DEVELOPMENT OF SAMPLE COLLECTION METHODS AND PRELIMINARY IDENTIFICATIONS OF APHID SALIVARY PROTEINS

by

MANUJA R LAMABADUSURIYA

B.Sc. University of Colombo, Sri Lanka, 2004
M.Sc. University of New South Wales, Australia, 2006

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCICENCE

Department of Chemistry
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2009

Approved by:

Major Professor
Dr. Christopher T Culbertson

Abstract

The study of aphid salivary secretome has practical importance on understanding interactions of aphids and their host plants. Around 250 species of aphids out of the identified 4000 aphid species are considered as serious pests. The experiments were performed with pea aphids (*Acyrthosiphon pisum*) that were feeding on bean plants (*Vivia fabe*). Pea aphids feed on plant phloem sap by probing their stylet into the sieve elements of the plant and secreting saliva for external digestion. In order to collect aphid salivary proteins from the secreted saliva, small scale and large scale sample collection methods were carried out. The small scale sample method was performed in microfluidic devices using 10-25 aphids. Aphids were able to feed on the artificial diet by probing through a stretched Parafilm^{TM} and survived for 2-3 days in the microfluidic devices. The experiments proved that the aphid survival and feeding rate could be improved with the factors such as ventilation, light intensity and increasing diet volume. However it was difficult to collect sufficient amounts of aphid saliva for detection using small scale devices. The large scale sample collection method was performed by feeding 8000 aphids in large screened chamber for 24/48h. The collected salivary samples after undergone a concentration process was capable of collecting detectable aphid salivary secretions. The experimental conditions were adjusted to obtain optimized HPLC separations. Finally, LC/MS/MS followed by peptide sequence database searching were able to identify potential aphid salivary proteins.

Table of Contents

List of Figures

Acknowledgements

This research project and the thesis would have not been completed if not for the immense support and guidance rendered to me by my research supervisor, Dr. Christopher T. Culbertson, Associate professor, Department of Chemistry, Kansas State University. Thank you very much for your valuable advice and encouragement given to me throughout my research project.

I also extend my deepest gratitude to Dr. Gerald Reek, Professor, Department of Biochemistry, Kansas State University, for his valuable assistance and advice given me as the research collaborator, during this project.I'm grateful to Dr. John Reese, Professor, Department of Entomology, Kansas State University, for his valuable assistance and advice given to me and especially for his great effort in maintaining the aphid colony and providing the specimens for my experiments.

I also extend my sincere thanks to Dr. Amanda Meyer, for her valuable assistance on chromatographic analysis and all her generous support given to me during my research period.I also acknowledge Dr Scott Bean for the opportunity provided us to carry out HPLC in USDA GMPRC and all his valuable assistances and advices given during HPLC analysis. My special thanks convey to Yasuaki Hiromasa and Dr. John M. Tomich of Biotechnology Core Laboratory for all the assistance given during mass spectrometric analysis.

Sincere thanks go out to Mr. Ron Jackson, Research Technologist, Department of Chemistry, KSU, for his assistance on development of aphid feeding devices. Special thanks go out to all the Culbertson group members who helped in various ways during my work in the research laboratory.

Last but not least I thank my family and all my true friends for their helpfulness and understanding, without which this research project would not have been a success.

Dedication

This thesis is dedicated to my late father, my mother, grand mother and sister who gave me unconditional love throughout all my life. This is also dedicated to my wonderful husband who has been a great source of motivation and encouragement in so many difficult situations during my studies.

CHAPTER 1: Research background

1.1 Introduction

This dissertation is focused on the development of novel analytical techniques for the pre-concentration, separation and identification of aphid salivary proteins. In order to generate sufficient amounts of proteins for analysis, salivary samples were collected from several thousand aphids and concentrated using molecular cut off filters. The concentrated proteins were then subjected to separation by High Performance Liquid Chromatography (HPLC) and detection using Mass Spectrometry (MS). Protein identification was performed subsequently through database searching. In order to develop an efficient sample protocol, one must understand the aphid life cycle, its method of feeding on plants and its dietary requirements. This background, along with the instrumental methods used or developed for the sampling is discussed in the introduction.

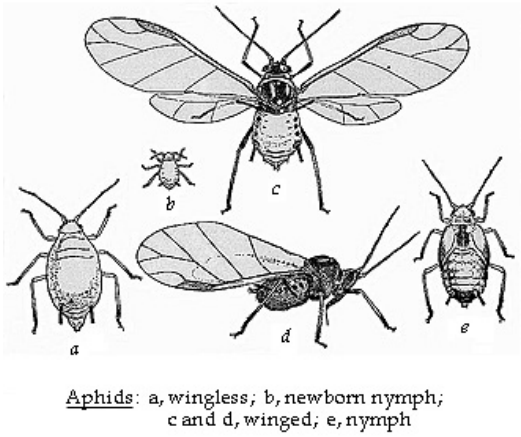
1.1.1 Pea aphid

As an important agricultural pest, studies of pea aphids, *Acyrthosiphon pisum*, and their interactions with host plant are very important. Around 250 out of 4000 species of aphids are considered pests.[1] They are commonly found in agricultural fields in the southeastern and northeastern United States. Aphids usually grow in colonies and prefer to live on the younger parts of plants.[2]

1.1.2 Life cycle and Reproduction

Aphids go through nymph, adult and winged stages during their life cycle (Figure 1.1). Nymphs are smaller in size but closely resemble the adults. A new born nymph molts by shedding its skin three to four times before becoming an adult. This process takes 10-14 days.[4] Most of the adult aphids are wingless, but some adults in the colony have two pairs of wings so that they can migrate to another colony or host plant if the current one is overcrowded.

Aphids reproduce both sexually and asexually. During the summer a single adult female can reproduce over several generations by asexual parthenogenesis. In general an adult can reproduce 4-12 nymphs per day. During winter, sexual reproduction is common by egg hatching (Figure 1.2).[2]



Aphids: a, wingless; b, newborn nymph; c and d, winged; e, nymph

Figure 1.1: Outer appearance of aphids in young, adult and winged stages.[3]

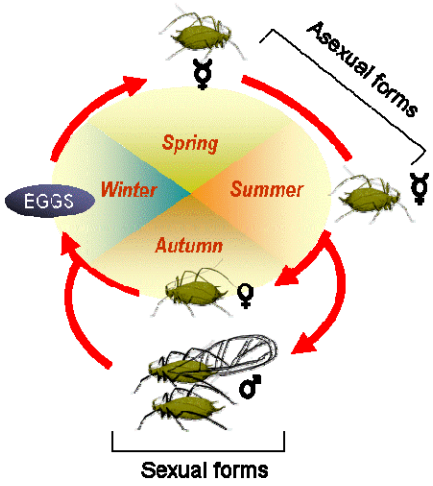


Figure 1.2: Life cycle of aphids.[5]

1.1.3 General Morphology of pea aphid

An adult pea aphid is around 2.0-4.0 mm long. The body is segmented with an elongated shape.[2] They also have a pair of antenna. The experiments reported below were performed with pink colored pea aphids (Figure 1.3) with black eyes, but there are other strains with different colors such as green and pale brown.

Figure 1.3: Live image of pea aphid taken under light microscope.

1.1.4 Feeding behavior of aphids

Aphids feed on plants in the genus *Leguminoseae.* This genus includes plants such as alfalfas, peas and clovers. Aphid probing not only leads to physical damages to plant tissues but also leads to lower crop yields as the aphids can consume significant amounts of phloem sap. In addition, aphids are responsible for transferring plant virus diseases. Currently 275 plant viruses are known to be spread via aphids.[6] However, there are some varieties of pea plants that show resistance to the aphids.

The sucking mouthpart of an aphid called the stylet is formed by a modified mandible and a maxilla. The stylet contains two channels, one for imbibing food and the other for salivation. The salivary canal is connected to a pair of salivary glands to secrete the saliva during feeding.[7] The aphid generally feeds from the phloem sieve elements in plants. To reach these sieve elements the aphid uses its elongated stylet as a probe.[7]

Phloem sap that is contained inside the phloem sieve tubes is the main nutrient of aphids. Phloem sap consists of several different types of carbohydrates, but mainly sucrose. Sucrose makes the phloem sap more viscous with high osmotic pressure. In addition to sucrose there are amino acids, proteins and other minerals. There are also low molecular weight proteins specific to the phloem.[8,9] Phloem sap flows under turgor pressure differences that are created between the source and sink ends of a sieve tube. Due to this internal turgor pressure, aphids can feed on the phloem sap passively. Experiments on

wheat aphids (*Rhopalosiphum padi*) have suggested that an average of 46 aphids can consume 10 μL of phloem sap within 6 h.[10]

Aphids excrete sticky droplets of honey dew that consists of highly concentrated sugar.[11] The amount of honeydew production can be an indication of whether the aphids are feeding well on a diet. In addition, there are some ant species that feed on honeydew and show mutual relationships with aphids such as protecting aphids from enemies and spreading aphid eggs to different locations.[11]

1.1.5 Impact on salivary secretion on host plant

There is evidence that the salivary proteins that aphids secrete into the host plants act as elicitors and play a major role in changing the plant physiology[1]. In general, plants react immediately to damaged phloem tissues by clogging the sieve elements to prevent the plant sap from flowing out. During feeding the aphids have the ability to prevent the sieve elements from clogging by remodeling the cell walls. Aphids can continuously withdraw the phloem sap for several hours without causing permanent physical damage. In other words, aphids secrete constituents that can act against the defensive interactions of the host plant which can otherwise stop the flowing out of the phloem sap.[10] Therefore, it is important to identify and analyze aphid salivary proteins in order to study the aphid-host plant interaction behavior.

1.2 Aphid salivary secretion

Aphid salivary secretion process can be broken down into three distinct phases. The first phase is the pathway phase where the aphid penetrates the stylet through the epidermis

and mesophyll tissues, until the stylet tip reaches a sieve element. The second phase is the phloem phase. In this phase, the aphid ejects saliva for external digestion and withdraws the phloem sap into the stylet food canal. In some cases, there can be a xylem phase where xylem sap ingestion is observed[12].

1.2.1 Aphid saliva

Aphids produce two different kinds of saliva, watery saliva and water insoluble stylet sheath or gelling saliva. Watery saliva contains soluble compounds that are discharged during brief intracellular probing in pathway phase, at the beginning of the phloem phase and during phloem feeding.[12] Watery saliva plays a major role in diluting the highly concentrated phloem sap as well as in external digestion. In contrast, gelling saliva is more viscous in nature and salivation occurs throughout the pathway phase. Gelling saliva creates a protective envelope for aphid stylet during probing. Hence, salivary sheath acts as lubricating tube to assist the stylet penetration through the plant tissues. This stylet sheath remains in the plant tissue even after the stylet is withdrawn[7].

Saliva is produced in the salivary glands connected to the stylet. Aphids have two different salivary glands, larger principle salivary glands and smaller accessory glands, each in pairs. The functions of these two glands are not yet precisely known. In general, the principle gland is mainly responsible for insoluble sheath salivary production while accessory glands involve watery saliva secretion. On the other hand it has been suggested that accessory glands are responsible in transferring viruses to the plant.[7]

1.2.2 Electrical Penetration Graph technique (EPG)

Aphid salivary secretion behavior has been recorded in terms of voltage fluctuations using an Electrical Penetration Graph (EPG) technique. In EPG, an electrical circuit is completed through an aphid, feeding on a plant, and the host plant via low DC voltage and high input resistor (1GΩ).[7] The resulting EPG signal enables a researcher to distinguish between intracellular and extracellular probing. The stylet penetration starts with the pathway phase which is due to extracellular probing shown in positive EPG signals (Figure 1.4). In addition occasionally cells are punctured during path phase and such events are indicated by sharp potential drops (negative signals) during the pathway phase. After a phloem sieve element is punctured, generally two types of signals are observed. These periodic waves are known as the E1 and E2 of phloem phases. In both the E1 and E2 phases, salivary injection occur into phloem sieve elements and appear as negative signals.[12,13] In the E1 phase the watery saliva is secreted into the sieve elements of the plant causing external digestion of phloem sap. In this phase, the ciberial valve of aphid is closed so that the food canal is not exposed to outside. Then the E2 phase begins by opening the ciberial valve. As a result, the phloem sap is ejected into the food canal and mixed with watery saliva in the food canal for internal digestion. E1 and E2 salivary phases are alternatively repeated in different time intervals throughout the feeding process.[12]

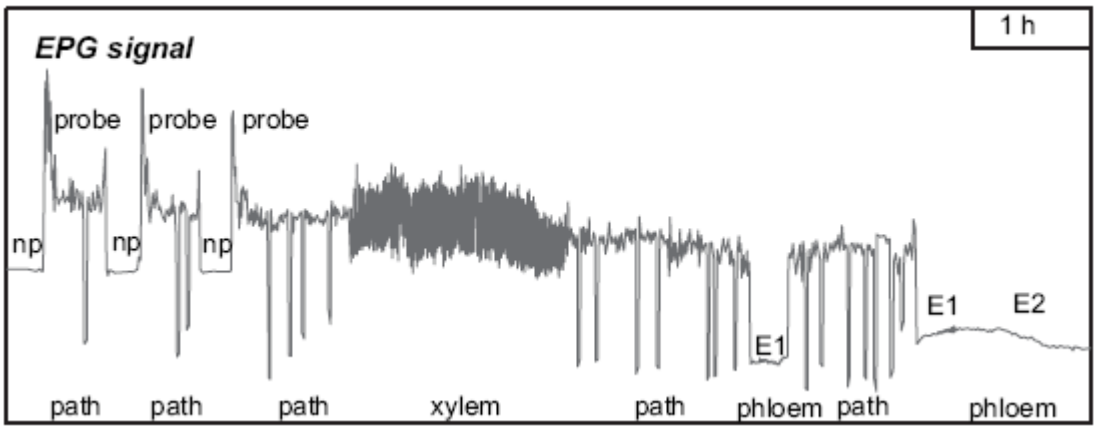


Figure 1.4: The EPG signal represented by an example of a 1 h recording.[12]

1.2.3 Salivary constituents

The watery saliva of aphids is hypothesized to contain many amino acids, proteins and carbohydrates. It is believed there are two types of proteins in the watery saliva: enzymatic proteins such as pectinases, cellulases, phenol oxidases and structural proteins. Previous studies have revealed that the composition of the saliva changes with different salivary phases as well as different diet compositions.[7]

1.2.4 Salivary proteins of other insects

Previous papers have reported the presence of several enzymatic proteins in different aphid species. Many Hemipterial insects that feed on plant sap contain oxidoreductases and hydrolases. These previous experiments, excised aphid heads were used to analyze salivary glands[14]. Studies on rose aphids (*Macrosiphum rosea*) have shown that salivary glands contain peroxidases and catechol oxidases while the saliva of cereal aphids (*Schizaphis aestivum*) contains polysaccharases such as pectinase and cellulase.[7] Recent

salivary gland analysis of the green peach aphid (*Myzus persicae*) identified enzymes such as glucose dehydrogenase, NADH dehydrogenase and glucose oxidase[1].As a whole, enzymatic studies reveal that different aphid species have different enzymes according to their host plant.

