In: Leukocytes: Biology, Classification and Role in Disease ISBN: 978-1-62081-404-8
Editors: Giles I. Henderson and Patricia M. Adams

Chapter 1

MASS SPECTROMETRY-BASED PROTEOMICS FOR STRUCTURE AND FUNCTION DETERMINATION OF ALLIGATOR LEUKOCYTE PROTEINS

Lancia N. F. Darville[1], Mark E. Merchant[2] and Kermit K. Murray[1]

[1]Louisiana State University, Baton Rouge, LA, US
[2]McNeese State University, Lake Charles, LA, US

ABSTRACT

Mass spectrometry-based proteomics can be used to investigate the properties of alligator leukocytes and better understand the alligator innate immune system. Proteomics is critical to understanding complex biological systems through the elucidation of protein expression, function, modifications, and interactions. Separation techniques, including liquid chromatography, gel electrophoresis, and ion mobility as well as mass spectrometry tools such soft ionization and high resolution mass separation are available for proteomic applications. With organisms such as reptiles, for which there is limited genomic and proteomic data, a *de novo* sequencing approach can be implemented. In this chapter, we describe the use of mass spectrometry-based proteomics to determine peptide sequences from alligator leukocyte proteins and identify proteins based on sequence homology. A description of the instruments and methods used for analyzing peptides and proteins isolated from alligator leukocytes is provided.

1. BIOCHEMISTRY OF THE IMMUNE SYSTEM

Crocodilians are vertebrates with a complex immune system comprising antimicrobial peptides, macrophages, heterophils, neutrophils, basophils, eosinophils, phagocytic B cells, and proteins of the complement system. Like other vertebrates, they have both innate and an adaptive immune system. [1, 2] Crocodilians thrive in microbe containing environments but exhibit a strong resistance to infections.

Alligator serum has been shown to have antibacterial, [3] antiviral, [4] and antiamoebacidal properties. [5] Antibacterial activity has also been observed in crocodile serum, particularly that of *Crocodylus siamensis*. [6, 7] In addition, the leukocytes of *Alligator mississippiensis* have been shown to produce a broad antimicrobial activity spectrum. [8] The alligator complement system, which is part of the innate immune system, has also been shown to be effective against gram-positive bacteria. [3, 9] When alligator serum was compared to human serum, the alligator serum was effective against different strains of Gram-positive bacteria, whereas human serum had no antibacterial activity. It was proposed that the complement is responsible for antiviral activity of 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. [4] These studies suggest that crocodilian leukocytes play a central role in their strong innate immune system.

1.1. Innate Immune System

The innate immune system comprises different biomolecules including lysozymes, proteins of the complement system, non-specific leukocytes, and antimicrobial peptides. [1] Lysozymes are enzymes that lyse bacterial cells by the hydrolysis of their cell wall. [10] Lysis entails the complement proteins rupturing the bacterial membrane therefore killing the invading bacteria. [11] Lysozymes have been isolated from several reptilian organisms such as, lizards, [12] turtles, [12] crocodiles, [13] and alligators. [14] The complement system consists of various proteins found in plasma that kill bacteria via lysis or opsonization. [1] 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, which then engulf the bacteria through phagocytosis. [1]

The complement immune system has been characterized in the American alligator [9] and is believed to be responsible for antiviral activity exhibited by alligator serum. [4] There are different pathways to the complement immune system: classical, alternative, and lectin. [15] The classical pathway is activated by the immunoglobulins, immunoglobulin G (IgG) and immunoglobulin M (IgM) that activate an immune response. [1, 15] The alternative pathway does not require antibodies but is activated by viruses or lipopolysaccharides (LPS) on the surface of bacteria. [11] Finally, the lectin pathway is activated by mannose sugars of proteins that are on the cell surface of bacteria. [15]

Non-specific leukocytes in reptiles include eosinophils, herterophils, basophils, monocytes, and macrophages. [1] Limited information is available about the function of eosinophils in reptiles; however, in mammals they play a key role in the defense against parasitic infections. [15] Heterophils are involved in the inflammatory response in reptiles and are also responsible for suppressing microbial invasion. [16] Basophils contain immunoglobulins on their surface and, when triggered by an antigen, release histamine. [17] Monocytes and macrophages are phagocytic cells that are responsible for processing and releasing antigens as well as releasing cytokines, [15] regulatory proteins that generate an immune response. [18]

Antimicrobial peptides and proteins are important components of the innate immune system. Antimicrobial peptides are found in the host defense system [19] and are typically amphipathic and cationic and less than 10 kDa in mass. [20] However, they can also be

anionic peptides [21] and proteins. [22] Antimicrobial peptides can be linear α-helical with12–25 residues [20] or cysteine containing and β-sheet with several antiparallel β-strands that are stabilized with up to six disulfide bonds. [20] There are also antimicrobial peptides that are rich in specific residues such as tryptophan, [23] proline, arginine, [24, 25] and histidine. [26]

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. [19] They are cationic peptides that bind to microbes via electrostatic interactions. [27] Defensins have been found in mammals and birds [28] and recently the first reptilian defensin was discovered in the European pond turtle *Emys orbicularis*. [29] Defensins contain between 38-42 residues and are found in cells and tissues involved in the host defense system, exhibit antibacterial, antifungal, and antiviral activity. [19] In many animals, the highest concentration of defensins is found in the granules, where the leukocytes are stored. [19] Another family of antimicrobial peptides is cathelicidins, linear molecules that range in size from 12 to 80 amino acids that have antifungal [30] and antibacterial activity. [31] Unlike the β-defensins, they lack disulfide bridges. [32] Cathelicidins are produced in the myeloid cells in the bone marrow and stored in the neutrophil granules. [33] They have been isolated from fish, [34] birds, [35] mammals, [36] and reptiles. [37] Cathelicidins have also been found in monocytes, epithelial cells of the skin, respiratory tract, urogenital tract, as well as T and B lymphocytes. [33]

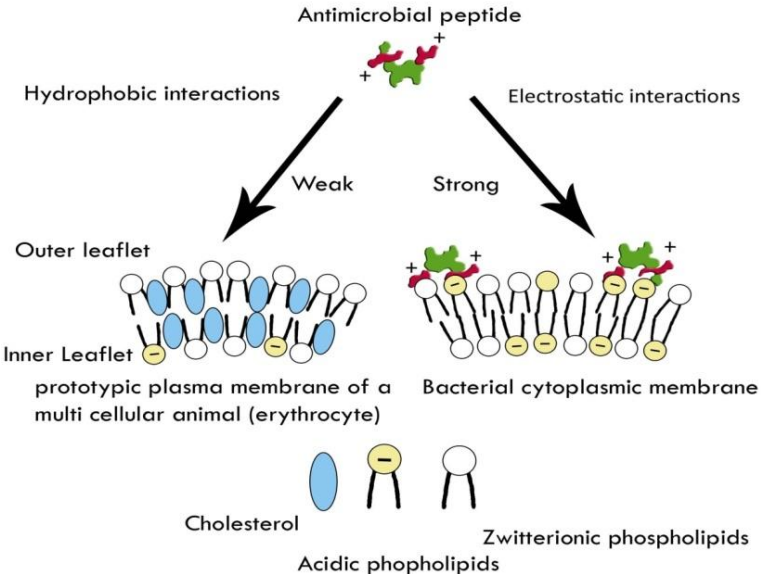

Figure 1. Comparison of antimicrobial peptide interaction with animal and bacterial membranes. (Modified from ref 38).

