered alternatives to procedures that may cause more than momentary or slight pain or distress to the animal, and has provided a written narrative description of the methods and sources used to determine that alternatives were not available".

B. "The principal investigator/instructor has provided written assurance that these activities do not unnecessarily duplicate previous experiments.

March 2006

McNeese State University
Office of Research Services and Sponsored Programs

ASSURANCES

The University has adopted on an institution-wide basis the principles regarding animal care as stated in the Animal Welfare Act (www.nal.usda.gov/awic/legislat/awa.htm) and the Guide for the Care and Use of Laboratory Animals (www.nap.edu/readingroom/books/labrats/) and is guided by the U.S. Government Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research, and Training (http://grants1.nih.gov/grants/olaw/references/phspol.htm). All university employees and students are responsible for adherence to this policy. All animals owned and cared for by the university are covered by this policy.

I agree to comply with all Federal, State, and University animal welfare laws and policies during this project. I am aware that the new animal use protocol form must be submitted to RSSP and the IACUC if the protocol described above is altered after it is begun. The information provided is accurate, complete and true to the best of my knowledge.

Name (Print): Mark Merchant

Title: Associate Professor

Signature: _____
Investigator/Instructor

Date: 10/09/09

March 2006

McNEESE STATE UNIVERSITY

Mark E. Merchant, Ph. D.
Associate Professor of Biochemistry
P.O. Box 90455, Lake Charles, LA 70609-0455
Phone (337) 475-5773, Fax (337) 475-5950
mmerchan@mail.mcneese.edu

Chemistry Department

October 10, 2009

McNeese ACUC Committee,

This letter is written to serve as a resubmission of protocols used to collect serum, plasma, or whole blood from alligators (*Alligator mississippiensis*), and other crocodilian species (enumerated on ACUC form).

It is common practice for research personnel to collect blood from the spinal vein of crocodilians (Olson *et al.*, 1975, Zippel *et al.*, 2003). We typically use a 2.75 cm 20 ga. needle. The needle is inserted into the dorsal side of the neck, approximately halfway between the back of the skull and the first set of dorsal scutes. I have bled literally thousands of wild and captive alligators and crocodiles using this method during the past eight years. In addition, we bleed alligators *via* direct cardiac puncture on rare occasions using 26 ga. needles. The requirement to collect several blood samples in a relatively short time frame from small alligators during kinetic studies necessitates the cardiac bleed. We have found that repeated blood collection from the spinal vein of small alligators (< 1 m) sometimes compromises the integrity of the spine of the smaller animals. Cardiac puncture has been described by Lloyd and Morris (1999) and is used widely by other researchers using crocodilians as animals models (Busk *et al.*, 2003, Stevenson *et al.*, 1957).

If you have any questions or need additional information, please contact me at x5773.

Thank you,

Mark Merchant

Busk M., Overgaard J., Hicks J., Bennet A., Wang T. (2003). Effects of feeding on arterial blood gases in the American alligator Alligator mississippiensis. *J. Exper. Biol.* 203 (20), 3117-3124.

Lloyd M, Morris, P. (1999). Phlebotemy techniques in crocodilians. *Bull. Assoc. Rept. Amphib. Vet.* 9, 12-13.

Olson G. A., Hessler J. R., and Faith R. E. (1975). Technics for blood collection and intravascular infusion of reptiles. *Lab. Anim. Sci.* 25(6), 783-6.

Stevenson, O., Coulson, R., and Hernandez, T. (1957). Effects of Hormones on Carbohydrate Metabolism in the Alligator. *Amer. J. Physiol.* 191, 95-102.

Zippel K., Lillywhite H., and Mladnich. (2003). Anatomy of the crocodilian spinal vein. *J. Morphology* 258, 327-335.

VITA

Lancia Nadina Fallen Darville-Bowleg was born in Freeport, Bahamas, to her parents, Christine Darville and Lancelot Darville Sr. Lancia graduated with a Bachelor of Science degree in 2003 from University of Nebraska at Kearney and a Master of Science degree in chemistry in 2005 from University of Nebraska-Lincoln. During her graduate study, her research was focused on the classification and identification of explosives to be developed into a chemical library for use by the Nebraska State Crime Lab.

In 2005, Lancia enrolled in the doctoral program at Louisiana State University and A&M College, where she joined Dr. Kermit K. Murray's research group. Lancia's research included application of mass spectrometry-based proteomics for identification, determination and characterization of immunological proteins from alligators. As a graduate student, Lancia has worked in collaboration with Dr. Mark E. Merchant and has submitted three articles to date. Her work has been presented at local and regional conferences, where she has given four oral presentations, twelve first-author and two collaborative poster presentations. Lancia's research has also been highlighted in both national and international science journals and magazines as well as local and national news broadcasts.

While attending LSU she received several fellowships and awards: Graduate Alliance for Education in Louisiana (GAELA) supplement, the Charles E. Coates travel award, Teaching Scholar Award for Excellence in Teaching and Outstanding Researcher. In 2009, she became affiliated with the Howard Hughes Medical Institute (HHMI) Math and Science Program as a science mentor where she executed leadership in the development and academic guidance of students interested in STEM areas. Her professional affiliations include the American Chemical Society (ACS), American Society of Mass Spectrometry (ASMS), the National Organization for

the Advancement of Black Chemists and Chemical Engineers (NOBCChE) and Iota Sigma Pi (ISP) National Honor Society of Women in Chemistry. Lancia is currently a candidate for the degree of Doctor of Philosophy in chemistry, which will be conferred at the December 2011 commencement.