1.2.5 Effect of E_1 saliva on plant defense responses

Previously biochemical studies have shown that this watery saliva contains many families of enzymes and proteins that have not been properly identified, and that might influence aphid plant interactions. Since the saliva secreted in E1 phase is the one that is secreted only into the phloem tissues, it has been suggested that the E1 saliva constituents have the control to prevent the plant wound responses as well as interaction of phloem proteins.[7]

1.2.6 Calcium binding salivary proteins

One group of proteins believed to be secreted is calcium binding proteins. Calcium binding salivary proteins are believed to penetrate the phloem tissues and suppress calcium regulatory defensive mechanisms.[15] There is evidence that binding of salivary proteins with calcium in the phloem sieve tubes can control the plugging mechanism.[16]

1.3 Aphids in the laboratory environment

An average adult aphid can survive and reproduce for 2-3 days on an artificial diet in the laboratory environment. Several studies have shown that aphid stylets can penetrate through a membrane. Furthermore, aphids can puncture stretched Parafilm™ membranes and feed on the diet on the other side of the parafilm.[1] For this research, stretch Parafilm™ was used to collect aphid salivary secretions after the aphids had fed for 24-48 h.

Parafilm™ consists of paraffin wax, butadiene and polyethylene. It is semi transparent and colorless. It is versatile due to its nature of stretchability and self sealing. In addition, it is waterproof and permeable to air. Furthermore, Parafilm™ with thicknesses less than 0.013cm is permeable to moisture[17]. Because of these properties Parafilm™ is frequently used for wrapping and coverings especially in laboratory purposes.

1.3.1 Artificial Aphid diets

Many different artificial aphid diets have been reported. One diet, often referred to as the complete diet consist of 15% sucrose, 23 amino acids and amides, several vitamins, trace metals and other salts at pH ~ 7. The complete diet constituents and preparation protocol is shown in Appendix A. The diet is highly nutritious and aphids survive for 2-3 days feeding on it. However, since the complete diet is also a good growth medium for bacteria, it easily gets contaminated.

A second diet called the 3 amino acid diet or "diet 3" consists of 15% sucrose and three essential amino acids, 100 mM methionine, 150 mM aspartic acid and 100 mM serine. It has a neutral pH (pH ~ 7.5). This diet is less nutritious but not as likely to get bacterial contamination.

A third type of diet that can be fed to aphids consists of only 15% sucrose. This diet is sufficient enough to keep aphids alive for 2 days with no other nutrients such as amino acids and vitamins. In order to measure the amount of diet consumed by aphids an artificial coloring agent can be added. Successful consumption of the diet can be monitored because the aphids and their honeydew that they excrete will turn the color of the diet.

1.3.2 Other factor affected on aphid behavior

Aphids will live longer if they are exposed to proper light intensities and light durations. They are attracted especially to yellow light because yellow is an indication of highly nutritious young plants.[18] Good ventilation is necessary for aphid survival and growth. In addition, in a natural environment they feed on the phloem sap which flows under a pressure of about 15-30 atmospheres.[19] This phloem sap flowing under turgor pressure might be helpful for them to feed better, than feeding on stagnant diet.

1.4 Detection of salivary components in pea aphid- earlier studies

Earlier studies have suggested that pea aphid salivary components consist of different classes of enzymes such as oxidoreductases and hydrolases. The existence of oxidoreductases such as phenol oxidase is likely to occur since phloem sap may contain phytochemicals that are produced by the plant for used in defensive purposes. In addition, hydrolases such as celluloses, pectinases, sucrases may be available for phloem sap digestion.[8] Constituents in the saliva may change with the diet composition that they are feeding on. In addition to enzymes there are other salivary gland proteins that have been isolated using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE).[1,16] However, they have been not uniquely identified as of yet.

1.4.1 Limitations in previous studies of salivary secretion analysis

Watery saliva can be collected from the artificial diet upon which aphids are fed. However, identifying salivary proteins is not straight forward because of the difficulty in collecting a sufficient amount of proteins that can be detected by chemical analysis techniques. Collecting an aphid salivary protein sample without contamination with other bacterial or human proteins also is a challenge.

1.5 Aphid salivary collection

Specific Aphid protein collection has previously been performed by placing aphids on fresh host plants for certain period, removing the aphids and then extracting the leaf tissues or excised aphid heads.[15] However, this method is not applicable for general

analysis of aphid secretome. In order to collect the aphid saliva from an artificial diet, two methods have been reported, namely large scale sample collection method and small scale sample collection method.

Large scale sampling has been reported previously using various different aphid chambers[1,16]. In general, aphids are put into the chamber and the chamber is covered with stretched Parafilm^{TM} layers. Diet is added on top of a first layer and a second layer is stretched over the first to seal the media between the two layers of Parafilm^{TM}. Different sizes of chambers have been used that can hold different numbers of aphids, varying from 30 to 8000[1,16]. Aphids have been fed previously for 24 h or 48 h periods after which the diet/saliva sample was collected for analysis.

This method is useful to collect both watery saliva as well as some sheath saliva on inner Parafilm^{TM} layer. It eliminates the extraction step that is necessary when aphids are fed on plants. Concentrating the resulting saliva thus becomes simpler.

However, an artificial diet can affect the salivation phases of aphids so that there might be a difference in salivation both quantitatively and qualitatively. In addition, since a large volume of diet is added there is the need to concentrate the sample in order to isolate the protein species.

1.5.1 Small scale sample collection

Up to now, a small scale sample collection has not been reported using a few micro liters of diet given to feed the aphids. This can be demonstrated using a glass chip with an etched channel to add the diet. This method can be useful to collect highly concentrated saliva samples that do not require lengthy pre concentration steps. On the other hand, it is difficult to attract aphids to a small volume of diet and still collect a sufficient amount of saliva.

1.6 Sample filtering and concentration

The collected salivary protein sample needs to be purified and concentrated by eliminating other components from the fed diet such as sugar, amino acids and vitamins and by decreasing the overall volume of collected sample. This can be done by diluting the samples with acetonitrile/water or sodium phosphate buffer followed by ultrafiltration. Lower molecular weight compounds such as amino acids and sugar molecules in the diet are washed through the filter while higher molecular weight species are retained and concentrated.

1.6.1 Molecular weight cut off filters

One method of protein concentration is centrifuge ultrafiltration using molecular weight cut off filters. Generally molecular weight cut off filters consist of filter unit, cap and centrifuge tube (Figure 1.5).These filter membranes are usually made up of a polymer resin with a narrow distribution of pore sizes. When analyte (solute) is added, the molecules larger than the pore size will be retained, while molecules smaller than the

pore size will permeate through the membrane. The retained molecules are called the retentate while the molecules that pass through the filter are called filtrate.

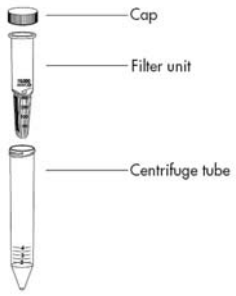


Figure 1.5: General diagram of a molecular weight cut off filter.

Selection of pore size of the membrane depends on the molecular size that needs to be concentrated. The minimum hydrodynamic volume (V_{min}) of a globular protein is related to molecular weight according to the equation (1.1).[20]

$$V_{\min} = \frac{M v_2}{N_A} \qquad (1.1)$$

Where, N_A = Avogadro number and v_2 = specific volume which changes with the hydrodynamic radius of the protein molecule. This indicates that molecular weight of a protein is proportional to Stokes volume. Based on this fact, molecular weight cut off filters (MWCO) can be used to retain proteins with defined molecular size. In MWCO filtration, the ultrafiltration occurs by rejecting 90% of the solute molecules based on the molecular weight.[21] For example 3000 MWCO filter membrane can retain 90% of proteins with a size that corresponds to an average molecular weight of 3000 Daltons, while filtering out the smaller solute molecules such as sugar, salts and amino acids.

Ideally, MWCO filter characterization works for spherical shape molecules that are soluble in water and mild buffer.

Protein ultrafiltration can be mathematically modeled using Kedem-Katchalsky formulation as follows.

$$J_v = L_p\left(\Delta P - \sigma_d \Delta\pi\right)^{22} \tag{1.2}$$

$$J_s = P\Delta C + J_v\left(1-\sigma_f\right)\overline{C}^{23} \tag{1.3}$$

where, L_p is the hydraulic permeability ($mPa^{-1}s^{-1}$); σ_d is the Osmotic reflection coefficient; J_s is the solute flux ($kgm^{-2}s^{-1}$); P is the solute permeability (ms^{-1}); ΔC is the concentration difference (kgm^{-3}) between two sides of membrane; σ_f is the solvent drag reflection coefficient; $\overline{C}$ is the average concentration in the layer (kgm^{-3}); $\Delta\pi$ is the osmotic pressure difference between two sides of membrane.

Kedem-Katchalsky formulation describes the pressure and concentration driving forces that affect on the solvent and solute flux during ultrafiltration. According to the equation, the solute flux mainly depends on hydraulic conductivity, osmotic reflection coefficient and solute permeability. Hydraulic permeability measures the flux due to the hydrostatic pressure differences between either sides of the membrane while the solute permeability measures solute flux induced by concentration difference.[21] The osmotic reflection coefficient expresses the contribution of membrane selectivity for the osmosis pressure difference. J_v is the convective flow, i.e. the flow of solutes coupled with the solvent molecules. 1- σ_f is known as rejection coefficient (R), which describes the boundary layer

mass transfer effects according to the film model.[24] R can be also defined as (equation (1.4)),

$$R = \frac{\left(1 - C_p\right)}{C_F} \tag{1.4}$$

where, C_p = solute concentration in permeate solution and C_F = solute concentration in feed stream. High C_p results in high rejection capabilities due to adsorption, deposition and solute/membrane interactions. In addition, C_p changes with degree of polarization at membrane boundary layer.[24]

During ultrafiltration, molecules are filtered through semi-permeable membrane due to osmotic pressure difference. External pressure is applied to generate an osmotic pressure difference between either sides of the membrane creating a water flux.[21] One effective method for vertical filtering through a membrane is centrifuge filtration in which centrifugal force generate a pressure towards membrane surface. The gravitational (g) force or the spin speed can be controlled according to the required filtering speed and the mechanical strength of membrane filters.

For protein ultrafiltration, polyethersulphone is commonly used, as it binds less protein than other types of membrane materials. Polyethersulphone is a rigid polymer with high dimensional stability.[25]

The pore size of a membrane can be controlled during the manufacturing process of the membrane filters. Required pore sizes can be adjusted during the solvent casting steps,

where the solvent is evaporated and during the etching process, where pores are opened up. In addition, the pore size can be changed by using different pore forming additives.

In reality, ultrafiltration using molecular weight cut off filters will not be uniform for all types of membranes and for all different macromolecules. In practice, the membrane consists of pore size in a distribution shown in Figure 1.6 which is a slight deviation of ideal membrane. The percentage rejection in the graph refers to the percentage number of retained molecules. The absolute rating is the largest pore size found in the membrane that has 100% rejection. Generally, the 'nominal' rating which was tested with a known size solute is 90%-95% under standard conditions. The nominal rating value can, however, vary according to the type of membrane[21].

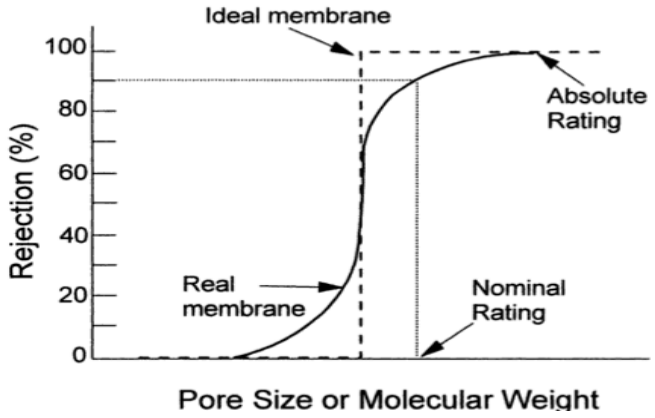


Figure 1.6: Relationship between pore size, molecular weight of ideal solutes and rating of ideal and real membranes.[21]

There are several factors that affect retention of these membranes. Essentially, the size of the target molecules needs to be larger than the pore size. Experimentally it has been

found that the best separation can be achieved if the target molecular size is approximately 10-fold larger than the rest of the molecules or hydrodynamic radius is 3 fold larger.[21] As a result, molecular cut off filters are ideal for separating proteins from solutes, but not for fractionation of different protein molecules with a narrow size distribution. Secondly, filtration depends on the shape of the molecules. Linear molecules are penetrated through the membrane more easily than globular molecules, so that the rejection profile will be deviated from the ideal for linear molecules. The shape of the molecules can also vary as a function of their interactions with other components, temperature, and ionic strength. Third, the type of membrane material affects solute rejection. Under conditions where high solute-membrane interactions occur there will be a deviation between observed and true rejection. These interactions depend on hydrophobicity, charge and surface roughness of membrane. Hydrophilic materials such as polysulfones have high flux and low adsorption, resulting in less deviation in rejection profile. Fourth, the type of solutes affects the filtration. Small solutes such as sugar or salts have high permeability and less interactions, so that they have almost zero rejection compared to larger solutes. There are other operating parameters that influence the filtration such as solution concentration, temperature, pH, turbulence near membrane, ionic strength, polarization and mass transfer effects. Polarization and mass transfer effects occur due to the concentration difference near the membrane than the bulk solution under high pressure conditions. Apart from these factors there is lot-to-lot variability of pore size and membrane composition in different manufacturing materials.[21]

There are several advantages to using molecular cut off filters for protein concentration. It is a fast and simple method of concentration. Generally, up to 300-fold concentration can be achieved within few hours. However, the filtrate time may vary with the type of filtering tubes.[26] According to the size of target molecules, filtering membranes with different pore sizes can be selected, so that the solute loss can be minimized, while optimizing other factors, such as filtrate time and final volume. During centrifugation, temperature can be lowered to minimize microbial contaminations and to lower the denaturation rate.