The membranes of microbes and multicellular animals differ in that microbes, such as 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); the negatively charged head groups are oriented towards the cytoplasm. [38, 39]

Table 1. Antimicrobial peptides available for therapeutic use. [45, 49]

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

The primary model that explains the mechanism of antimicrobial peptides is the Shai-Matsuzaki-Huang model, [39, 40] which 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. [40] The membrane of multicellular animal cells contains cholesterol, which reduces the activity of the antimicrobial peptide via interaction with the cholesterol or stabilization of the lipid bilayer. In addition to this mechanism, studies suggest that there are other mechanisms involved: 1) depolarization of the bacterial membrane leading to death, [41] 2) creation of holes in the cell wall causing cellular leakage, [42] 3) activation of processes that cause cell death (e.g. destroys the cell wall by hydrolases), [43] 4) disruption of the membrane function by rearrangement of the lipids on the outer cell membrane, [39] and 5) destruction of the internal cellular components after internalization of the antimicrobial peptide. [44]

Some antimicrobial peptides have been tested for pharmaceutical use, primarily as a topical treatment. The first antimicrobial peptide to undergo clinical trial was magainin isolated from frog skin, which was developed as a topical treatment for diabetic patients suffering from foot ulcers. [45, 46] Phase III clinical trials showed the magainin peptide to be as effective as the oral antibiotic ofloxacin. [47] A variety of antimicrobial peptides are currently being developed as potential antibiotics as indicated in Table 1.

The broad spectrum of the antimicrobial peptides allows them to be used for diverse applications; however, a major drawback of for clinical use is the level of toxicity: the level at which the antimicrobial peptides are effective *in vivo* are usually toxic. [48] Other factors are stability and immunogenicity. [45]

1.2. Adaptive Immune System

The adaptive immune system is activated following the innate immune response. There are two adaptive immune responses: cell-mediated and humoral adaptive. [1] 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. TH helps control other immune cells. T-cells also release cytokines that affect the humoral response (the immune responses mediated by antibodies). [1] Humoral adaptive immunity relies on B-cells that recognize antigens and initiate responses to protect the body from foreign material. [15]

1.3. Study of the Immune System

Serological assays are often used to study the immune system of vertebrates. [50] There are many different types of serological assays including agglutination, precipitation, immunoassays, immunofluorescence, fluorescence-activated cell sorting analysis, and lymphocyte function in which *in vitro* reactions between antigen and serum antibodies are studied. 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. [50] 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 chemical biology, which is discussed in Section 2. Genomics and proteomics provide a molecular level view of the immune system and its function as compared to serological assays and hematology tests.

2. CHEMICAL ANALYSIS METHODS FOR PROTEOMICS

Proteomics is the study of protein structure and function and is critical to the understanding of complex biological systems in terms of protein expression, function, modifications, and interactions. [51] Proteomics is related to genomics, which is the study of the genetic make-up of an organism. [52] An important step towards understanding an organism's biology is to determine its genome sequence. However, that is not enough to provide information on complex cellular processes; the complement of proteins associated with a particular genome is essential to this understanding. [53]

Proteomics is complimentary to genomics and provides an additional component to the understanding of biological systems. However, there are significant challenges such as

limited sample quantity, sample degradation, broad dynamic range (>10^6 for protein abundance), post-translational modifications, and disease changes. [54] Protein concentrations typically exceed the dynamic range of a single analytical instrument or method necessitating the use of one or more dimensions of separation. [51]

Proteomics can be divided into three major branches: structural proteomics, [55] expression proteomics, [56] and functional proteomics. [56] Structural proteomics involves determining the structures of proteins, such as their shape (secondary and tertiary structure) and amino acid sequence (primary structure). [55] A commonly used approach for protein sequencing is Edman degradation, which was developed by Pehr Edman in 1950 and is one of the oldest and most developed techniques for protein sequencing. [57] However, Edman sequencing has largely been replaced by mass spectrometry for protein sequencing and identification. [58, 59] Protein secondary and tertiary structures can be characterized by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy in addition to mass spectrometry. [55]

Expression proteomics involves the quantitative and qualitative analysis of proteins under different conditions. [56] 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, [60] multi-dimensional chromatography with mass spectrometry, [61] and protein micro-arrays. [62] Two-dimensional electrophoresis suffers from limitations such as the large dynamic range of protein expression in biological systems and difficulty in sequencing proteins that are post-translationally modified. [56] Limitations of microarray technology include the sensitivity of the arrays for detection of low abundance genes [63] as well as its inability to identify post-translational modifications. [64] Unlike 2-D electrophoresis and micro-array techniques, multi-dimensional chromatography with mass spectrometry can identify post-translational modifications in expression proteomics. [65]

Functional proteomics is an approach to analyze and understand macromolecular networks in cells. [56] Proteins and their specific roles in metabolic activities can be identified.

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 [66] and many proteins of interest are present in low abundance. Plasma comprises ~97% high abundance proteins including albumin (57–71%) and immunoglobulins (8–26%), [67] hence plasma samples are difficult to analyze directly and purification is required. In addition, samples usually contain contaminants such as lipids, nucleic acids, and surfactants. [68] The dynamic range of expressed proteins is greater than six orders of magnitude [69] and protein mixtures can be quite complex and contain proteins with different solubility, hydrophobicity, pI, and molecular masses. [70] Therefore, sample purification is necessary before mass spectrometry analysis.

Separation methods such as affinity-based techniques, chromatography, and centrifugation have been employed. [71] 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 [72] and protein A or protein G that selectively capture IgG on the columns. [73] The removal of salts and other contaminants from protein samples has been accomplished using precipitation with centrifugation. [68] Commonly used precipitation methods include acetone, trichloroacetic acid (TCA), ammonium sulfate, and chloroform/methanol precipitation.

2.2. Digestion

Protein digestion can be performed using three common approaches: in-gel, [74, 75] in-solution, [76] or solid phase. [77, 78] With 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. [75] A major advantage of in-gel digestion is that it removes detergents and salts that can be interferants in the mass spectrometer. [75] However, a limitation to this method is the loss of peptides during in-gel digestion through binding to the polyacrylamide. [75] Another approach is in-solution digestion which entails digesting proteins directly in buffers or solvents such as ammonium bicarbonate or acetonitrile. [79] This approach is advantageous in that low abundance molecules that may otherwise be lost in the gel can be detected; however, it has longer incubation times due to lower enzymatic concentrations. [80] Solid phase digestion is another approach which includes immobilization of an endoprotease on a solid support, for example monolithic columns [81] or microfluidic devises with integrated trypsin digestion. [78] Solid phase digestion offers the advantages of speed, reduced interference from trypsin autolysis products, and low sample consumption. [82, 83] A limitation to this method is the use of organic solvents which improve digestion efficiency but can damage the immobilized enzyme. [84]

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. [85]

Most 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. [86] Endoproteases and chemicals used for protein analysis are indicated in Table 2. The most commonly used endoprotease for proteomics is trypsin. Trypsin cleaves at lysine and arginine residues, unless followed by a proline in the C-terminus direction [87] and has good activity in both in-gel and in-solution digests.

Table 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. [88]

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[89]	C
Hydroxylamine, pH 9[90]	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. [91, 92] 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 information for database searching or for *de novo* sequencing if the protein sequence is not known.

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. [86] 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. [93]

2.3. Separations

Two commonly used separation methods in proteomics are gel electrophoresis and liquid chromatography (LC). Gel electrophoresis is an efficient method for separating and identifying proteins in a gel matrix such as agarose or polyacrylamide. [68] Agarose is typically used to separate larger macromolecules such as nucleic acids and polyacrylamide is used to separate proteins. Polyacrylamide gel electrophoresis can be used to determine the size, isoelectric point, and purity of proteins. [68] 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 and ratio of monomer to crosslinker. [94]

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 applied electric field and the electrophoretic mobility, which is in turn proportional to the ion charge and inversely proportional to the frictional forces. The frictional forces depend on the analyte's mass and the viscosity of the solvent. Smaller analytes have a greater mobility and migrate farther through 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. [95]

Two-dimensional gel electrophoresis (2D-GE) is used to separate proteins based on their isoelectric point (pI) and mass. [95, 96] The first dimension separates proteins based on their isoelectric point using isoelectric focusing. A protein's pI is determined by the type and number of acidic and basic residues it contains. Protein separation is performed in a pH gradient gel; the proteins migrate to the point in the gel at which their pI is identical to the pH at which point the protein has a net charge of zero. The second dimension in 2D-GE separates proteins based on their mass and is usually performed in a SDS gel (SDS-PAGE). Limitations of 2D-GE include difficulty resolving large proteins or those with extreme pI or hydrophobicity and lack of reproducibility. [97]

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 [98] and is one of the most powerful and commonly used liquid chromatography techniques. [99] Commonly used hydrocarbon ligands for reversed-phase resins include C_4 and C_{18}. [98] C_4 is commonly used for polar proteins and C_{18} is primarily used for peptides.