There are a few drawbacks to molecular cut filtration method using MWCO filter tubes such as fouling and adsorption. At high C_p values solute molecules may get trapped on the membrane surface and reduce the flow rate. This interaction of solute with the membrane surface is known as the fouling effect. Binding of solutes depends upon the nature of solute molecules and the membrane as well as the time duration for centrifuge filtration. Use of low molecular weight sugar can reduce the flux effects. In addition water flux can be increased by modifying the membrane. When hydrophilic groups such as carboxylic and sulfonic groups are added fouling can be reduced; e.g. polysulfone membrane[27]. On the other hand, high hydrophobicity broadens the pore size distribution and improves the percentage rejection; e.g. cellulose acetate and regenerated cellulose membranes. Another disadvantage is that, the proteins at their isoelectric point become unstable and generate low fluxes. Therefore, performance may change, depending on the solution conditions. Furthermore, ultrafiltration generally does not result in a dried retentate, so that a secondary step is required, such as lyophilization. Finally, the filter

tubes are not suitable for storing at low temperature for long hours as the membrane can be disrupted.[21]

"Vivaspin" ultra-filtration spin tubes are one of the types of concentrators that were used during these particular experiments reported below. The membrane of the filter in Vivaspin tube is made out of polyethersulphone, a low binding polymer with no external wetting agents. The active area of the membrane is approximately 4 cm^2.[26] The volume of the tube indicates in the "Vivaspin number"; e.g. Vivaspin 15 tubes consist of 15 cm^2 volume. Aphid salivary proteins are estimated to be in the molecular weight range from 6 kDa to more than 700 kDa[9], which is compatible with 3000 and 5000 MWCO Vivaspin tubes.[26]

1.7 Analysis methods of aphid salivary proteins

1.7.1 SDS-PAGE (Sodium Dodecyle Sulfate Polyacrylamide Gel Electrophoresis)

Electrophoresis is a movement of dispersed particles under an applied electric field. The relative movement of a particle occurs as a result of several forces acting upon it. A charged dispersed particle feels a Coulomb force (equation (1.5)), an opposing electrostatic force due to the oppositely charged diffused double layer. In addition, a moving particle feels a frictional force, operating in opposition to electrostatic force.[28]In gel electrophoresis, the frictional force is a prominent factor that depends on gel pore size and particle size.[29] In the simplest case, the frictional force (f) can be defined in Stoke's law as equation (1.6),

$$f = QE \tag{1.5}$$

$$f = 6r\pi\eta v \tag{1.6}$$

r = Particle radius; η = viscosity of solution; v = velocity; E = applied field strength and Q = charge of the particle

At equilibrium, the coulomb and frictional forces are balanced so that the velocity of a particle can be found by equating equation (1.5) and (1.6) solving for v (equation (1.7)).

$$v = Q\,E / 6\pi\eta r \tag{1.7}$$

The term $Q / 6\pi\eta r$ is constant under a given set of buffer conditions. These parameters are generally represented by a simple term called the electrophoretic mobility (μ) (equation (1.8)).

$$\mu = Q / 6\pi\eta r \tag{1.8}$$

From equation (1.8), it is clear that the migration distance of the particle is proportional to the size and the charge. Therefore the electrophoretic migration distance can be measured to determine the relative size of a particle in a mixture that contains same charge.[29]

Macromolecules such as proteins are separated commonly using gel electrophoresis. In gel electrophoresis a polymer matrix is used to separate proteins according to their migration rates through the gel matrix pores. Polyacrylamide gel is commonly used in protein separations because it is a mechanically strong and chemically inert polymer

matrix. However, proteins in their native structure may not migrate through the polyacrylamide gel due to their large size. In addition, proteins of similar sizes will migrate at different rates due to their different surface charges. In order to generate proteins separations based upon size only, the proteins must all have the same charge density. To achieve this, proteins are dissolved in a detergent such as sodium dodecyl sulfate (SDS). SDS can denature the proteins, and cover the protein as a uniformly negatively charged coating.[28] Therefore sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) can be used to separate proteins according to their sizes regardless of the shape or the charges. Protein size can be estimated by the migration distance compared to distance traveled by molecular weight markers. The separated protein bands can be visualized by staining the gel with a dye (e.g. Coomassie blue), using silver staining or using fluorescent active analyte molecules. In two dimensional gel electrophoresis, two electric fields are applied in 90 degree angle of the first field. This method results in a better separation of a complex mixture of proteins. This method can achieve higher separation efficiencies because it separates proteins based upon two different molecular properties such as size and isoelectric point. However, during SDS PAGE, proteins get denatured and loose their secondary structures.[28]

1.7.2 LC/MS detection

HPLC can be coupled to mass spectrometer for protein separation and identification.

1.7.2.1. High Performance Liquid Chromatography (HPLC) – Introduction

Chromatography is a method of separation of a mixture of components that partition between two phases. The two phases are the immobile stationary phase and mobile phase. In HPLC the stationary phase is a solid or liquid while the mobile phase is a liquid. When the mobile phase with analytes flows through the stationary phase, the analyte components alternatively interact with stationary phase (sorption) and move back to mobile phase (desorption) for multiple times, resulting partitioning between stationary and mobile phase. The time required for sorption-desorption process throughout the column affect the retention time of a component. Different analyte components possess different degree of partitioning according to their interactions with the stationary and mobile phases resulting different retention times.[30] Therefore in HPLC, analyte components are separated based on their degree of interactions that causes different affinities to the stationary phase throughout the chromatographic system. The separated components according to their different elution times are illustrated as chromatographic peaks in the chromatogram (Figure 1.7).

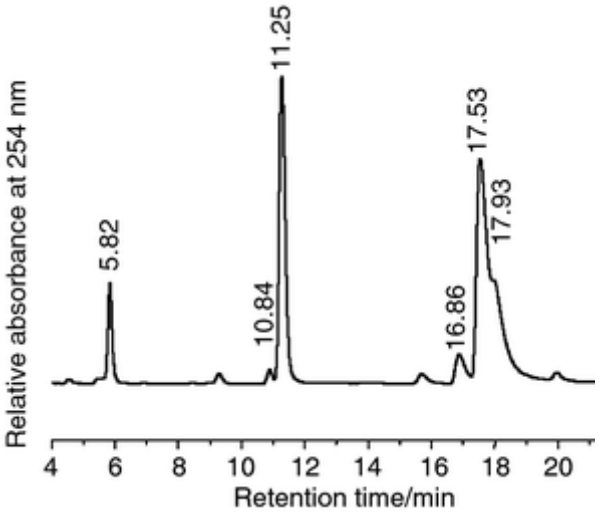


Figure 1.7: An Example of a typical chromatogram.[31]

Peak broadening as a kinetic phenomenon is expressed in van Deemter equation. According to the van Deemter equation (equation (1.9)) the plate height depends on three main constants namely, eddy diffusion (A term), longitudinal molecular diffusion (B term) and mass transfer in mobile and stationary phases (C term) as shown in following equation.[30]

$$H = A + B/u + Cu \qquad (1.9)$$

Where, u = average linear velocity of mobile phase. Conditions should be optimized to achieve fast separation with minimum band broadening in a chromatographic separation. The optimum conditions can be obtained in so many different modes such as selecting the appropriate stationary and mobile phases, changing the solvent gradient, adjusting the flow rate and adding ion pairing agents depending on the analyte sample.

There are factors that need to be considered such as the plate height, capacity factor, resolution and peak capacity, in order to achieve efficient separation with well resolved peaks.

1.7.2.2 Factors affecting HPLC separations

Capacity factor (k') is a term that is used for the measurement of analyte retention relative to the retention of the solvent which has zero retention.[30] In terms of retention time,

$$k' = \frac{t_r - t_o}{t_o} \qquad (1.10)$$

Where t_r = time spent by the solute molecules in the column and t_o = time spent by the solvent molecules in the column.[32] Since (t_r- t_o) = time spent by the analyte in the stationary phase, k' is also define as the time spent by the analyte in the stationary phase / time spent by the analyte in the mobile phase. Alternatively,

$$k' = \left(\frac{X_s . V_s}{X_m . V_m} \right) \qquad (1.11)$$

where, X_s / X_m = K (Equilibrium Distribution Coefficient of solutes in stationary and mobile phase) and V_s/ V_m = volume ratio of stationary and mobile phase that also known as the phase ratio of the separation column (equation 1.11).[30] Moreover, the ratio of capacity factors of adjacent peaks is known as selective factor α. Having a high α value indicates good separation of the two peaks.

The efficiency of the column is expressed in number of theoretical plates (N). N is high if the peaks have less peak broadening. It can be expressed in terms of peak width (σ_r) and retention time (t_r) as shown in equation (1.12),

$$N = \frac{(t_r)^2}{(\sigma_r)^2} \tag{1.12}$$

As the retention time is higher along the separation the column the peak width increases. Therefore N is a measure of relative peak broadening as the analyte pass through the column. N is also defined as (1.13),

$$N = \frac{(H.L)}{d_p} \tag{1.13}$$

H is the theoretical plate height, L is the Column length and d_p is the particle diameter of stationary phase.

The Equation (1.13) shows that N changes according to the column length. However the term H is independent of the column. The Plate height is the height of a theoretical plate that is proportional to peak variance. Lower the plate height higher the separation efficiency. In general, the N term together with h term express the goodness or the quality of separation of a chromatographic system. On the other hand, the equation (1.13)

indicates that column parameters such as length and particle diameter of stationary phase have a direct relationship with the separation efficiency.[32]

The resolution (R_s) is defined as degree separation of two adjacent peaks. . Larger the resolution better the separation. In general, resolution affects the peak width and distance between the two peaks according to the equation. Resolution is measured in a chromatogram according to the equation 1.14.[30]

$$R_s = \frac{2d}{\{(W_b)_B + (W_b)_A\}} \tag{1.14}$$

Where, W_b is the peak width at the base for A and B adjacent peaks and d is distance between the two peaks. However, resolution is affected by peak tailing and therefore equation 1.12 is not useful for peaks with different peak heights.

By considering all the factors affecting, Rs is expressed as,

$$R_s = \tfrac{1}{4}\sqrt{N}(\alpha - 1)\left[\frac{1}{(1+k')}\right] \tag{1.15}$$

The equation (1.15) shows that the resolution depend three factors such as peak width, selectivity and relative retention. Resolution can be improved by raising the selectivity, maintaining the capacity factor (k') in the range of 1-10 or by increasing N by using smaller particle sizes in the stationary phase.[32]

1.7.2.3 Column parameters

To improve chromatographic separations, several parameters (variables) can be adjusted. The main variables of a column are column length, diameter, stationary phase composition and packing.[32]

Conventional column lengths are around 3-25 cm, with 4.6 mm internal diameters (i.d.). These columns are packed with particles 3-10 μm in diameter.[30] The flow rate of conventional columns is around 1 mL/min. Currently, shorter columns are more commonly used as they can generate fast separations and consume lower volume of mobile phase. Microbore columns have smaller i.d. varying from 0.2-1 mm and provide high detection sensitivity due to smaller elution volumes and flow rates. The flow rate throughout these columns range from 1-20 μL/min.[30] In addition, there are nanobore columns with 25-100 μm i.d. having 25-4000 nL/min range of flow rates.[33] Since injection volumes are very small on nanobore columns may lead to external peak broadening and therefore sample pre concentration may be needed.

Stationary phases are generally attached to particles packed in HPLC column. In reverse phase liquid chromatography, mobile phase is polar and the stationary phase consists of non polar molecules attached to the particles. The column is usually fabricated with silica based sorbent with modified silanol groups. One limitation of silica column is ionization of silanol groups in low pH conditions. Ionized silanol groups strongly interact with

solutes to generate long separation times.[30] This can be sometimes avoided by end capping.

Porous packing materials are used for complex mixtures of biomolecular separations to obtain high efficiency and to increase the loading capacity. The solute molecules must be less than one tenth of the pore diameter in order to avoid restricted diffusion and to get a high accessible surface area. Column packing should be done by considering both factors of good separation efficiencies and high column loadability. For example, thinner columns packed with porous shell particles can generate high separation efficiency due to faster mass transfer but with reduced loadability.[32] In contrast when poroshell particles are used the column loadability is improved while maintaining the high efficiency.[34]

Commonly used non polar stationary phases are hydrocarbon alkyl chains with different chain lengths such as C_4, C_8 and C_{18}.[32] The longer the alkyl chain the higher the retention of hydrophobic solutes.[30] Non polar stationary phases can retain proteins with hydrophobic regions followed by a separation according to the degree of polarity.

The mobile phase in chromatographic column consists of one or more solvents. The composition of mobile phase can be kept constant or changed throughout the separation, depending on the analyte sample. If the elution is isocratic there is no variation in solvent composition throughout the separation, whereas in gradient elution the mobile phase percentage composition is changed gradually. In gradient elution, solvent designated as A

is usually water (aquase solvent) and B is an organic solvent such as acetonitrile or methanol. Percentage of B is changed to generate the gradient.[30]

1.7.2.4 Protein separations in HPLC

HPLC is widely used for the separation and purification of proteins because it can separate complex mixtures quickly with a high resolution. Proteins are three dimensional structures with unique orientations specific to each species. The protein surface consists of different polarity regions with different affinity for certain solvents. On the other hand, interactions of large proteins with a stationary phase are restricted to certain regions unlike small peptides.[32] These specific interactions are useful in optimizing the separation conditions. Protein samples are often separated by reverse phase liquid chromatography (RPLC), ion exchange chromatography and affinity chromatography.

The affinity chromatographic column consists of immobilized ligands bonded to the solid stationary phase that can interact and make stable complexes specifically with certain biomolecules. These interactions are highly specific the ligands in the affinity chromatographic column can separate specific molecules or group of molecules from rest of the non bonding species.[30] Therefore, affinity chromatography is more specifically useful during protein purification rather than separation since it selectively binds a specific protein out of a mixture.

In ion exchange chromatography, the stationary phase is immobilized with a charged resin (R) that binds either to opposite charged anion or cation. The resin can selectively

exchange ions or polar molecules from the aquase mobile phase by formation of weak ionic forces. General reaction of cation exchange resin can be expressed as equation (1.16).[30]

$$R^-A^+ + B^+ = R^-B^+ + A^+ \qquad (1.16)$$

Ion exchange chromatography is effective to separate proteins with surface charges. However it has limited applications on proteins that contain similar surface charges. Out of all, RPLC is the most effective and versatile separation method since it interacts with proteins according to the protein's hydrophobicity and enables to change the degree of interactions by changing the elution conditions. Hence, RPLC is very useful for separating mixtures that contain closely related proteins as well as widely different molecules in structure.[32]

Column parameters and conditions can be varied according to the type of protein mixtures that are analyzed. Usually for small proteins, long columns (15-25 cm) are used to generate high resolution. In contrast, for larger proteins 2-20 cm long columns are used to avoid low mass recovery due to irreversible binding with the stationary phase.[31] The choice of the internal diameter of the column is based on the sample capacity. Narrow bore columns are widely used because small volumes can be injected and because they generate high separation efficiency. However nano bore columns have limited stationary phase material and therefore consists of low loading capacity which also reduce the separation rate. In addition, Nano LC is difficult to operate in practice due to sample

clogging and less reproducibility. Reducing column i.d. also affects extra column band spreading due to high permeability. Microbore columns with 1-2 mm i.d., on the other hand, are more convenient and can be operated at 50-250 μL/min flow rates.[32] Micro bore columns are more suitable in protein separation because fast separations are important in protein analysis to avoid loss of biological activity and irreversible denaturation. Poroshell columns are more effective and fast in separating proteins as they have high mass transfer kinetic properties, high capacity and improved reproducibility than totally porous materials.[34] Small particle sizes (3-5 μm) with pore size ~300 ˚A to achieve short diffusion time that also influences fast separations.[32]

Selection of n-alkyl chain lengths in the stationary phase depends on the polarity of the protein sample. The longer the chain length the more the hydrophobic proteins are retained while highly polar peptides and amino acids elute fast. If the hydrophobicity is too high such as in C_{18} column (e.g.octadecysilyl) mass overloading and protein denaturation can occur whereas shorter alkyl columns (e.g. C_4 columns such as butylsilyl) can result in low resolution.[32]

By manipulating the mobile phase conditions fast separation can be achieved maintaining good peak capacity. Generally acidic mobile phase is favorable in protein elution since hydrophobic interactions between proteins and silanol groups are weakened by acids. Triflouroacetic acid (TFA) is commonly used in the mobile phase as a volatile acid but alternatively phosphoric acid, perchloric acid, formic acid, acetic acid and hydrochloric acid are also used. Acids can also act as a proton donors as well as ion pairing agents.[30]

Ion pairing agents can increase the hydrophobic interactions of highly hydrophobic molecules such as proteins with the non polar stationary phase and therefore improve the selectivity.