Ultra performance liquid chromatography (UPLC) uses smaller particles and has high speed and peak capacity (the number of peaks that can be resolved). [100] In contrast to conventional HPLC columns that are packed with 3.5 to 5 μm particles, UPLC columns are packed with 1.7 μm particles. [101] Smaller particles shorten the analyte's diffusion path which improves separation efficiency, speed, and resolution. [102, 103]

Ion exchange is a form of chromatography that separates proteins and peptides based on charge-charge interactions. [98] 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 and the stationary phase is polar. [104] Another separation technique is affinity chromatography, which separates proteins and peptides based on their specific ligand-binding affinity. [98] 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. [105] Therefore, affinity-based approaches can be used either to remove high abundant proteins or to enrich low abundant proteins. [97]

Due to the complexity of the protein samples, one-dimensional separation techniques are usually insufficient and multi-dimensional separations are employed. In multi-dimensional separations, two or more methods are coupled to improve the separation efficiency.

2.4. Mass Spectrometry

Mass spectrometry (MS) is an analytical technique used for measuring the mass and structure of molecules and is widely used for proteome analysis. [71] A variety of ionization techniques can be used for mass spectrometry but the most commonly used for the analysis of biomolecules are electrospray ionization (ESI) [106] and matrix assisted laser desorption ionization (MALDI). [107-109] MALDI uses a matrix that absorbs laser energy and aids in ionization of the analyte; the ions generated are typically singly charged. ESI uses a high voltage applied to a capillary to produce highly charged ions 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). [110] Tandem mass spectrometry is often used to determine peptide sequences from protein digests. [71] A peptide is selected in the first stage of mass spectrometry and dissociated by collision with an inert gas. The fragments are then separated in the second stage of mass spectrometry. [71, 111]

There are three MS-based 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 by 1 or 2-dimensional electrophoresis and the individual protein bands or spots are cut and digested with an enzyme such as trypsin to produce peptides. The peptides are analyzed by mass spectrometry using MALDI peptide mass fingerprinting or with liquid chromatography and ESI tandem mass spectrometry (MS/MS) to create sequence tags for database searching. [112-114] 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. Bottom-up the most widely used technique in proteomics, [115] hence many bioinformatics tools are available. In addition, proteins can be separated from a complex mixture before digestion, aiding in identification. Drawbacks of this approach are the limited dynamic range [116] and difficulty separating membrane proteins. [117, 118]

In shotgun proteomics, a mixture of proteins is enzymatically digested and separated using strong cation-exchange chromatography (SCX) followed by reversed-phase liquid chromatography (RPLC). [97, 119] The separated peptides are subjected to tandem mass spectrometry and database searching. [119] A major advantage of the shotgun technique is that thousands of proteins can be identified in a single analysis and it is better suited to membrane proteins. However, limitations include the need for complex mixtures to be purified prior to separation, [120] limited dynamic range, [51] and bioinformatics challenges in identification of peptide and protein sequences from a large number of acquired spectra. [51]

In the top-down approach, intact proteins are separated by gel electrophoresis or HPLC before being introduced into the mass spectrometer. [112, 114] 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. [114] Top-down sequencing can be used to locate and characterize post-translational modifications, determine the complete protein sequence, and it minimizes time-consuming preparation steps such as digestion and separation of peptides. Conversely, spectra

generated by multiply charged proteins can be very complex and the suite of bioinformatics tools for protein identification is limited.

A unique approach to biomolecule separation is ion mobility spectrometry, which can be combined with MS in ion mobility mass spectrometry (IM-MS). [121-123] Ion mobility is a gas-phase technique that separates ions based on their drift velocity through a buffer gas in the presence of an electric field. [121, 122] 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. [121] Ions with larger cross-section 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. [121] 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. [124] However, many IM studies of peptides and proteins reported in literature use ESI. [125-127] IM can also be coupled with tandem MS to obtain additional peptide and protein information. [128]

IM-MS separations can produce similar separation efficiency compared to HPLC and CE. [122] LC-MS has a high dynamic range; [129] however, the LC separation limits sample throughput and the demands of the MS ion source can limit optimization of the LC separation. [130] IM has two major advantages over LC: it reduces separation time and it separates biomolecules into chemical classes based on their high order structure; [131, 132] for example peptides, DNA, oligonucleotides, and lipids or protein conformational classes such as α-helix and random coil. [133] However, IM-MS has its shortcomings for proteomic applications, specifically poor sensitivity and limited peak capacity. [122] Despite the limitations, IM-MS, has proven useful for proteomics, [122] metabolomics, [134] and glycomics. [135]

There are four commonly used mass analyzers for proteomics, quadrupoles, time-of-flight (TOF), ion trap, and Fourier transform ion cyclotron (FT-MS). [113] These mass analyzers can be used in tandem (MS/MS) for peptide and protein identification. Examples of tandem mass analyzers are linear ion trap (LIT), linear ion trap - orbitrap (LTQ-Orbitrap), quadrupole-fourier transform ion cyclotron resonance mass spectrometer (Q-FTICR), quadrupole-time of flight (Q-TOF), ion trap-time of flight (IT-TOF) and time-of-flight/time-of-flight (TOF/TOF) [136, 137].

Tandem mass spectrometry is often used to determine the sequence of peptide from protein digests. [138] A peptide is separated from a mixture of peptides in the first mass spectrometry stage and dissociated by collision with an inert gas or other means. The resulting fragments are separated in the second mass spectrometry stage producing a mass spectrum (MS/MS). [111, 138] Cleavage of the peptide backbone by the collisions usually occurs at the amide bond. A nomenclature has been proposed by Roepstorff and Fohlman [139] and later modified by Biemann [140] 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. The subscript denotes the residue counting from either the N or C terminus (Figure 2). The most common ions produced by low energy collisions are b and y-ions [141].

The most commonly used enzyme for digesting proteins is trypsin, which produces peptides with arginine or lysine residues at the C-terminus, therefore y-ions predominate.

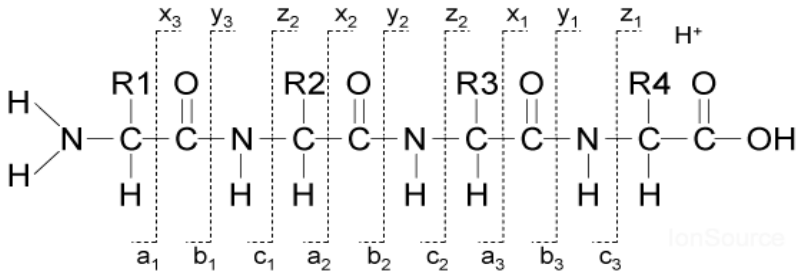

Figure 2. Common ions produced from peptide fragmentation.

Time-of-flight (TOF) is a commonly used mass analyzer that offers several advantages: 1) theoretically unlimited mass range which is limited in practice by detector, 2) high ion transmission, 3) simple to operate, and 4) low cost. TOF can be coupled to both electrospray and MALDI ion sources for analysis of biomolecules from a few hundred Da to greater than 150 kDa. [106, 142] MALDI-TOF is commonly used for protein identification via peptide-mass fingerprinting. [113] Instruments including Q-TOF and IT-TOF have been coupled with TOF mass analyzers for proteomic applications. Q-TOF and IT-TOF can be used for bottom-up and top-down proteomics as well as PTM identification. [137] Both instruments provide mass resolution above 10,000, mass accuracy between 2-5 ppm and detection limits at the attomole level. [137, 143] Hence, these instruments are applicable to *de novo* sequencing.