In order to achieve a good resolution, gradient elution is more often used in protein separation since proteins have different polarities. During the gradient elution, the non polar, organic component of the mobile phase is gradually increased. Since proteins have multi-site interactions, organic solvents can generate conformational changes of proteins that enhance selective elution from stationary phase.[32] The longer the gradient time the better the separation. However, organic solvents can increase the degree of denaturation of proteins.

The chromatographic systems for protein separations are usually generated under ambient temperature conditions to minimize denaturation or conformational changes. However some conformational changes at high temperatures can generate better separations.[32] In addition, as the solvent viscosity decreases at high temperatures. This generates faster separations.

UV detection is the most commonly used detection method for protein separations. At 210 nm and 220 nm absorbance wavelengths are specific for peptide bonds while at 280 nm wavelength the aromatic amino acids tryptophan and tyrosine are detectable.[32]

As a whole, RPLC is commonly used in separating protein mixtures because separation conditions can be easily changed to enhance selectivity in addition to achieve high recovery and reproducible results. However, this method is not suitable for proteins for which bioactivity must be conserved.

1.7.3 Size exclusion chromatography (SEC)

In order to determine the size of proteins in a sample, SEC can be useful as it separates the sample molecules according to the size. Unlike HPLC SEC is based on physical sieving rather than chemical interactions. Typically porous materials are used for the stationary phase with different pore sizes depending on the sample. Larger molecules will be excluded from pores and elute from the column earlier. This is known as total exclusion. The smallest molecules can penetrate through all pore sizes which known as total permeation. The molecular sizes that falls in between total exclusion and total permeation are partially retained in the stationary phase and partition according to the degree of penetration. The retention volume (V_R) for an analyte can be illustrated according to the equation (1.17).

$$V_R = V_M + KV_S \tag{1.17}$$

V_M = mobile phase volume, V_S = volume of the stationary phase pores and K = partition coefficient. The higher the K, the higher the penetration as well as the higher the retention volume. The retention volume is analogous to retention time of RPLC chromatogram. Therefore SEC is able to achieve peak separation according to different

molecular sizes that contain different retention volumes. In protein analysis, SEC helps to separate the smaller molecular amino acid peaks that have higher retention time than that of protein peaks. In addition molecular weights of the separated protein molecules can be estimated running biomarker protein mixture separation under the same conditions.[30]

1.7.4 Mass spectrometry

Mass spectrometry is an analytical technique that can be used for determination of elemental compositions and chemical structure elucidations. In mass spectrometry the analyte molecules are ionized using various ionization methods to generate charged molecules or molecular fragments and separate the ions according to their mass to charge ratio (*m/z*). The fragmented ions are then subjected to move through electro-magnetic field so that each ionic particle is achieved a certain kinetic energy. The separation is achieved in the mass analyzer according to the variation in ionic trajectory of fragmented ions that traveled through the electro-magnetic field which proportionate to the mass to charge ratio.[35]

Mass spectrometry is also a common method for identifying and characterizing protein samples. Mass spectrometers can be linked to HPLC column for direct analysis known as LC/MS. In order to perform LC/MS, apparent flow rate of HPLC and the mobile phase composition should be compatible with MS. For example low flow rate is required for MS with electrospray ionization because the liquid that flowing into the capillary that is undergo nebulization can handle only a low volume. Usually nano and microbore HPLC columns with low flow rates are compatible with MS. Mobile phase composition can be

modulated by increasing the volatility of buffer and changing the pH. Reversed phase LC is relatively compatible with the volatile components of acetonitrile or methanol. On the other hand, ion paring acids such as TFA in RPLC column reduces the MS detectability. This can be avoided by adding propionic acid and 2-propanol in 3 to 1 ratio to break the ion pairs before reaching the mass spectrometer.[30] In addition the analyte should be non-volatile during ionization step of mass spectrometry.[30]

Separated protein samples need to be ionized before entering mass spectrometric analyzer. Ionization is commonly done for whole proteins using electrospray ionization (ESI) or matrix assisted laser desorption ionization (MALDI).

Electrospray ionization can directly be coupled with LC at low flow rates (0.5-2 μL) or otherwise eluent splitting and nebulization are required. In ESI, an electric field is generated at capillary tip (Figure 1.7). The liquid analyte sample with volatile solvent is pushed through the capillary. A high voltage field (~3-5 kV relative to the ground) is maintained at the tip of the capillary.[36] When the liquid analyte reaches the tip, it gets sprayed as an ion solution under the high electric field while N_2 gas is used to nebulize the liquid. The solvent molecules evaporate fast and at the same time the charged sample ions repel each other to get tiny droplets.[36] Most commonly ionization occurs by protonation (M+H) or deprotonation (M-H). Multiply charged ions are commonly produced by ESI.

$$M + e \longrightarrow M^{(n-1).+} + ne$$

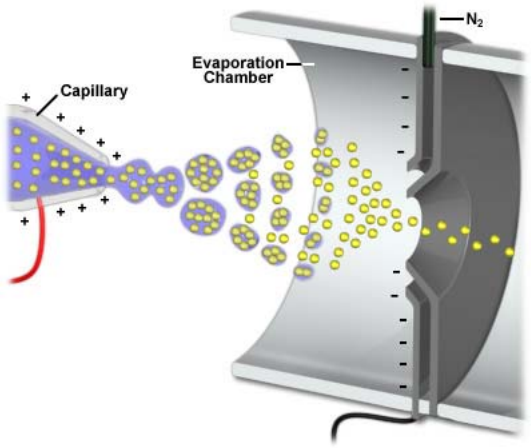


Figure 1.8: Schematic representation of main components of Electrospray ion source.[37]

It is a suitable ionization method for identification of molecular ions and one can obtain useful structural information from fragmentation. In addition, it has high mass range, upto 10^5 Daltons.[30] Therefore, it is ideal for analyzing macromolecules such as proteins. Ideally, the response is linearly proportional to analyte concentration. Linear relationship is however, can be achieved by optimizing the sensitivity and reducing the background.[36] The Detection limit for ESI is within the range of 0.1-1 pmol/µL.[41]

The resolution of ESI mainly depends on operating pressure, typically within the range of 10^{-4}-10^{-7} torr.[35] The lower the pressure, the longer the mean free path of electrons that leads to high resolution. Mass resolution is the ability of the analyzer to separate adjacent ions. The higher the resolution, the higher the selectivity can be achieved. This depends on type of analyzer used.[36] A quadrapole analyzer is commonly used with ESI to get high selectivity. In reality however, ESI has limited sensitivity and limited transfer of ions to

MS for detection.[36,41] Due to low fragmentation, limited structural information can be obtained. In addition ionization is difficult in large non volatile molecules.

MALDI is another soft ionization technique that can generate high molecular weight ions. The analyte sample is co-crystallized with a matrix solution that contains a matrix compound (e.g. 3,5-dimethoxy-4-hydroxycinnamic acid and α-cyano-4-hydroxy cinnamic acid) dissolved in organic solvent (e.g. acetonitrile). The co-crystallized MALDI spot is subjected to a focused laser beam for ionization (Figure 1.8). Since the analyte is within the matrix, there is less chance for degradation. Resulting ions are mainly with a singly charged, but a few multiply charged ions may also produced.[41] MALDI-laser spot size is in the range of 100-500 μm which mainly affects the optimal MALDI performance. Sensitivity is measured in terms of the slope of peak intensity vs. moles of sample. Sensitivity is reduced when the sample is not clean and contains high surfactant or salt concentrations. Less fragmentation is generated using in MALDI and it is easier to identify molecular ions.[41] Time of flight (TOF) is the commonly used analyzer with MALDI for biological samples, which results in high resolution and fast scans. In TOF, the time taken for ions to fly from source to detector under electric field is calculated accurately.[40] Typical flight time ranges between 5-100 μs. Flight time vary with factors such as the initial kinetic energy and location of ions in space. The longer the ion paths the higher the resolution. Generally, a tube with a length of 1m, can obtain a resolving power of 1000. However, there is a practical limitation in using long tubes. The *m/z* ratio is an intrinsic property of an analyte ion that varies with flight time. Therefore

TOF analyzer generates highly accurate separations. The resolution can be improved by optimizing the sample preparation and the MALDI spot size.[42]

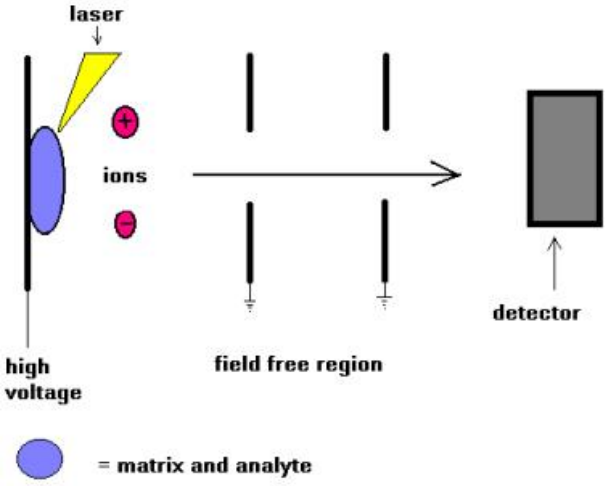


Figure 1.9: Simplified representation of MALDI-TOF arrangement.[38]

ESI is effective in identifying hydrophobic proteins, whereas MALDI is fine for basic peptides and aromatic residues. When LC-MS/MS is performed MALDI ionization is ideal, since it has limited molecular disruptions. However, ESI is ideal for sample separation by RPLC where MS/MS is not required.[42] As a whole, both these ionization techniques give limited information for whole protein samples, especially when analyte concentration is low in proportion to the other contaminants leading to low signal: noise ratio. On the other hand, data interpretation is difficult, when analyzing complex spectrums of mixtures of proteins.[43]

Better results can be achieved by protein degradation followed by peptide mass finger printing of protein fragments. The tryptic digestion is commonly used for protein

degradation while Tandem mass spectrometry is commonly used for proteins fragmentation analysis.[43]

In Tandem Mass Spectrometry or MS/MS includes multiple steps of separation of ions according to *m/z*. Typically, the first separation of precursor ions are selectively trapped and transmitted to collision cell where the precursor ions undergo fragmentation. In the collision cell, the precursor ions undergo collisions with an inert gas. During the collision, the kinetic energy of the gas molecules is converted to vibrational energy of analyte molecules which eventually releases by fragmentation. The resulting fragments with charged ions are distinguished according to *m/z*. Either quadrapole or TOF analyzers could be used for this. There can be multiple trapping steps according to the requirements. This improves the sensitivity and selectivity of a complex molecule sample analysis. During protein analysis peptide chain can be cleaved in different locations in the fragmentation process. The resulting peptide sequence tags can be used for peptide sequencing with the use of protein databases.[44] This is performed by tagging the peptide fragments with specific notations. Typically, backbone peptide ions are tagged with a, b, c or x, y, z, depending on the fact that the charged terminal is N terminal or C terminal respectively (Figure 1.9).[44] Out of all, b and y cleavages are the most common charge directed cleavages. In addition, there are other possible neutral fragments with loss of water or ammonia. Usually, when trpysin digestion is performed, the resulting fragment ions are charged more than one. The mass difference of two adjacent mass peaks (eg. b_n and b_{n+1}) should be equal to the corresponding residual mass (b_{n+1}). Likewise peptide sequence can be revealed by comparing residues with the mass peak differences both for

C terminal and N terminal residues.[45] However, it is complicated to manually identify peptide sequences from mass spectral data of a complex protein mixture. Therefore peptide fingerprinting can be accomplished using peptide sequence database algorithms.

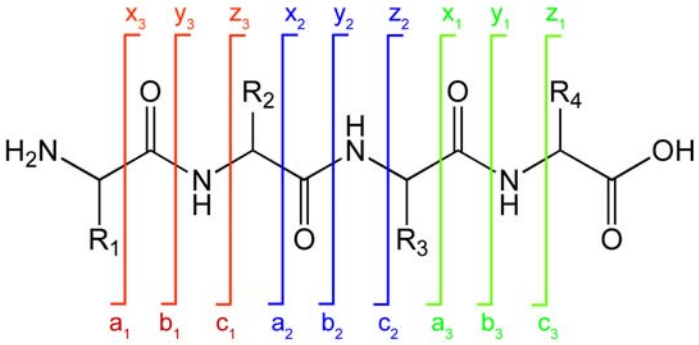


Figure 1.10: Peptide fragmentation notation.[44]

1.7.5 Peptide fingerprinting and database search

One approach for the protein identification is peptide fingerprinting sequence analysis through library database searching. MS/MS spectral data can be correlated with the theoretical fragmentation patterns to locate similar peptide sequence from the known peptides.[45] General process of tandem mass spectrometry database searching can be illustrated as Figure 1.12.

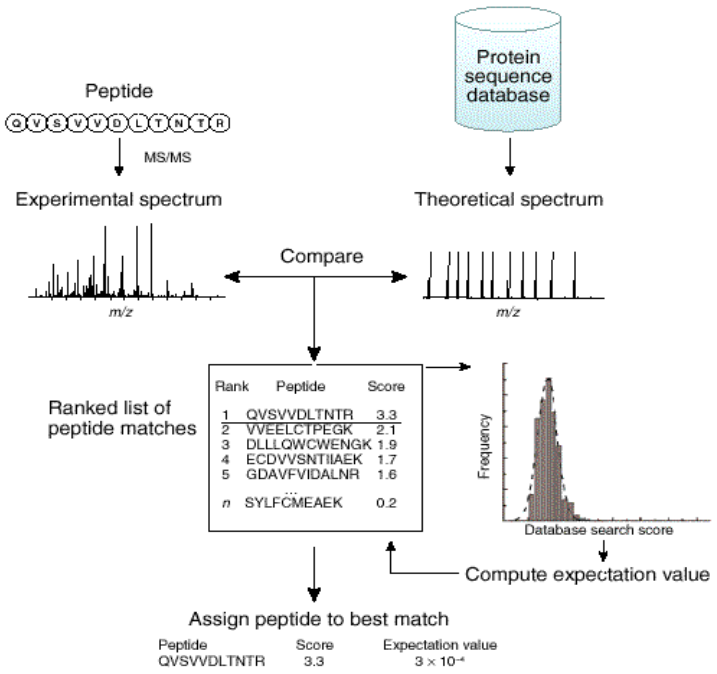


Figure 1.11: Peptide identification by MS/MS database searching.[46]

There are several database searching algorithms or computer software are available such as Mascot. Mascot algorithm can search for possible peptide sequences of different cleavage methods that match with the experimental mass values. In practice, it is highly unlikely to achieve all the peptide mass values coincide with the library databases. If the exact peptide sequence is not found in the library database, the closest peptide sequence matching can be obtained.[47] Statistical validation is necessary to select the highest correlated peptide matching out of all the possible matching sequences generated from the program. The statistical probability of the each random peptide sequence match is given as a percentage score using a scoring algorithm. If the database contains large number of entries matches with the experimental data, the significant probability (P) becomes low and therefore a high score can be obtained, indicating a good peptide match (Equation (1.18)). Usually a good peptide match has a score of 70 or above at $P < 0.05$.[47]

$$\text{Mascot score} = -10 \log_{10} (P) \qquad (1.18)$$

During fragmentation process, proteins can undergo post translational modifications such as phosphorelation, methylation etc. The post translational modification of a peptide occurs only in one or few amino acids of the peptide fragments that cause slight deviations in mass of the fragments. Mascot algorithm can be used to interpret the modified peptide sequences, by comparing with the databases with possible modifications in amino acid residues.[48]

Mascot also can display the high scored peptide matched within the amino acid sequence and determine the percentage coverage. Percentage coverage is the percentage availability of matching peptide fragments in the amino acid sequence of the possible protein.