Tandem time-of-flight mass spectrometers can be thought of as two time-of-flight mass spectrometers tandem in space. Like other hybrid TOF instruments TOF/TOF can be used for protein identification and *de novo* sequencing. TOF/TOF instruments have been designed with different collision cell configurations. The Applied Biosystems (ABI) TOF-TOF instrument configuration consist of ions being decelerated into a collision cell where they collide with an inert gas and the products are reaccelerated from a second pulsed ion extraction source. [144] With the Bruker TOF-TOF instrument, ions are initially accelerated at lower energy, then collide with an inert gas in a collision cell that is then "lifted" to a high potential. [145] The "LIFT" cell increases the ion kinetic energy and a metastable suppressor removes unfragmented precursor ions. [145].

Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry is used for top-down proteomics and identification of modified proteins. It can perform multiple stages of mass spectrometry sequentially, and provides high mass resolution and mass accuracy. [146] FT-ICR has the capability to measure thousands of peptides and proteins in a complex mixture, provides low limit of detection for proteins (attomole level), high resolution, and MS/MS for proteins >100 kDa. [147, 148]

Ion trap instruments trap ions in a dynamic electric field and have the advantage of fast scan rates, MS^n scans, high sensitivity, and mass accuracy of 100 ppm. [137] Ion trap instruments can be used for both bottom-up [149] and top-down proteomics [150] due to their fast scan rates and sensitivity. [137] Ion traps can also be used as hybrid instruments. A commonly used hybrid ion trap is the LTQ-Orbitrap. In an orbitrap, ions circle between two electrodes and their axial motion is detected. [151] The LTQ-Orbitrap can be used for protein identification, quantification and identification of post-translational modification. [137] Therefore, the LTQ-Orbitrap is a more compact alternative for top-down sequencing of

proteins with detection limit ranging between attomole to femtomole range, mass accuracy of 2 ppm and isotopic resolution of small proteins. [137, 150]

3. PROTEIN BIOINFORMATICS

Bioinformatics is the approach used to analyze large numbers of genes and proteins [152] and is important for the analysis of mass spectrometry data due to the large quantities of data produced. [153] It is used in proteomics to provide functional analysis and mining of data sets. [152] Peptide and protein data can be interpreted via peptide mass fingerprinting, database searching, or *de novo* sequencing.

Peptide mass fingerprinting (PMF) is a protein identification method based on the accurate identification of peptide masses. [154] In this method proteins are separated and individually digested 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 protein matches are scored based. [154, 155] Several programs have been developed for peptide mass fingerprinting including MassSearch, MS-FIT, PepMAPPER, PepSea, PeptideSearch, ProFound and PeptIdent. [156] Peptide mass fingerprinting only provides 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. [157] 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 tandem mass spectra to theoretical spectra to identify peptides in the protein database. Theoretical tandem mass spectra are produced from fragmentation propensities that are known for a specific series of amino acids. Search engines that are used for database searching includes Mascot, [158], SEQUEST, [159] X!TANDEM, [160] Open mass spectrometry search algorithm (OMSSA), [161] SONAR, [162] ProbID, [163] PeptideProphet [164] and OLAV-PMF. [165] Two commonly used search engines include Mascot and SEQUEST. The latter uses a cross-correlation score to match hypothetical spectra to experimental spectra [159] whereas Mascot uses a score that indicates the probability of whether or not a spectral match was random. [158] 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. When some commonly used search engines were compared, including SEQUEST and Mascot, the latter proved to be able to better discriminate between a correct and incorrect hit as compared to SEQUEST. [166] An overall evaluation showed that Mascot outperformed the other algorithms used in the study, which included PeptideProphet, Spectrum Mill, SONAR and X!TANDEM.

Database searching offers several advantages including high-throughput, robustness, and annotated proteins (detailed information on each protein). Despite these advantages there are

also some disadvantages, 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 data (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. [167]

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 fragmentation method, for example collision induced dissociation (CID), [168] electron-transfer dissociation (ETD), [169] or electron-capture dissociation (ECD), [168, 170] the type of enzyme used, as well as any chemical modifications. Some commonly used *de novo* sequencing programs are PEAKS [171] Mascot Distiller [85], Lutefisk, [172] PepNovo, [173] and SHERENGA. [174] 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. [175] These nucleotide sequences are translated into protein sequences for protein identification from tandem mass spectra.

Under CID, peptides cleave 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. [176] *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. [176] 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. [167] 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. [168] 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. [167]

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. [177] This approach is useful for the identification of proteins from organisms that have unsequenced genomes. [178, 179] 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. [180, 181]

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,

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

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. [180]

4. ELUCIDATION OF ALLIGATOR LEUKOCYTE PROTEINS

The study of alligator leukocytes at the molecular level is key to understanding the function of the immune system and the proteins involved in its potency. Although this area of study is relatively new, there have been several recent studies in which the utility of proteomics approaches to the study of the crocodilian immune system has been demonstrated.

In a recent work from our group, proteins from the leukocytes of the American alligator (*Alligator mississippiensis)* were characterized using a bottom-up proteomics approach with bioinformatics for protein identification. [14] A three-step strategy was performed to identify similar proteins in the gel bands and spots: *de novo* sequencing, Mascot search, and BLAST search. As noted above, one of the major challenges in the study of the alligator blood proteome is the limited information available on the reptilian genome and proteome. [29, 182] Therefore, in this study, proteins from the alligator leukocytes were identified based on sequence similarity. Forty-three proteins with sequence similarity to the alligator leukocyte proteins were identified and found to be common among eukaryotes and associated with the immune system. The searches showed that the alligator proteins matched similar proteins in the database. The proteins were grouped based on their functionality where protein functions were divided into six groups: cytoskeletal proteins, immune proteins, enzymes, DNA/synthesis proteins, other function, and unknown function. Proteins involved in the cytoskeletal system, immune system, and other systems made up the three most abundant groups of peptides with 37%, 23% and 23% contribution, respectively. Examples of immune related proteins identified were myeloid protein, cathepsin C, and complement component c3.

Similar to alligators, crocodiles have also shown to have a similar and potent immune system and this potency is believed to be contributed by the presence of antimicrobial compounds.[8] In the leukocyte extract of crocodiles (*Crocodylus siamensis*), four novel antimicrobial peptides leucrocin I, II, III, and IV were purified using reversed-phase chromatography and sequenced using mass spectrometry. [183] The leucrocin peptides

showed to have different primary structures and were 7 to 10 amino acids long. The amino acid sequence of two of the leucrocin peptides, leucrocin I and leucrocin II whose masses were 804.9 and 847.9 Da were determined using tandem mass spectrometry. The sequences of leucrocin III and IV were not determined, due to the purity of those two peptides. The leucrocin peptides exhibit antibacterial activity against gram-negative bacteria, including Staphylococcus *epidermidis, Salmonella typhi* and *Vibrio cholerae*. The toxicity of the leucrocin peptides towards human red blood cells *in vitro* was also studied. Leucrocin I was shown to have mild toxic effects on human red blood cells while leucrocin III and IV peptides did not exhibit toxicity to human red blood cells at the tested concentrations. However, leucrocin II exhibited the highest toxicity toward human red blood cells. The interaction between leucrocin peptides and bacterial membranes were also studied to determine the peptide's bactericidal mechanism. The results showed that the four leucrocin peptides permeabilize the outer membrane of bacterial cells and leucrocin II, III, and IV disrupt the liposome membrane. The leucrocin peptides isolated from crocodile leukocytes were fully characterized for antimicrobial behavior and leucrocin I and II show unique sequence structure when compared to other antimicrobial peptides previously reported.

There is limited information available on alligator and crocodile leukocytes at the molecular level. However, another compound related to the immune system of crocodiles has been characterized at the molecular level using a proteomics approach. An antibacterial compound, crocosin, was isolated from crocodile (*Crocodylus siamensis)* blood plasma. Crocosin was partially purified using reversed-phase liquid chromatography and characterized using different biological assays. [183] Crocosin exhibited antimicrobial activity towards both gram-negative and gram-positive bacteria, *Salmonella typhi* and *Staphylococcus aureus* respectively. [184] Tandem mass spectrometry was used to elucidate the structure of crocosin; however, the tandem mass spectrometry results showed that crocosin may not be a peptide because the molecular masses did not correspond to the common amino acids. [184] Therefore, the structure of crocosin has not been determined.