The basic local alignment search tool (BLAST) program is an algorithm that contains a library database with a known amino acid or genetic sequence of a protein and allows matching the known peptide sequences with the target peptide sequences. By following BLAST procedure, set of amino acid or nucleotide sequences that consists of significant regions to the target peptide sequence can be obtained. It also can provide functional information that matches with the target peptide sequence. There are number of protein databases that can be used for the peptide sequencing, such as National Center for Biotechnology Information (NCBI) and expressed sequence tag (EST). BLAST algorithm

also can provide functional information that matches with the target protein and therefore can identify desired genetic family of the target protein with similar functional characterizations. However, protein database search can be helpful only if the genome code of desired species is already identified and has been fed to the database.

1.8 Summary

In summary, aphid salivary protein analysis required multiple steps such as, feeding, sample collection, pre-concentration, separation and identification. Earlier research work indicates that aphid saliva can be collected by feeding the aphids on an artificial diet. There are special techniques and conditions that need to be followed to obtain reproducible results for analysis process. Miniaturized devices such as micro chip and nano LC can be useful in enhancing the sensitivity of minute samples of aphid saliva with development of optimized conditions. Alternatively, large number of aphid feeding samples collection and pre-concentration can be used for preliminary studies. Protein identification can be performed by fractionate samples obtained from LC/MS followed by proteomics analysis. The information and techniques discussed above will be used in the following chapters to develop methods in order to effectively collect samples and identify the components of the aphid saliva.

CHAPTER 2: Small scale sample collection

2.1 Introduction

The aphid saliva samples were collected for analysis mainly in two different methods, small scale sample collection using microfluidic devices and large scale sample collection using screened chambers. The chapter 2 describes mainly the small scale salivary sample collection methods and data analysis. In this chapter microfluidic devices were used to optimize aphid feeding and to investigate the amounts of fluids conserved by aphids over a 48-72 h period of time. In chapter 3, large scale sample collection methods will be discussed.

2.2 Materials and methods

2.2.1 The aphid colony

The original pea aphid colony was grown on alfalfa plants (Dr. Marina Caillaud; Cornell University) and transferred to Kansas State University to Dr. Reese's lab. Before taken for experiments, they were grown on fresh fava bean *(Vivia fabe*) 10 cm height plants, under sodium yellow lights for 16 h day.

2.2.2 Artificial diet

Two different artificial diets were used: complete diet and 3 amino acid diet (diet 3). Complete diet consists of 22 amino acid, 15% sucrose, vitamins (ascorbic acid, thiamine

etc) and metal salts. Appendix A describes the compositions in detail and the protocol of preparation. Diet 3 contains 100 mM serine, 100 mM methionine, 100 mM ascorbic acid and 15% sucrose by weight. The solution was adjusted to pH 7.5 by addition 0.75 M potassium hydroxide (KOH).

2.2.3 Glass chip microfluidic device

In order to hold the artificial diet in a confined volume, a glass chip was etched channel was developed using photolithography. The glass chip for the microfluidic device was constructed using a glass slide with photoresist fabricated on one side of the glass. Magnetic stirrers, stirring rods and clamps were used to dip the glass slides in different solutions during the etching process. There were several solutions used for the photolithography process. Those are, potassium hydroxide as the photo resists developer, ceric sulfate as the chrome mask etchant and ammonium fluoride and hydrofluoric acid (NH_4F/HF) with hydrochloric acid as the buffered oxide etch solution. In addition acetone, ethanol and distilled water were used as wash solutions.

Initially, the photolithography to produce the chip with channels for the aphid feeding experiment was carried out under a red light to avoid decomposing the photoresist spun on the glass. A photo mask with the spiral shape geometric pattern was laid on the photoresist layer of the glass slide. Using the flood exposure system, photoresist glass was exposed to ultraviolet light for 4 s. Then it was submerged in KOH for 1.30 min. followed by ceric sulfate solution for 3 min. while keeping in dark. At this point, the exposed photoresist was developed. The lights in the room were switched on after this

step. Then the glass slide was rinsed thoroughly with distilled water and dried with nitrogen gas. Then glass chip was baked in oven at 100 $^{\circ}$C for 20 min to remove the solvents and solidify the photoresist so that more durable glass etching can be achieved.

The pattern on the glass was then etched to make the channel. The etching of the glass plate was carried out by submerging the glass in a buffered oxide etch solution, while stirring the solution with a magnetic stirrer. The plate was turned 90° every 5 min. This was done until channel depth reached 100 μm. The glass was then rinsed thoroughly with distilled water. The etched channel depth on the glass slide was measured periodically using a stylus based surface profiler. When the desired depth was achieved, etching was stopped. Then the glass plate was washed with acetone, ethanol and distilled water and was dried with nitrogen gas. Finally the glass slide was sliced into ten small chips by cutting the glass by dicing saw.

The final aphid feeding chip had dimensions 2.5 cm×5 cm×0.5 cm (1"×2"×0.2") and with a serpentine channel (average depth of 100 μm), etched on one side of the glass. In order to add the diet, two holes were drilled at two ends of the channel and two upchurch fittings were glued using JB Weld, covering the holes, on the opposite side of the glass from the etched channel. In addition a line was painted with yellow paint on the same side of the glass, outlining the channel. The complete aphid chip is shown in Figure 2.1.

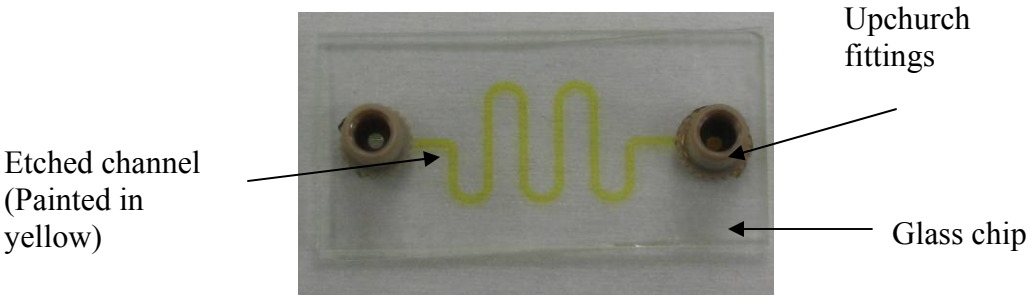


Figure 2.1: Aphid feeding chip device.

2.2.4 PDMS (Polydimethylsiloxane) cage

A cage made up of solidified polydimethylsiloxane (PDMS) elastomer was used to hold the aphids in the glass chip microfluidic device. PDMS is an organosilicon polymer compound. A corral to hold the aphid was constructed using the Sylgard 184 elastomer kit supplied separately as the elastomer base and the curing agent. The Sylgard base and curing agent were added together in 10 in 1 ratio by weight respectively and was mixed thoroughly using a stirring rod. Then the air bubbles in the polymer solution were removed using the vacuum pump. Finally it was poured on to a rectangular mold and kept in an 80 $^{\circ}$C oven for 30 min. Once the PDMS was solidified it was removed from the oven and cooled. Then a rectangular cube was cut out from the middle of the PDMS to make the cage. Sylgard base and curing agent was added in 1:20 ratio by weight and repeated the same procedure in order to make thin PDMS cover slips.

2.2.5 Experiments on glass chip microfluidic device

The etched channel of the aphid glass chip was covered with a piece of stretched Parafilm^{TM} in order to make an enclosed channel to fill the aphid diet. Then the PDMS cage was attached on the Parafilm^{TM} layer on the chip. Ten similar aphid chips were used for one experiment. All ten aphid chips were filled with complete diet (described in chapter 1), approximately 13 μL in each chip, by passing the liquid through the channel by vacuum pump. Ten aphids were added inside the PDMS cage of each chip and covered with the thin PDMS covering slip. This glass chip devices with the aphids were kept as the glass chip on the top of the PDMS cage, so that aphids were hanging onto the channel for feeding. Aphid feeding behavior of the aphids in the device was observed for 4 days.

2.2.5.1 Observing the feeding behavior of aphids

Feeding behavior was observed in aphids feeding in the glass chip device, using confocal light microscope. In addition, the experiment was carried out by adding 1% blue colored food coloring to the aphid diet. If the aphids feed on the diet their abdomens turn blue and they excrete blue colored honeydew. The number of blue aphids and honeydew drops were recorded for 4 days for in each of the chip.

2.2.5.2 Aphid feeding and survival rate using even aged aphid cohort

The survival rate of the aphids was observed. Even aged, aphid cohort was used for this experiment. Ten aphids were put in one aphid chip with complete aphid diet and recorded the number of live aphids and number of nymphs reproduced.

2.2.5.3 Aphid feeding level measurements with capillary tubes

Feeding rate was observed by clamping two capillary tubes on top of two ends of the channel and observing change in diet level with time.

2.2.5.4 Collection of aphid diet/saliva sample

Twenty aphids were added in each of the PDMS cages with the aphid chip covered with parafilm and were kept feeding for 24 h. After 24 h, parafilm on the aphid chip was removed and approximately 15-20 μL of fresh complete aphid diet was added on to the channels in order to dissolve the remaining diet and saliva sample in the channels. After 30 s the fluid on the chip was pipetted out. This was repeated for all 10 aphid chips and all the samples were collected to the same vial. The total volume of the sample collected was in the range of 200-220 μL. Not more than 80% of the contents could be recovered. The collected samples were centrifuge filtered by vivaspin 2 (2 mL volume) 3000 MWCO filter tubes, centrifuging at 3000 rpm. Before centrifugal filtration, the sample was added with a wash solution of 10% acetonitrile/water + 0.02% sodium azide in ten fold dilution. The centrifugal filtration was repeated for three times, after adding the wash solution in each time. The collected sample was stored at -80 $^{\circ}$C. HPLC analysis was carried out for collected aphid diet/saliva samples.

2.2.6 Plexiglas® microfluidic device

A Plexiglas® sheet with dimensions of 7.5cm×2.5cm×0.5cm was constructed. A channel was milled on one side, at the center of the Plexiglas® that can hold around 50 μL diet. Two holes were drilled at the two ends of the channel and upchurch fittings were attached on each hole at opposite ends of the channel (Figure 2.7). A yellow theatre gel stripe was attached on outside of the device right above the channel to attract the aphids to the channel. Two different aphid chambers were used to hold the aphids during different experiments: PDMS cage and a large screened box. During the initial experiments with the PDMS cage, a PDMS sheet was made with the same dimensions as the Plexiglas® device, following the same procedure in section 2.2.4. The channel was covered with a stretched Parafilm™ and the complete aphid diet was added to the channel. Then the PDMS cage was kept on the Plexiglas® sheet to make the complete aphid chamber. During the latter experiments Screened box was used instead of PDMS cage. A lid of a (100-1000 μL) eppendorf pipette fillers were used for to make the screened chamber. The chamber was closed with a Parafilm™ and a vertical cut was made on the Parafilm™, long enough to fit the channel of the Plexiglas® device. The device was kept on the cut facing the channel down so that aphids can feed on it (Figure 2.8).

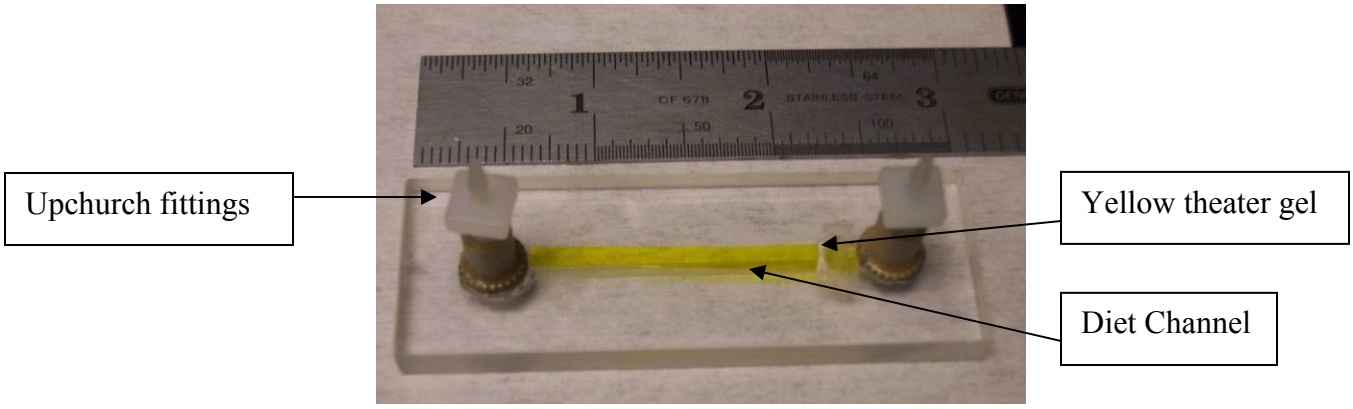


Figure 2.2: Plexiglas® aphid device

2.2.7 Pumping the aphid diet to the channel while feeding.

To fill and replenish the aphid feeding media, it was pumped using Instech's p720 peristaltic pump (Instech Solomon, Plymouth Meeting PA, USA) through the channels. Initially a single layer of Parafilm™ was used to cover the channel but it was damaged and diet was leaked out during pumping. This was avoided with double layers of Parafilm™. A cut was made on the inner Parafilm™ layer on the channel so that aphids probe through outer Parafilm™ layer only. Thereafter, the inner Parafilm™ was covered with an outer Parafilm™ layer. For the experiments reported below, two methods for pumping the aphid feeding media were explored: circulating the same diet and pumping new diet continuously.

2.2.8 Experiments on plexiglas® device

Twenty five aphids were added to the device and complete diet with 1% blue food coloring was added to the channel. Number of live aphids and number of aphids feeding

on the diet were recorded each day, for 3 days. Two capillary tubes were fixed on the upchurch fittings to measure the change in diet level during feeding. Aphid feeding and survival rate was recorded in Plexiglas® devices using PDMS cage and screened box. The experiments were carried out using both pumping diet media and stagnant diet.

2.3 Results and discussion

2.3.1 Glass chip aphid device

Aphids were attracted to the yellow line on the aphid glass chip device and were feeding on the diet in the channel as shown in Figure 2.2. Microscopic images showed how the aphids probe their stylets through the membrane film (Figure 2.3).