Some studies have been performed to characterize the proteins and peptides from alligator leukocytes; [185] however, none have been fully characterized. Alligator leukocytes have been isolated and its extracts were separated by reversed phase chromatography and analyzed using tandem mass spectrometry. The antimicrobial activity of the peptides from the chromatographic fractions was tested for growth inhibition of various microbes and antibacterial activity was observed for Escherichia coli, *Enterobacter* cloacae and *Klebsiella oxytoca*, indicating that the antimicrobial peptides from the alligator are active against gram-negative bacteria. The masses of two molecules were identified in the fraction exhibiting antimicrobial activity using mass spectrometry. These molecules were subjected to tandem mass spectrometry for structure elucidation. The tandem mass spectrometry results showed that these molecules are peptides because of the amino acid sequences that were predicted. Peptide sequences were determined using *de novo* and manual sequencing. This work confirms that antimicrobial peptides are present in *Alligator mississippiensis* leukocytes and can be isolated and partially characterized. Based on preliminary results, these peptides exhibit characteristics similar to previously identified antimicrobial peptides.

Proteomic analysis of alligator leukocytes helps provide a better understanding of alligator and crocodile immunity as well as provide more insight on the evolutionary development of reptile's immune system. The initial characterization of the alligator leukocyte proteins [14] has paved the way for further investigations. Therefore, a more

exhaustive MS-based proteomics study should be performed to obtain a comprehensive overview of the proteins involved in the alligator's immune system. Proteomic studies of the alligator leukocytes may also contribute to drug discovery.

Alligator leukocyte extracts have been shown to inhibit the growth of bacterial, viral and fungal pathogens. [186] Inoculation of the Candida yeast species with alligator leukocyte extract showed rapid antifungal activity. There was strong antifungal activity observed for *Candida parapsilosis*, *Candida lusitinae*, and *Candida utilis.* Alligator leukocyte extract also showed strong antibacterial activity against both Gram-negative and Gram-positive species. The strongest antibacterial activity was observed for *Shigella flexneri*, *Citrobacter freundii*, *Streptococcus faecalis* and *Streptococcus pyogenes*, Moderate antibacterial activity was also observed for Escherichia coli, Pseudomonas aeruginosa, and *Salmonella* choleraesuis. Alligator leukocyte extract showed moderate antiviral activity against human immunodeficiency virus-1 (HIV-1) and herpes simplex virus-1 (HSV-1). However, the antiviral activity was compromised by the cytotoxicity of the leukocyte extract; therefore the antiviral activity can potentially be higher. This study indicates that alligators have molecules in their leukocytes that have broad-spectrum antimicrobial activities which can have significant clinical applications.

Investigation of the alligator leukocytes is ongoing. Given the strong antimicrobial characteristics of the alligator leukocytes and the significant information that can be obtained from using mass spectrometry-based proteomics, it is projected that proteomics will help unravel the proteome of alligator's immune system.

REFERENCES

[1] Zimmerman LM, Vogel LA, Bowden RM. Understanding the vertebrate immune system: insights from the reptilian perspective. J. Exp. Biol. 2010; 213:661-671.

[2] Merchant ME, Mills K, Leger N *et al.* Comparisons of innate immune activity of all known living crocodylian species. Comp. Biochem. Physiol. B 2006; 143:133-137.

[3] Merchant ME, Roche C, Elsey RM, Prudhomme J. Antibacterial properties of serum from the American alligator (Alligator mississippiensis). Comp. Biochem. Physiol. B 2003; 136:505-513.

[4] Merchant ME, Pallansch M, Paulman RL *et al.* Antiviral activity of serum from the American alligator (Alligator mississippiensis). Antiviral Res. 2005; 66:35-38.

[5] Merchant M, Thibodeaux D, Loubser K, Elsey RM. Amoebacidal effects of serum from the American alligator (alligator mississippiensis). J. Parasitol. 2004; 90:1480-1483.

[6] Preecharram S, Daduang S, Bunyatratchata W *et al.* Antibacterial activity from Siamese crocodile (Crocodylus siamensis) serum. Afr. J. Biotechnol. 2008; 7:3121-3128.

[7] Preecharram S, Jearranaiprepame P, Daduang S *et al.* Isolation and characterisation of crocosin, an antibacterial compound from crocodile (Crocodylus siamensis) plasma. Anim. Sci. J. 2010; 81:393-401.

[8] Merchant ME, Leger N, Jerkins E *et al.* Broad spectrum antimicrobial activity of leukocyte extracts from the American alligator (Alligator mississipiensis). Vet. Immunol. Immunopathol. 2006; 110:221-228.

[9] Merchant ME, Roche CM, Thibodeaux D, Elsey RM. Identification of alternative pathway serum complement activity in the blood of the American alligator (Alligator mississippiensis). Comp. Biochem. Physiol. B 2005; 141:281-288.

[10] Salton MRJ. Properties of lysozyme and its action on microorganisms Bacteriol. Rev. 1957; 21:82-100.

[11] Seelen MAJ, Roos A, Daha MR. Role of complement in innate and autoimmunity. J. Nephrol. 2005; 18:642-653.

[12] Thammasirirak S, Ponkham P, Preecharram S *et al.* Purification, characterization and comparison of reptile lysozymes. Comp. Biochem. Physiol. C 2006; 143:209-217.

[13] Supawadee Pata SD, Jisnuson Svasti and Sompong Thammasirirak Isolation of lysozyme like protein from crocodile leukocyte extract (Crocodylus siamensis) KMITL Sci. Technol. J. 2007; 7:70-85.

[14] Darville LNF, Merchant ME, Hasan A, Murray KK. Proteome analysis of the leukocytes from the American alligator (Alligator mississippiensis) using mass spectrometry. Comp. Biochem. Physiol. D 2010; 5:308-316.

[15] Richard Coico GS, and Eli Benjamini. Immunology, A Short Course. Hoboken, NJ: Wiley-Liss Publications; 2003.

[16] Montali RJ. Comparative pathology of inflammation in the higher vertebrates (reptiles, birds and mammals). J. Comp. Pathol. 1988; 99:1-26.

[17] Sypek JP, Borysenko M, Findlay SR. Anti-immunoglobulin induced histamine-release from naturally abundant basophils in the snapping turtle, Chelydra serpentina Dev. Comp. Immunol. 1984; 8:359-366.

[18] Gilman A GL, Hardman JG, Limbird LE. Goodman & Gilman's the pharmacological basis of therapeutics. New York: McGraw-Hill; 2001.

[19] Ganz T. Defensins: Antimicrobial peptides of innate immunity. Nature Rev. Immunol. 2003; 3:710-720.

[20] Yount NY, Bayer AS, Xiong YQ, Yeaman MR. Advances in antimicrobial peptide immunobiology. Biopolymers 2006; 84:435-458.

[21] Lai R, Liu H, Lee WH, Zhang Y. An anionic antimicrobial peptide from toad Bombina maxima. Biochem. Biophys. Res. Commun. 2002; 295:796-799.

[22] Silphaduang U, Hincke MT, Nys Y, Mine Y. Antimicrobial proteins in chicken reproductive system. Biochem. Biophys. Res. Commun. 2006; 340:648-655.

[23] Selsted ME, Novotny MJ, Morris WL *et al.* Indolicidin, a novel bactericidal tridecapeptide amide from neutrophils. J. Biol. Chem. 1992; 267:4292-4295.

[24] Frank RW, Gennaro R, Schneider K *et al.* Amino acid sequences of two proline-rich bactenecins - Antimicrobial peptides of bovine neutrophils. J. Biol. Chem. 1990; 265:18871-18874.

[25] Gudmundsson GH, Magnusson KP, Chowdhary BP *et al.* 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.

[26] Kacprzyk L, Rydengard V, Morgelin M *et al.* Antimicrobial activity of histidine-rich peptides is dependent on acidic conditions. Biochim. Biophys. Acta - Biomembranes 2007; 1768:2667-2680.

[27] Lehrer RI, Ganz T. Defensins of vertebrate animals. Curr. Opin. Immunol. 2002; 14:96-102.

[28] Ganz T. Defensins: antimicrobial peptides of vertebrates. C.R. Biol. 2004; 327:539-549.

[29] Stegemann C, Kolobov A, Leonova YF *et al.* Isolation, purification and de novo sequencing of TBD-1, the first beta-defensin from leukocytes of reptiles. Proteomics 2009; 9:1364-1373.

[30] Lopez-Garcia B, Lee PHA, Yamasaki K, Gallo RL. Anti-fungal activity of cathelicidins and their potential role in Candida albicans skin infection. J. Invest. Dermatol. 2005; 125:108-115.

[31] Wang YP, Hong J, Liu XH *et al.* Snake Cathelicidin from Bungarus fasciatus Is a Potent Peptide Antibiotics. Plos One 2008; 3.

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

[33] Zanetti M. Cathelicidins, multifunctional peptides of the innate immunity. J. Leukocyte Biol. 2004; 75:39-48.

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

[35] Xiao YJ, Cai YB, Bommineni YR *et al.* Identification and functional characterization of three chicken cathelicidins with potent antimicrobial activity. J. Biol. Chem. 2006; 281:2858-2867.

[36] 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.

[37] Zhao H, Gan TX, Liu XD *et al.* Identification and characterization of novel reptile cathelicidins from elapid snakes. Peptides 2008; 29:1685-1691.

[38] Zasloff M. Antimicrobial peptides of multicellular organisms. Nature 2002; 415:389-395.

[39] 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.

[40] 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.

[41] Westerhoff HV, Juretic D, Hendler RW, Zasloff M. Magainins and the disruption of membrane-linked free-energy transduction Proc. Natl. Acad. Sci. U S A 1989; 86:6597-6601.

[42] Yang L, Weiss TM, Lehrer RI, Huang HW. Crystallization of antimicrobial pores in membranes: Magainin and protegrin. Biophys. J. 2000; 79:2002-2009.

[43] Bierbaum G, Sahl HG. Induction of autolysis of staphylococci by the basic peptide antibiotics Pep 5 and nisin and their influence on the activity of autolytic enzymes Archives of Microbiology 1985; 141:249-254.

[44] Kragol G, Lovas S, Varadi G *et al.* The antibacterial peptide pyrrhocoricin inhibits the ATPase actions of DnaK and prevents chaperone-assisted protein folding. Biochemistry 2001; 40:3016-3026.
[45] Hancock REW. Peptide antibiotics. Lancet 1997; 349:418-422.
[46] Jacob L, Zasloff M. Potential therapeutic applications of magainins and other antimicrobial agents of animal origin. In: Antimicrobial Peptides. Marsh J, Goode JA (Editors). 1994. pp. 197-216.
[47] Hancock REW, Lehrer R. Cationic peptides: a new source of antibiotics. Trends Biotechnol. 1998; 16:82-88.
[48] Darveau RP, Cunningham MD, Seachord CL *et al.* Beta-lactam antibiotics potentiate magainin 2 antimicrobial activity in vitro and in vivo. Antimicrob. Agents Chemother. 1991; 35:1153-1159.
[49] Lazarev VN, Govorun VM. Antimicrobial peptides and their use in medicine. Appl. Biochem. Microbiol. 2010; 46:803-814.
[50] Coico R, Sunshine G. Immunology : a short course. Hoboken, N.J.: Wiley-Blackwell; 2009.
[51] Domon B, Aebersold R. Review - Mass spectrometry and protein analysis. Science 2006; 312:212-217.
[52] Ganten DR, Klaus Encyclopedic reference of genomics and proteomics in molecular medicine. In: Berlin Heidelberg New York: Springer; 2006.
[53] Yarmush ML, Jayaraman A. Advances in proteomic technologies. Annu. Rev. Biomed. Eng. 2002; 4:349-373.
[54] Tyers M, Mann M. From genomics to proteomics. Nature 2003; 422:193-197.
[55] Norin M, Sundström M. Structural proteomics: developments in structure-to-function predictions. Trends Biotechnol. 2002; 20:79-84.
[56] Kocher T, Superti-Furga G. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. Nat. Methods 2007; 4:807-815.
[57] Edman P. Method for determination of the amino acid sequence in peptides Acta Chem. Scand. 1950; 4:283-293.
[58] Fenselau C. Beyond Gene Sequencing: Analysis of Protein Structure with Mass Spectrometry. Annu. Rev. Biophys. Biophys. Chem. 1991; 20:205-220.
[59] Yates JR. Mass spectrometry and the age of the proteome. J Mass Spectrom 1998; 33:1-19.
[60] Görg A, Weiss W, Dunn MJ. Current two-dimensional electrophoresis technology for proteomics. Proteomics 2004; 4:3665-3685.
[61] Zheng SP, Schneider KA, Barder TJ, Lubman DM. Two-dimensional liquid chromatography protein expression mapping for differential proteomic analysis of normal and O157 : H7 Escherichia coli. BioTechniques 2003; 35:1202.
[62] Debouck C, Goodfellow PN. DNA microarrays in drug discovery and development. Nat. Genet. 1999; 21:48-50.
[63] Bunney WE, Bunney BG, Vawter MP *et al.* Microarray Technology: A Review of New Strategies to Discover Candidate Vulnerability Genes in Psychiatric Disorders. Am. J. Psychiat. 2003; 160:657-666.
[64] Luo Z, Geschwind DH. Microarray applications in neuroscience. Neurobiol. Dis. 2001; 8:183-193.

[65] Romijn EP, Krijgsveld J, Heck AJR. Recent liquid chromatographic-(tandem) mass spectrometric applications in proteomics. J. Chromatogr. A 2003; 1000:589-608.
[66] Adkins JN, Varnum SM, Auberry KJ *et al.* Toward a human blood serum proteome - Analysis by multidimensional separation coupled with mass spectrometry. Mol. Cell Proteomics 2002; 1:947-955.
[67] Anderson NL, Anderson NG. The human plasma proteome - History, character, and diagnostic prospects. Mol. Cell Proteomics 2002; 1:845-867.
[68] 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.
[69] Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez JC. The dynamic range of protein expression: A challenge for proteomic research. Electrophoresis 2000; 21:1104-1115.
[70] Shen YF, Smith RD. Proteomics based on high-efficiency capillary separations. Electrophoresis 2002; 23:3106-3124.
[71] Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass spectrometry. Annu. Rev. Biochem. 2001; 70:437-473.
[72] 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.
[73] Bjorck L, Kronvall G. Purification and some properties of streptococcal protein G, protein-A novel IgG-binding reagent. J. Immunol. 1984; 133:969-974.
[74] Shevchenko A, Tomas H, Havlis J *et al.* In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat. Protoc. 2006; 1:2856-2860.
[75] Rosenfeld J, Capdevielle J, Guillemot JC, Ferrara P. In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. Anal. Biochem. 1992; 203:173-179.
[76] Medzihradszky KF. In-solution digestion of proteins for mass spectrometry. In: Methods Enzymol. Burlingame AL (Editor) Academic Press; 2005. pp. 50-65.
[77] Wang S, Regnier FE. Proteolysis of whole cell extracts with immobilized enzyme columns as part of multidimensional chromatography. J. Chromatogr. A 2001; 913:429-436.
[78] Lee J, Soper SA, Murray KK. Development of an efficient on-chip digestion system for protein analysis using MALDI-TOF MS. Analyst 2009; 134:2426-2433.
[79] KR SKW. Enzymatic digestion of proteins in solution and in SDS polyacrylamide gels. In: The Protein Protocols Handbook. 1996.
[80] Klammer AA, MacCoss MJ. Effects of Modified Digestion Schemes on the Identification of Proteins from Complex Mixtures. J. Proteome Res. 2006; 5:695-700.
[81] Peterson DS, Rohr T, Svec F, Frechet JMJ. 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.
[82] Peterson DS, Rohr T, Svec F, Fréchet JMJ. 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.
[83] Sanders GHW, Manz A. Chip-based microsystems for genomic and proteomic analysis. TrAC, Trends Anal. Chem. 2000; 19:364-378.