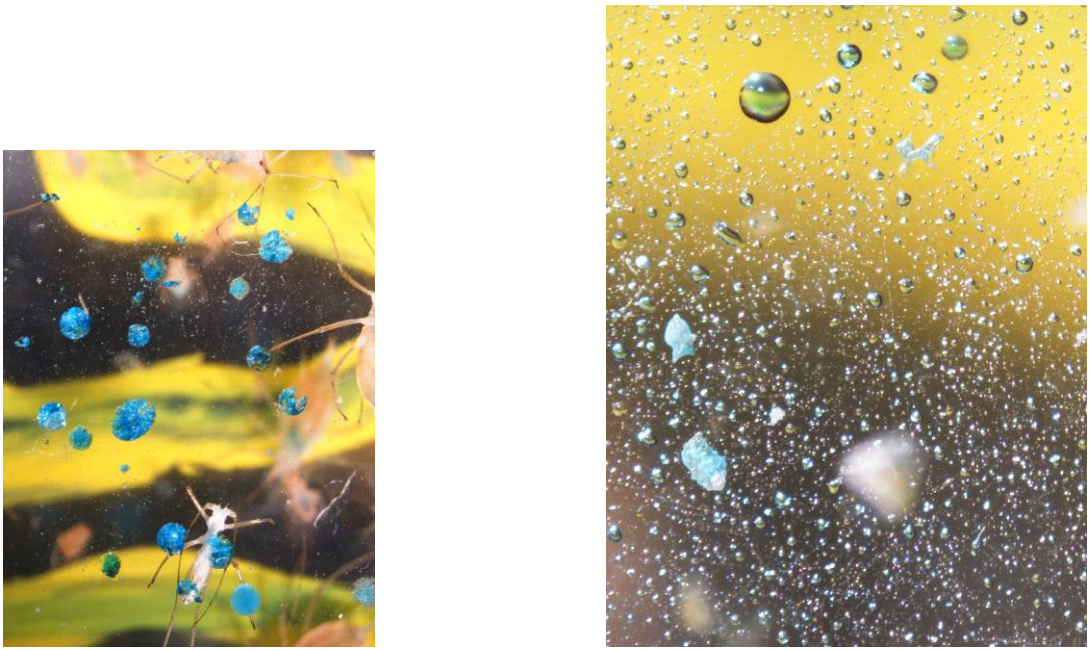
Figure 2.3: Aphid feeding on the diet in the aphid chip device with yellow painted line.

Figure 2.4: Microscopic image of an aphid penetrating its stylet through parafilm.

Figure 2.5: Microscopic image of aphid fed on the diet mixed with blue coloring.

When aphids were fed on the diet with food coloring after 24 h, their abdomen turned blue, confirming that they had fed on the diet (Figure 2.4). In addition, they have excreted blue colored honeydew, confirming the feeding on the diet (Figure 2.5 (a) & (b)). There was some difference in the size of the honeydew droplets when feeding on 20 amino acid

diet vs. the 3 amino acid diet. Aphids fed on 3 amino acid diet produced fewer but larger honeydew drops while aphids fed on the 20 amino acid diet produced more of smaller sized drops.



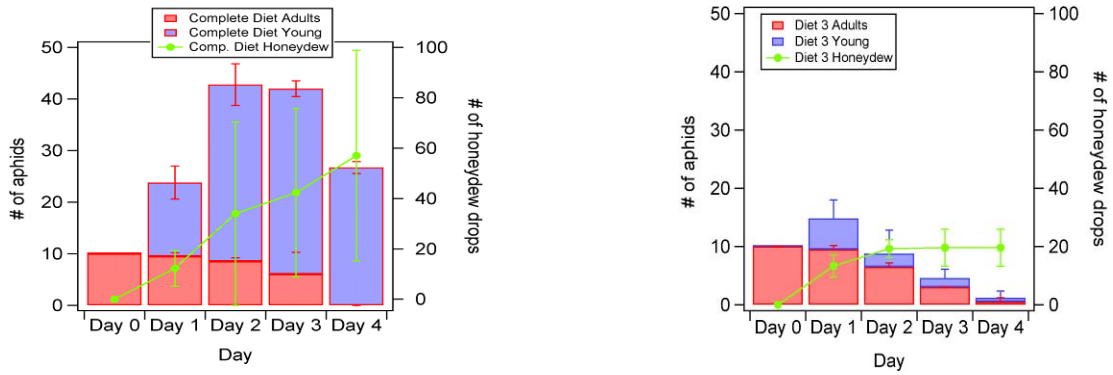
(a) (b)

Figure 2.6: Microscopic images of honeydew droplets of aphids fed on (a) Diet 3 and (b) Complete diet mixed with blue coloring.

2.3.2 The mortality rate for even aged cohort aphids

The number of living aphids feeding on the complete diet was higher than that of diet 3 aphids. (Figure 2.6 (a) & (b)). A similar trend was seen with more honeydew production on the complete diet than on diet 3. During the first two days aphids seem to feed well and an increasing number of nymphs are seen for both the diets. After the second day the

number of living aphids started decreasing. However, the nymphs seem to survive better than the adults because on the 4th day of the experiment, only the nymphs were still alive. When the aphids were fed on the complete diet, the honeydew production did not reduce, even though the aphid numbers declined. This indicates that the young aphids were continuously feeding on it, even after 4 days. However, when feeding on diet 3, the aphids stopped feeding after the third day and the honeydew production rate remained constant.



(a) **(b)**

Figure 2.7: Graphs of number of live adult and young aphids and number of honeydew droplets feeding on (a) Complete diet and (b) Diet 3 throughout four days.

2.3.3 HPLC Analysis of collected aphid salivary samples

No reproducible data was obtained as the sample concentration was too low.

2.3.4 Plexiglas® aphid device

More aphids survived (~ 10% increase) on the plexiglas® device than on the glass chip device (figure 2.9 (a) & (b)). This may be due to the larger space in the Plexiglas® chamber and higher volume of diet in the channel. However, no reproducible data was obtained in the diet level in capillary tubes to indicate the feeding rate of the aphids. This was mainly because the high rate of evaporation of the diet when they were kept under the clean hood with air blowing. Aphid feeding further improved, approximately 20%, (Figure 2.9 (b) & (c)) when they were kept in the screened box instead of the small PDMS cage. This may be due to better ventilation and space.

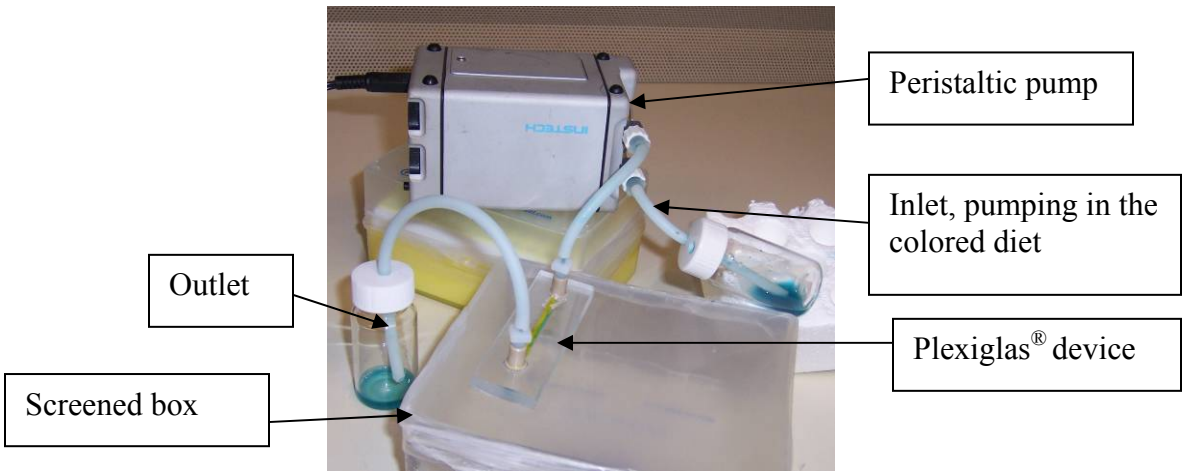
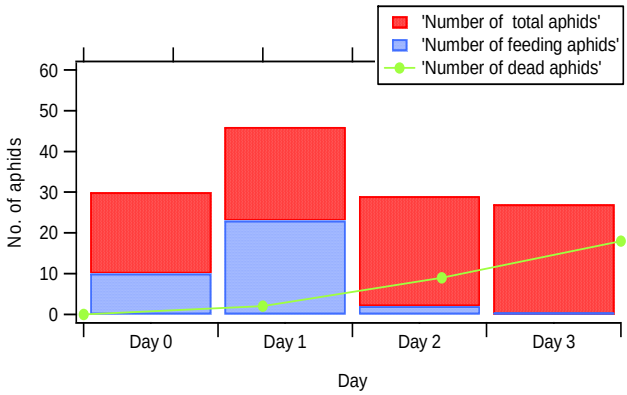
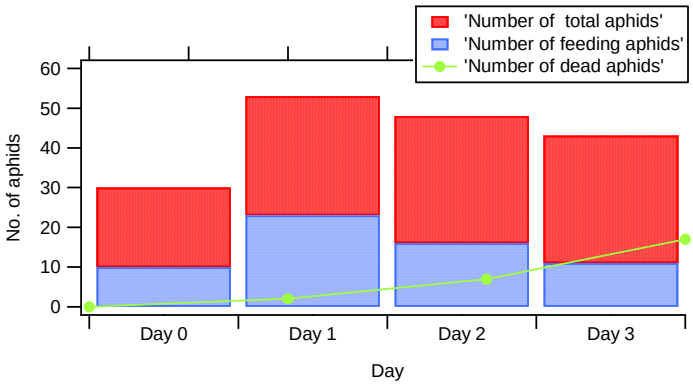


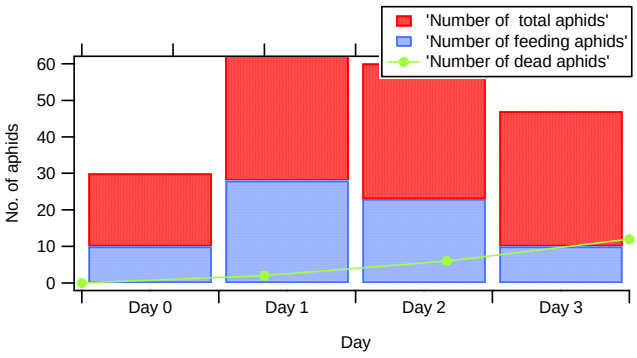
Figure 2.8: Experiment setup of Plexiglas® device on screened box covered with Parafilm™**. The diet is pumped with peristaltic pump from inlet diet reservoir to outlet.**



(a)



(b)



(c)

Figure 2.9: Graphs of Number of total aphids, Number of feeding aphids and Number of dead aphids feeding on complete diet for three days in (a) Glass chip microfluidic device (b) Plexiglas® microfluidic device with the PDMS cage and (c) Plexiglas® microfluidic device with the screened box.

2.3.5 Effect of pumped aphid diet vs stagnant diet

The number of aphids that were feeding on the diet was declined when the diet was pumped through the channel (regardless whether same diet was circulating or new diet was pumping continuously). The number of live aphids also decreased when the diet was flowing, compared to the stagnant diet. The same results were obtained even after the flow rate of peristaltic pump was reduced. This is curious because the aphids feed well on the flowing diet from the phloem tubes (Figure 2.9 (a) & (b)).

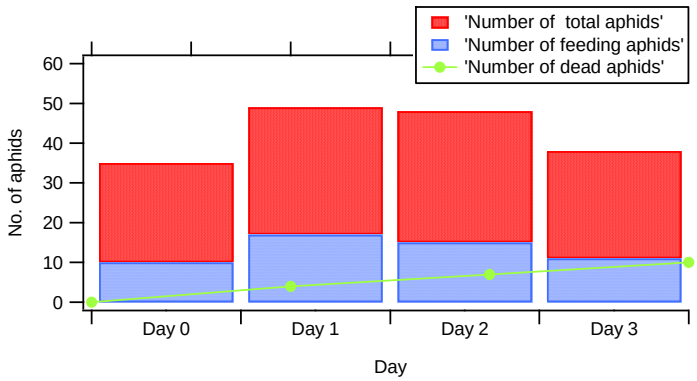


Figure 2.10 (a): Aphid survival rate and feeding with stagnant diet, on plexiglas® device.

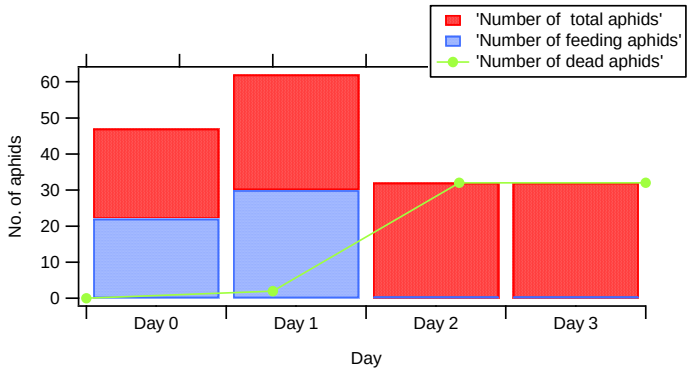


Figure 2.10 (b): Aphid survival rate and feeding with flowing diet on plexiglas® device.

2.4 Summary

Aphids were able to feed on an artificial diet contained in microfluidic devices. Survival rates of aphids were measured in both glass chip and Plexiglas® devices. The feeding behavior on two different diets, complete diet and 3 amino acid diet were also compared. In general, aphids were able to survive in microfluidic devices for two days while feeding on the diet. Aphid survival rate was higher on complete diet with higher nutrition content than when using the 3 amino acid diet. The collected salivary samples from the glass chip devices did not generate any reproducible results in HPLC analysis. The survival rate was higher in Plexiglas® device than the glass chip device which might be due to larger space in Plexiglas® device. On the other hand aphid survival was decreased when the diet was pumped, compared to the stagnant diet.

CHAPTER 3: Large scale sample collection

3.1 Introduction

This chapter describes the salivary secretion collection method that was performed from large number of aphids after feeding on a large volume of artificial diet. The collected samples were washed and concentrated prior to HPLC analysis.

3.2 Materials and Methods

3.2.1 Harvesting aphids

The pea aphid colony was grown as described in chapter 2. 8000 pea aphids were harvested from the fava bean plants (*Vivia fabe*) from Dr. John Reese's lab. The removed trunks of the bean plants were scraped with a razor blade and collected the aphids adhered on the plant were collected. It was found that 1000 aphids weighed approximately 0.795 g. 8 samples of 1000 aphids were weighed (0.795 g each) and collected separately to eight jack-pots. The jack-pots with aphids were covered with pieces of parafilm to keep the aphids from escaping, until they are transferred to the feeding chambers.

3.2.2 Artificial aphid diet

Three different artificial diets were used to feed aphids for the experiments reported below as described in chapter 1. The complete diet consisted of 22 amino acid, 15% sucrose, vitamins (ascorbic acid, thiamine etc) and metal salts. Appendix A describes the compositions in detail and the protocol for preparation. Diet 3 contained 100 mM serine,

100 mM methionine, 100 mM ascorbic acid and 15% sucrose by weight. The solution was adjusted to pH 7.5 by addition 0.75 M KOH. Finally, 15% sucrose by weight was dissolved in distilled water to make the 15% sucrose diet.

3.2.3 Aphid chamber

Two different aphid chambers were used throughout the experiments. Initially, large size (150 mm×15 mm) petri dishes (Fisher Scientific; Pittsburgh, PA) with square holes cut (Dimensions ~ 9 cm×9 cm) into the lid were used to feed the aphids. It was found that aphid feeding could be improved when a larger feeding chamber was used with better ventilation than a fully covered petri dish. Lids of 100-1000 μL pipette tip boxes were used to make larger feeding chambers. Two windows were made by cutting square pieces (Dimensions ~ 3.5 cm×12.5 cm) in two opposing faces of the box and a screen was glued on each window. Type 304, stainless steel woven wire cloth (McMaster-Carr; Princeton, NJ) was used for the screen. The screen was adhered to the box with hot glue. The two windows allowed steady air flow circulation in the chamber when the chambers were kept in the clean hood with blown air. 8 similar chambers were prepared by following the above protocol to keep approximately 1000 aphids in each chamber.

3.2.4 Assembling aphid samples

1000 aphids were put into each of the boxes were then closed with a stretched parafilm layer. A known volume of aphid diet was added onto the Parafilm^{TM} using a 100-1000 μL eppendorf pipette. The volume of the diet was changed according to the diet and the feeding time. The diet with saliva samples were collected after 24 h/48 h. For the 48 h

feeding samples, the 3 amino acid diet and the 15% sucrose diet volume used was 1mL per aphid box because amount of feeding is low in those diets. However, when the complete diet was used, the diet volume was increased to 3.5 mL per box so that the survival rate of the aphids was improved. In addition, higher diet volumes could help to prevent the aphids from feeding on saliva that their neighbors had already spit. The diet volume was kept as 1mL for all three diets when the samples were collected in 24 h.

After the diet was added onto the Parafilm^{TM} it was covered with the second layer of stretched Parafilm^{TM}. The edges of each box were sealed with additional Parafilm^{TM} stripes to prevent the aphids escape. The diet was spread in-between two Parafilm^{TM} layers covering as much as possible (Figure 3.1). Then the box was turned upside down for 20 min so that the aphids at the bottom of the box would fall onto the Parafilm^{TM} and the aphids could cling on the Parafilm^{TM} and start feeding on the diet. After 20 min the box turned upright without aphids falling back to the bottom of the box (Figure 3.2 (a) & (b)). A blank sample was prepared with a 1 mL diet kept in between the two stretched Parafilm^{TM} on a petri dish with no aphids added to it.

3.2.5 Method optimization for sample collection

The aphid samples were kept in the clean hood in order to minimize contamination. The samples were kept under a constant yellow light. Initially, a rack was used to stack the aphid boxes to save space in the clean hood, but the boxes in lower racks received less light than those above. As a result, the aphids started to move to the front of the box to receive more light, instead of consuming the diet over the entire

parafilm surface. Thereafter, the boxes were spread out over the clean hood floor without using the rack so that all the aphid boxes received the same amount of light.

Each aphid chamber was covered with a square piece of yellow theater lighting gel (Roscoe 10: A to Z Theatrical: Kansas City, MO) to keep the aphids attracted to the diet within the Parafilm^{TM} layers. Theater gel allows 92% of the light to penetrate through it and hence aphids were better attracted to the Parafilm^{TM} than when a regular yellow paper was used.

Finally, the air was circulated in the clean hood continuously so that the aphids had good ventilation throughout the experiment.

3.2.6 Salivary sample collection and washing

The aphid feeding experiments were continued for 24 h or 48 h depending on the experiment. After 24/48 h time, the diet was removed from each aphid box, starting with the blank sample. First, the top layer of the Parafilm^{TM} was carefully removed. Then, using a new sterile syringe, the diet and the saliva secreted into the diet were collected from the inner Parafilm^{TM} layer and transferred to a Falcon conical tube (Fisher Scientific; Pittsburg, PA). Then ~ 8 mL of filtered (using 0.45 μm membrane filters) ultra pure distilled water (Barnstead Ultrapure Water System; Dubuque, IA) was used to rinse upper and lower Parafilm^{TM} layers. The washings were collected with the syringe. The wash water was reused to wash Parafilm^{TM} of all 8 aphid chambers. The washings were

collected into a separate Falcon conical tube. A clean spatula was used to gently scrape and gather the residual diet on the two Parafilm^{TM} layers of each chamber.

3.2.7 Sample pre concentration

The collected diet/saliva samples from all of the aphid chambers were combined in the Falcon conical tube and a wash solution was added before centrifuge filtration. Two different wash solutions were used, one with 10% acetonitrile/0.02% sodium phosphate in water and the other one with 10% acetonitrile/0.02% sodium phosphate in 100 mM sodium phosphate pH 7 buffer in different experiments.

The collected diet/saliva sample and washings from all 8 aphid chambers was centrifugally filtered in order to increase the aphid salivary protein concentration. This was performed using Vivaspin molecular weight cut off (MWCO) filters (Sartorius Stedim Biotech; Aubagne, France). Two different MWCO tubes were used as Vivaspin 20, 3000 MWCO and Vivaspin 15, 5000 MWCO in different experiments. First the collected samples and washings were added to filtration tubes. The blank sample which was collected from the chamber with no aphids was added to a separate Vivaspin tube. Then the wash solution was added to generate 10-fold dilutions in volume to each Vivaspin tube.

The collected aphid saliva samples were centrifuged using GP Centrifuge with GH 3.7 horizontal rotor (Beckman Instruments, Palo Alto, CA). The centrifugal-filtration was repeated three times for each sample. After each centrifugal-filtration step, the wash

solution was added to generate a 10-fold dilution. Each centrifugal-filtration step resulted in a retentate volume of approximately 200-100 μL. As a result, the final concentrated sample volume after the third centrifugation cycle was concentrated 1000 times from the starting collected volume. This sample was further concentrated by lyophilizing to dryness. The concentrated sample was reconstituted in small aliquots of wash solution just prior to HPLC analysis.

The centrifugally filtered samples were transferred back to the clean hood and filtered through a 0.45 μm pore size polyvinylidene difluoride (PVDF) membrane filter using a sterilized syringe pump. The volume of the sample before and after the membrane filtration was recorded. The final concentrated saliva sample was stored in sterilized PCR tubes and frozen at -70 $^{\circ}$C for analysis. The summarized sample collection and pre concentration protocol is given in the flow chart (Figure 3.3).

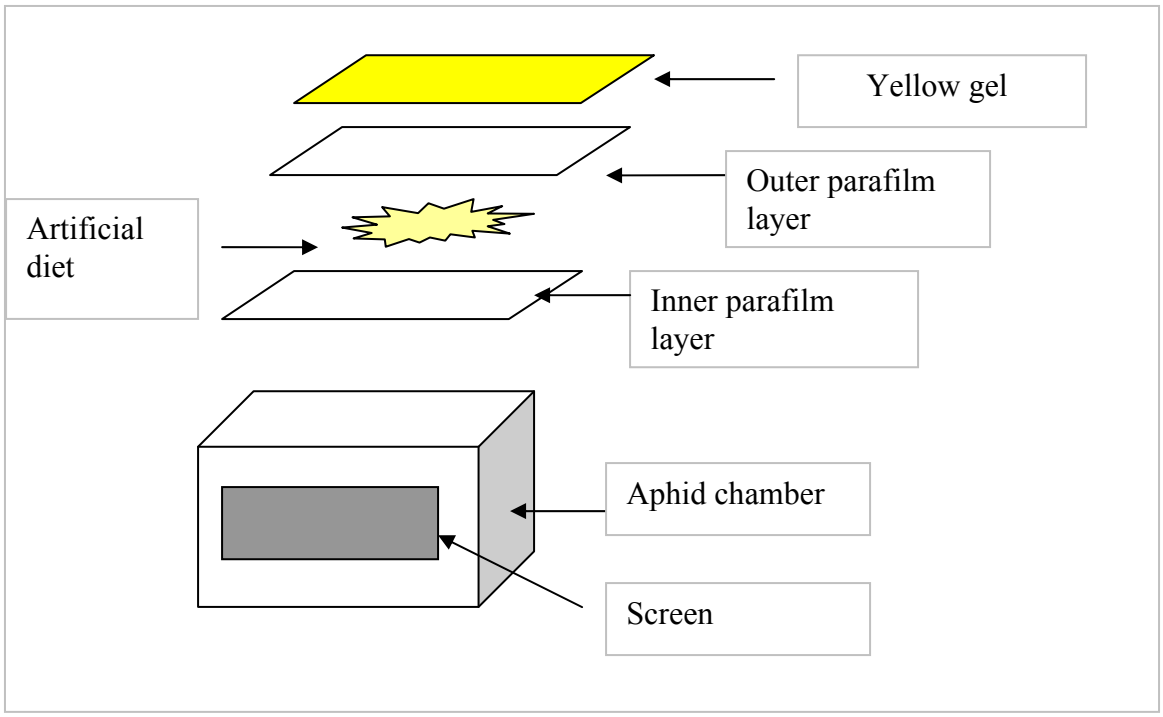
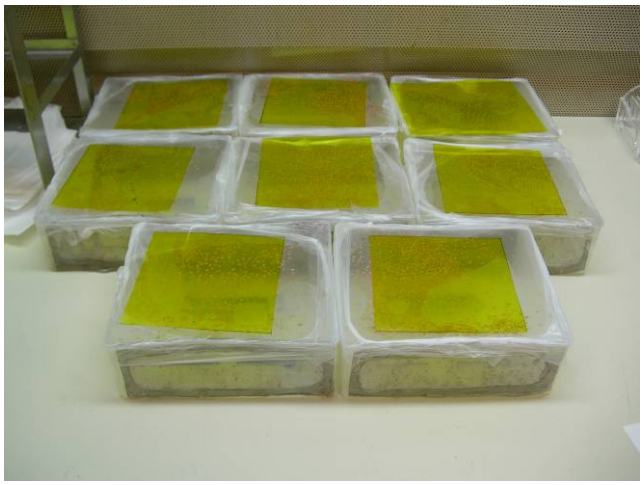
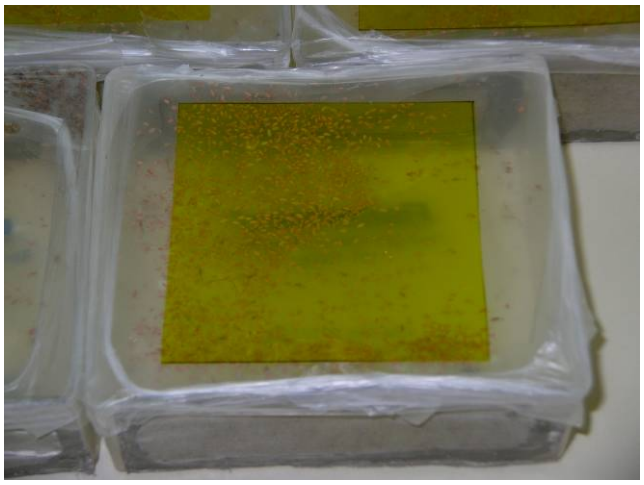


Figure 3.1: Sketched diagram of assembling the aphid screened chamber.

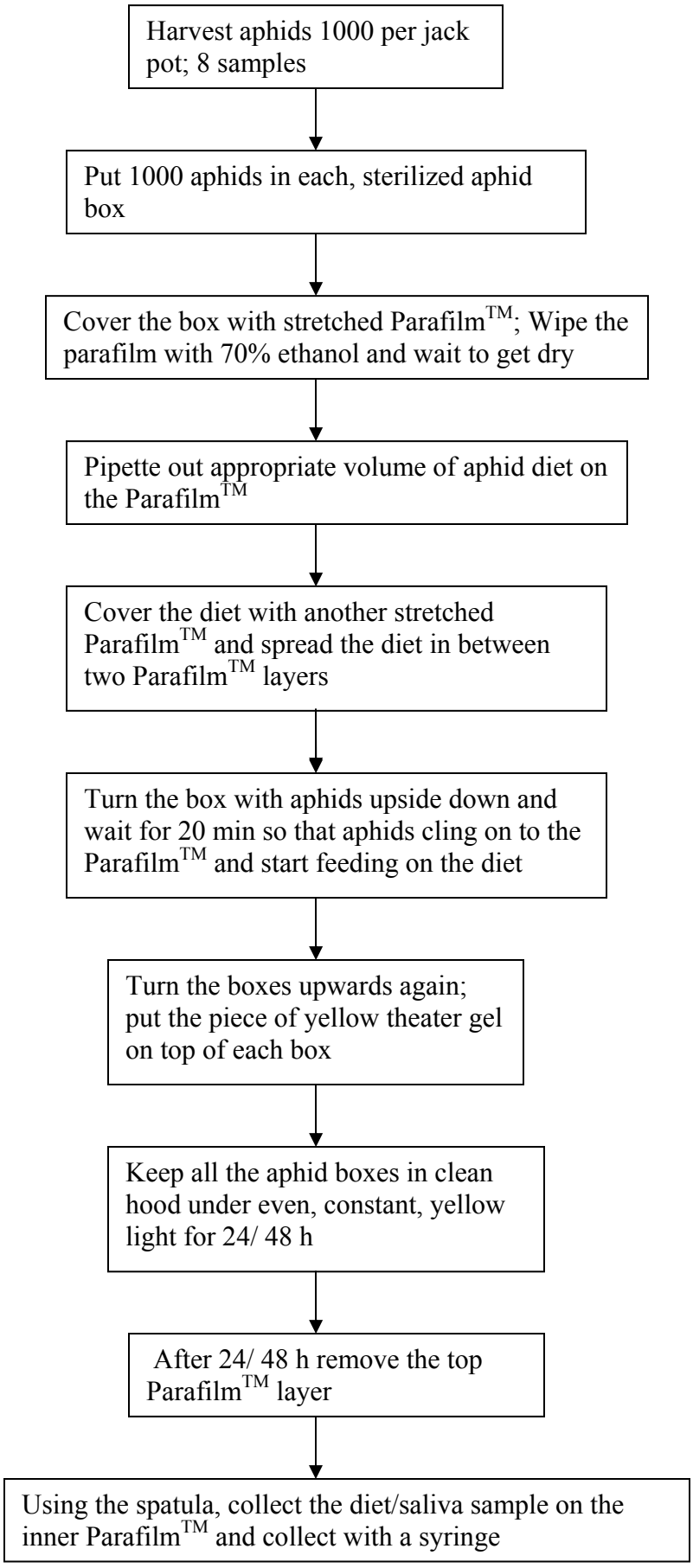


(a)



(b)

Figure 3.2: Photographs of Aphid screened chambers with 1000 aphid feeding on diet in each box (a) All 8 samples in one experiment (b) Closer view of one box.