[84] Liu JY, Lin S, Qi DW *et al.* On-chip enzymatic microreactor using trypsin-immobilized superparamagnetic nanoparticles for highly efficient proteolysis. J. Chromatogr. A 2007; 1176:169-177.
[85] Deutsch EW, Lam H, Aebersold R. Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. Phycol. Genomics 2008; 33:18-25.
[86] Liebler DC. Introduction to Proteomics: Tools for the New Biology. Totowa, NJ: Humana Press Inc; 2002.
[87] Aitken Aea. Protein sequencing: A practical approach. Oxford: IRL Press; 1989.
[88] Twyman RM. Principles of proteomics. New York, NY: BIOS Scientific 2004.
[89] Jacobson GR, Schaffer MH, Stark GR, Vanaman TC. Specific chemical cleavage in high-yield at amino peptide-bonds of cysteine and cysteine residues J. Biol. Chem. 1973; 248:6583-6591.
[90] Bornstei.P, Balian G. Specific nonenzymatic cleavage of bovine ribonuclease with hydroxylamine J. Biol. Chem. 1970; 245:4854-&.
[91] Drapeau GR. Protease from Staphyloccus aureus. In: Methods Enzymol. Laszlo L (Editor) Academic Press; 1976. pp. 469-475.
[92] Drapeau GR. Cleavage at glutamic acid with staphylococcal protease. In: Methods Enzymol. Hirs CHW, Serge NT (Editors). Academic Press; 1977. pp. 189-191.
[93] Villa S, De Fazio G, Canosi U. Cyanogen bromide cleavage at methionine residues of polypeptides containing disulfide bonds. Anal. Biochem. 1989; 177:161-164.
[94] Westermeier R, Naven T. Proteomics in practice : a laboratory manual of proteome analysis. Weinheim: Wiley-VCH; 2002.
[95] O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. 1975; 250:4007-4021.
[96] Gorg A, Obermaier C, Boguth G *et al.* The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 2000; 21:1037-1053.
[97] Ye ML, Jiang XG, Feng S *et al.* Advances in chromatographic techniques and methods in shotgun proteome analysis. Trends Anal. Chem. 2007; 26:80-84.
[98] Kastner M. Journal of chromatography library In: Protein Liquid Chromatography. Amsterdam, The Netherlands: Elsevier; 2000.
[99] Dong MW. Modern HPLC for practicing scientists. Hoboken, N.J.: Wiley-Interscience; 2006.
[100] Swartz M. UPLC ™ : An Introduction and Review. J. Liq. Chromatogr. Relat. Technol. 2005; 28:1253-1263.
[101] de Villiers A, Lestremau F, Szucs R *et al.* Evaluation of ultra performance liquid chromatography - Part I. Possibilities and limitations. J. Chromatogr. A 2006; 1127:60-69.
[102] Swartz ME. Ultra performance liquid chromatography (UPLC): An introduction. Lc Gc North America 2005:8-14.
[103] de Villiers A, Lestremau F, Szucs R *et al.* Evaluation of ultra performance liquid chromatography: Part I. Possibilities and limitations. Journal of Chromatography A 2006; 1127:60-69.
[104] Alpert AJ. Hydrophilic-interaction chromatography for the separation of peptides, nucleic-acids and other polar compounds J. Chromatogr. 1990; 499:177-196.

[105] Lee HJ, Lee EY, Kwon MS, Paik YK. Biomarker discovery from the plasma proteome using multidimensional fractionation proteomics. Curr. Opin. Chem. Biol. 2006; 10:42-49.
[106] Fenn JB, Mann M, Meng CK *et al.* Electrospray ionization for mass spectrometry of large biomolecules. Science 1989; 246:64-71.
[107] Hillenkamp F, Karas M. Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption ionization. Methods Enzymol. 1990; 193:280-295.
[108] Hillenkamp F, Karas M, Beavis RC, Chait BT. Matrix-assisted laser desorption ionization mass spectrometry of biopolymers. Anal. Chem. 1991; 63:A1193-A1202.
[109] Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal. Chem. 1988; 60:2299-2301.
[110] Dass C. Fundamentals of contemporary mass spectrometry. Hoboken, N.J.: Wiley-Interscience; 2007.
[111] Biemann K, Scoble HA. Characterization by tandem mass-spectrometry of structural modifications in proteins Science 1987; 237:992-998.
[112] Chait BT. Mass spectrometry: Bottom-up or top-down? Science 2006; 314:65-66.
[113] Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature 2003; 422:198-207.
[114] Reid GE, McLuckey SA. 'Top down' protein characterization via tandem mass spectrometry. J Mass Spectrom 2002; 37:663-675.
[115] Han X, Aslanian A, Yates Iii JR. Mass spectrometry for proteomics. Curr. Opin. Chem. Biol. 2008; 12:483-490.
[116] VerBerkmoes NC, Bundy JL, Hauser L *et al.* Integrating "top-down" and "bottom-up" mass spectrometric approaches for proteomic analysis of Shewanella oneidensis. J. Proteome Res. 2002; 1:239-252.
[117] Wu CC, Yates JR. The application of mass spectrometry to membrane proteomics. Nature Biotechnol. 2003; 21:262-267.
[118] Santoni V, Molloy M, Rabilloud T. Membrane proteins and proteomics: Un amour impossible? Electrophoresis 2000; 21:1054-1070.
[119] Wu CC, MacCoss MJ. Shotgun proteomics: Tools for the analysis of complex biological systems. Curr. Opin. Mol. Ther. 2002; 4:242-250.
[120] Wehr T. Top-down versus bottom-up approaches in proteomics. Lc Gc North America 2006; 24:1004.
[121] Uetrecht C, Rose RJ, van Duijn E *et al.* Ion mobility mass spectrometry of proteins and protein assemblies. Chem. Soc. Rev. 2010; 39:1633-1655.
[122] McLean JA, Ruotolo BT, Gillig KJ, Russell DH. Ion mobility-mass spectrometry: a new paradigm for proteomics. Int. J. Mass spectrom. 2005; 240:301-315.
[123] Kanu AB, Dwivedi P, Tam M *et al.* Ion mobility-mass spectrometry. J Mass Spectrom 2008; 43:1-22.
[124] Clemmer DE, Jarrold MF. Ion mobility measurements and their applications to clusters and biomolecules. J Mass Spectrom 1997; 32:577-592.
[125] Liu XY, Valentine SJ, Plasencia MD *et al.* Mapping the human plasma proteome by SCX-LC-IMS-MS. J. Am. Soc. Mass. Spectrom. 2007; 18:1249-1264.
[126] Hoaglund-Hyzer CS, Lee YJ, Counterman AE, Clemmer DE. Coupling ion mobility separations, collisional activation techniques, and multiple stages of MS for analysis of complex peptide mixtures. Anal. Chem. 2002; 74:992-1006.

[127] Wyttenbach T, Kemper PR, Bowers MT. Design of a new electrospray ion mobility mass spectrometer. Int. J. Mass spectrom. 2001; 212:13-23.
[128] Pringle SD, Giles K, Wildgoose JL *et al.* 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.
[129] Mo WJ, Karger BL. Analytical aspects of mass spectrometry and proteomics. Curr. Opin. Chem. Biol. 2002; 6:666-675.
[130] Holland JF, Enke CG, Allison J *et al.* Mass-spectrometry on the chromatographic time scale - realistic expectations Anal. Chem. 1983; 55:A997-&.
[131] Woods AS, Ugarov M, Egan T *et al.* Lipid/peptide/nucleotide separation with MALDI-ion mobility-TOF MS. Anal. Chem. 2004; 76:2187-2195.
[132] Koomen JM, Ruotolo BT, Gillig KJ *et al.* Oligonucleotide analysis with MALDI-ion-mobility-TOFMS. Anal. Bioanal. Chem. 2002; 373:612-617.
[133] Ruotolo BT, Verbeck GF, Thomson LM *et al.* Observation of conserved solution-phase secondary structure in gas-phase tryptic peptides. J. Am. Chem. Soc. 2002; 124:4214-4215.
[134] Dwivedi P, Wu P, Klopsch SJ *et al.* Metabolic profiling by ion mobility mass spectrometry (IMMS). Metabolomics 2008; 4:63-80.
[135] Jin L, Barran PE, Deakin JA *et al.* Conformation of glycosaminoglycans by ion mobility mass spectrometry and molecular modelling. Phys. Chem. Chem. Phys. 2005; 7:3464-3471.
[136] Yates JR, Ruse CI, Nakorchevsky A. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. Annual Review of Biomedical Engineering 2009; 11:49-79.
[137] Yates JR, Ruse CI, Nakorchevsky A. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. Annu. Rev. Biomed. Eng. 2009; 11:49-79.
[138] Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass spectrometry. Annual Review of Biochemistry 2001; 70:437-473.
[139] Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed. Mass Spectrom. 1984; 11:601-601.
[140] Biemann K. Contributions of mass-spectrometry to peptide and protein-structure Biomedical and Environmental Mass Spectrometry 1988; 16:99-111.
[141] Wysocki VH, Resing KA, Zhang QF, Cheng GL. Mass spectrometry of peptides and proteins. Methods 2005; 35:211-222.
[142] Fenn JB, Mann M, Meng CK *et al.* Electrospray ionization-principles and practice Mass Spectrom. Rev. 1990; 9:37-70.
[143] Gygi SP, Aebersold R. Mass spectrometry and proteomics. Curr. Opin. Chem. Biol. 2000; 4:489-494.
[144] Samyn B, Debyser G, Sergeant K *et al.* A case study of de novo sequence analysis of N-sulfonated peptides by MALDI TOF/TOF mass spectrometry. Journal of the American Society for Mass Spectrometry 2004; 15:1838-1852.
[145] Suckau D, Resemann A, Schuerenberg M *et al.* A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. Anal. Bioanal. Chem. 2003; 376:952-965.
[146] Hu QZ, Noll RJ, Li HY *et al.* The Orbitrap: a new mass spectrometer. J. Mass Spectrom. 2005; 40:430-443.

[147] Tolmachev AV, Robinson EW, Wu S *et al.* FT-ICR MS optimization for the analysis of intact proteins. International Journal of Mass Spectrometry 2009; 287:32-38.
[148] Patrie SM, Charlebois JP, Whipple D *et al.* Construction of a hybrid quadrupole/Fourier Transform Ion Cyclotron Resonance Mass Spectrometer for versatile MS/MS above 10 kDa. Journal of the American Society for Mass Spectrometry 2004; 15:1099-1108.
[149] Yates JR, Cociorva D, Liao LJ, Zabrouskov V. Performance of a linear ion trap-orbitrap hybrid for peptide analysis. Anal. Chem. 2006; 78:493-500.
[150] Macek B, Waanders LF, Olsen JV, Mann M. Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer. Mol Cell Proteomics 2006; 5:949-958.
[151] Komatsu S. Edman Sequencing of Proteins from 2D Gels In: Plant Proteomics Methods and Protocols. Thiellement H, Zivy M, Damerval C, Méchin V (Editors). Totowa, NJ: Humana Press; 2007. pp. 211-217.
[152] Kumar C, Mann M. Bioinformatics analysis of mass spectrometry-based proteomics data sets. FEBS Lett. 2009; 583:1703-1712.
[153] Cristoni S, Bernardi LR. Bioinformatics in mass spectrometry data analysis for proteomics studies. Expert Rev. Proteomic 2004; 1:469-483.
[154] Gevaert K, Vandekerckhove J. Protein identification methods in proteomics. Electrophoresis 2000; 21:1145-1154.
[155] Cottrell JS, Sutton CW. The identification of electrophoretically separated proteins by peptide mass fingerprinting. In: 1996. pp. 67-82.
[156] Beavis RC, Fenyö D. Database searching with mass-spectrometric information. Trends Biotechnol. 2000; 18:22-27.
[157] Lam H, Deutsch EW, Eddes JS *et al.* Development and validation of a spectral library searching method for peptide identification from MS/MS. Proteomics 2007; 7:655-667.
[158] Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 1999; 20:3551-3567.
[159] Eng JK, Fischer B, Grossmann J, MacCoss MJ. A fast SEQUEST cross correlation algorithm. J. Proteome Res. 2008; 7:4598-4602.
[160] Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. Bioinformatics 2004; 20:1466-1467.
[161] Geer LY, Markey SP, Kowalak JA *et al.* Open mass spectrometry search algorithm. J. Proteome Res. 2004; 3:958-964.
[162] Field HI, Fenyo D, Beavis RC. 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.
[163] 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.
[164] Keller A, Nesvizhskii AI, 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.
[165] 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.

[166] Kapp EA, Schutz F, Connolly LM *et al.* An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: Sensitivity and specificity analysis. Proteomics 2005; 5:3475-3490.

[167] Hughes C, Ma B, Lajoie GA. De Novo Sequencing Methods in Proteomics. Proteome Bioinf. 2010:105-121.

[168] Standing KG. Peptide and protein de novo sequencing by mass spectrometry. Current Opinion in Structural Biology 2003; 13:595-601.

[169] Bertsch A, Leinenbach A, Pervukhin A *et al.* De novo peptide sequencing by tandem MS using complementary CID and electron transfer dissociation. Electrophoresis 2009; 30:3736-3747.

[170] Syka JEP, Coon JJ, Schroeder MJ *et al.* Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc. Natl. Acad. Sci. U S A 2004; 101:9528-9533.

[171] Ma B, Zhang KZ, Hendrie C *et al.* PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom 2003; 17:2337-2342.

[172] Johnson RS, Taylor JA. Searching sequence databases via de novo peptide sequencing by tandem mass spectrometry. Molecular Biotechnology 2002; 22:301-315.

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

[174] Dancik V, Addona TA, Clauser KR *et al.* De novo peptide sequencing via tandem mass spectrometry. Journal of Computational Biology 1999; 6:327-342.

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

[176] Papayannopoulos IA. The Interpretation of Coliision-Induced Dissociation Tandem Mass-Spectra of Peptides Mass Spectrom. Rev. 1995; 14:49-73.

[177] McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 2004; 32:W20-W25.

[178] Shevchenko A, Sunyaev S, Loboda A *et al.* 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.

[179] Waridel P, Frank A, Thomas H *et al.* Sequence similarity-driven proteomics in organisms with unknown genomes by LC-MS/MS and automated de novo sequencing. Proteomics 2007; 7:2318-2329.

[180] Altschul SF, Madden TL, Schaffer AA *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389-3402.

[181] Altschul SF, Gish W. Local alignment statistics. In: Computer Methods for Macromolecular Sequence Analysis. 1996. pp. 460-480.

[182] Jacobson ER. Infectious diseases and pathology of reptiles: color atlas and text. Boca Raton: CRC Press; 2007.

[183] Pata S, Yaraksa N, Daduang S *et al.* Characterization of the novel antibacterial peptide Leucrocin from crocodile (Crocodylus siamensis) white blood cell extracts. Dev. Comp. Immunol. 2011; 35:545-553.

[184] Preecharram S, Jearranaiprepame P, Daduang S *et al.* Isolation and characterisation of crocosin, an antibacterial compound from crocodile (Crocodylus siamensis) plasma. Animal Science Journal 2010; 81:393-401.

[185] Darville LNF, Merchant ME, Hasan A, Murray KK. 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.

[186] Mark E. Merchant NL, 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.