Harvest aphids 1000 per jack pot; 8 samples
Put 1000 aphids in each, sterilized aphid box
Cover the box with stretched Parafilm™; Wipe the parafilm with 70% ethanol and wait to get dry
Pipette out appropriate volume of aphid diet on the Parafilm™
Cover the diet with another stretched Parafilm™ and spread the diet in between two Parafilm™ layers
Turn the box with aphids upside down and wait for 20 min so that aphids cling on to the Parafilm™ and start feeding on the diet
Turn the boxes upwards again; put the piece of yellow theater gel on top of each box
Keep all the aphid boxes in clean hood under even, constant, yellow light for 24/ 48 h
After 24/ 48 h remove the top Parafilm™ layer
Using the spatula, collect the diet/saliva sample on the inner Parafilm™ and collect with a syringe

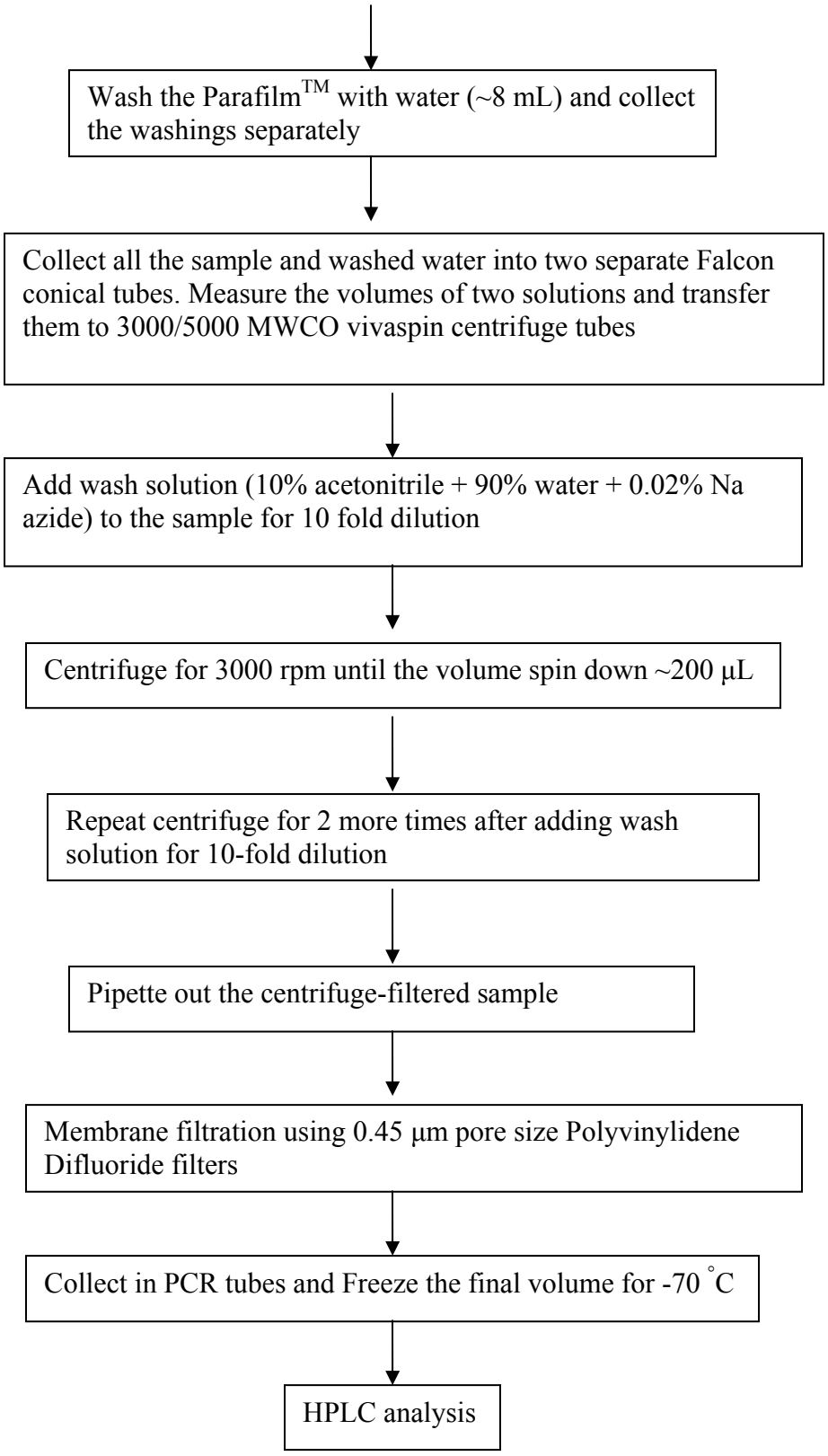


Figure 3.3: Aphid diet/saliva sample collection protocol.

3.2.8 Sterile conditions

All the steps in the sample collection protocol were carried out under sterile conditions to minimize contamination by human and bacterial proteins. The aphid feeding boxes were kept in a clean hood equipped with HEPA- filters (Purifier Clean Bench; Labconco; Kansas City, MO) during feeding. In addition, all the experimental steps were performed while wearing a clean lab coat and new gloves. The gloves taped to the sleeve of the lab coat, to avoid human skin protein contamination. In addition, the gloves were kept wet with 70% ethanol when handling the samples.

Before setting up the experiment, the aphid boxes, aphid diet and pipettes were sterilized by wiping with 70% ethanol and exposing to uv light in the biohood for 30 min. The centrifuge tubes, transferring tubes, spatula and PCR tubes were washed with 70% ethanol and dried before use. Transferring the samples from clean hood to centrifuger was carried out in a tightly closed box, wiped with 70% ethanol.

All the solutions were filtered using 0.45 μm pore size polyvinylidene difluoride (PVDF) membrane filters, including the aphid diet prior to feeding, wash solutions before addition to the collected samples and water that was used to wash the Parafilm^{TM}. The quantitative transfer of wash solution to the Vivaspin tubes was carried out using sterilized transfer bottles and sterile syringes and the transfer was performed under a clean hood.

3.2.9 HPLC separations

HPLC separations and data analysis were carried out by Amanda Meyer and it was performed at US Department of Agriculture, Grain Marketing Production Research Center (USDA GMPRC), Manhattan, Kansas. Initially, a reversed phase nano-HPLC column (1100 Series HPLC; Santa Clara, CA) was used for the separations since the sample concentration was very low and it was essential to have a highly sensitive column to achieve successful separations. A C-18, Jupiter column was used with the dimensions of 0.3 mm i.d. × 15 cm long. The particles were 5 μm with a 300 °A pore size (Phenomenax; Torrance, CA). The mobile phase composition was in the range of 20-95% acetonitrile (B)/water (A) and 0.01% TFA. The separations were performed at a 4 μL/min flow rate. The column temperature was 50 °C with an injection volume that varied from 3-8 μL. HPLC detection was carried out by absorption UV 214 nm, 254 nm and 280 nm. However, flow in the nano HPLC column was difficult to maintain during the separations of the unpurified protein sample mixture which included bacterial contaminant and solid residuals in the diet.

Because of the difficulty using the nano HPLC, most separations were carried out using, a conventional reverse phase HPLC (1100 Series HPLC; Santa Clara, CA) system. The column was a C-18, poroshell 300SB column, 2.3 mm i.d. and 75 mm long. The particles were 5 μm i.d. and the pore size was 300 °A. The column temperature was kept at 50 °C. The average injection volume was ~25 μL but the volume was normalized to account final pre concentrated sample volume differences. Absorbance detection was performed at 214 nm, 254 nm and 280 nm simultaneously.

3.3 Results and Discussion

3.3.1 Reproducible results

Reproducible results were obtained at 214 nm as shown in figure 3.4. The mobile phase was 10%/90% Acetonitrile/water. The HPLC peak intensities were improved by optimization the conditions as described below.

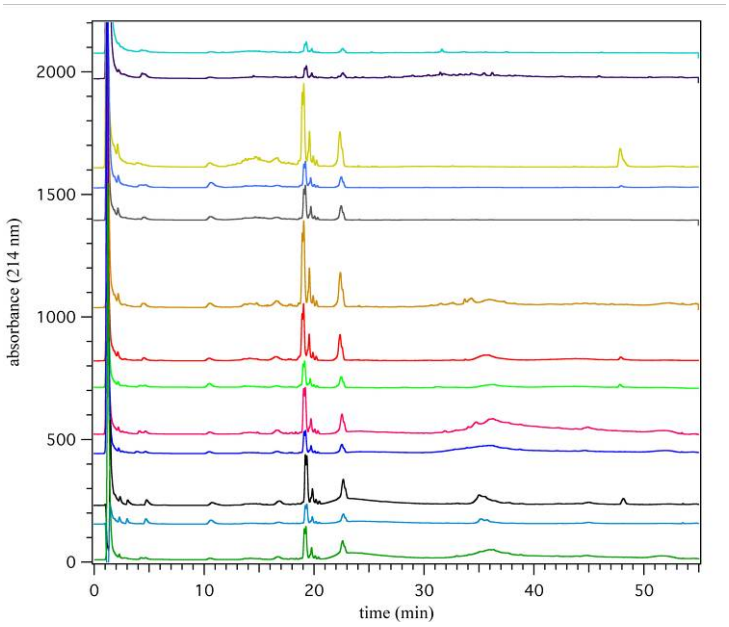


Figure 3.4: RP-HPLC chromatograms of pre concentrated aphid salivary secretions of different samples obtained at 214 nm indicating the reproducibility of multiple chromatograms. The mobile phase was 10%/ 90% Acetonitrile/ water[†].

The sample collection and pre concentration steps were optimized to obtain better separations with high reproducibility in HPLC chromatogram. It was observed that when

[†] Chromatographic experiments and analysis performed by Dr. Amanda Meyer

aphids were kept in the screened chambers, better separations could be achieved than when using the petri dishes. In addition, HPLC spectra were improved when the yellow theater gel was used to cover the chamber instead of yellow paper. The other conditions that were changed to optimize the sample collection protocol were, keeping the aphid feeding chambers under even light and rinsing the Parafilm™ with water during sample collection and adding the washings to the collected sample.

3.3.2 Method optimization

A variety of experimental conditions were modified to increase the concentration of aphid salivary secretions. When petri dishes were used as the feeding chamber, covered with a yellow paper, the chromatogram generated lower intensity peaks (chromatogram (1), Figure 3.5). When theater yellow gel was used higher intensity peaks were generated. The increased light intensity through the theater gel seems to be the reason for the improved peak heights (chromatogram (2), Figure 3.5).

Initially the dishes were stacked in a rack to save space in the clean hood. However, the shadow of the rack prevented light from evenly illuminating the dishes. The aphids on the top of the stack were fed well but the amount of feeding decreased in aphids in lower stacks, reducing the overall aphid feeding. The positive effect of even light is illustrated in the in chromatogram (3) in Figure 3.5, by keeping the dishes on clean hood floor instead of rack.

The peak amplitudes were further improved almost 3 fold by rinsing the inner Parafilm^{TM} layer after collecting the diet/saliva sample and adding the washed water to the sample as shown in chromatogram (4), Figure 3.5.

Finally when the screened chambers were used to keep the aphids instead of petri dishes while following the rest of the optimized conditions same, maximum peak amplitude with highest peak intensities were obtained as shown in chromatogram (5), figure 3.5. The screened chamber contained enough space, light and better air circulation so that optimized conditions were accomplished for aphid feeding. When aphids were fed on 1-2 mL of complete diet per screened chamber for 48 h, almost all of diet was consumed because the aphids had high reproduction and high feeding rates during the experiment. Therefore, the diet volume was increased to 3.5 mL per chamber in order to collect the extra residual diet.

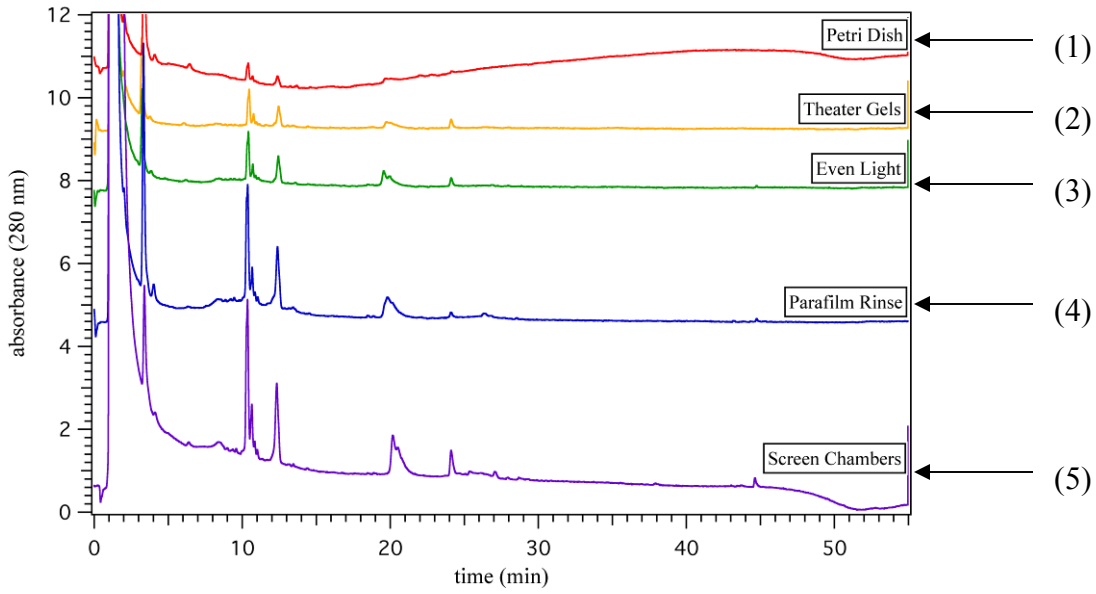


Figure 3.5: RP-HPLC chromatograms at 280 nm with improved signal strength due to various changes in the sample collection protocol. The top right corner box shows the changed step, while keeping the changes of the above chromatogram remain as the same[†].

3.3.3 Sample variables

The effect of different variables was investigated in the HPLC results. The variables that analyzed from HPLC chromatograms were, the different diet compositions (complete diet vs diet 3 vs. 15% sucrose), feeding time (24 h vs 48 h), the diet with extra calcium (1 mM) and without extra calcium, the diet volume per box, different wash solutions (10% acetonitrile in water vs 10% in phosphate buffer) and different molecular cut off filters (3000 MWCO vs. 5000 MSCO).

[†] Chromatographic experiments and analysis performed by Dr. Amanda Meyer

3.3.3.1 The effect of using different aphid diets

HPLC analysis showed that changing the aphid diet has significant impact on the amount of salivary secretion (Figure 3.6). It was found that the diet with 15% sucrose was the best out of the three diets tested for inducing salivary secretion, because those samples generated the highest amplitude peaks. The second best diet was found to be the 3 amino acid diet while the complete diet gave the least salivary secretion.

However, when comparing the number of aphids that survived on each diet after 48 h, the complete diet was the best generating a high aphid reproduction rate and the least number of dead aphids. The 15% sucrose diet had fewer live aphids than the 20 amino acid diet but a higher number than in the 3 amino acid diet. As a whole, the amount of salivation did not increase proportionately with amount of feeding or the survival rate of aphids.

3.3.3.2 The effect of time duration of feeding: 24 h vs 48 h

The aphids were fed on the diet for both 24 h and 48 h times in different experiments. The peak amplitudes were approximately two folds higher in 48 h feeding than that of 24 h (Figure 3.7). The Figure 3.7 shows chromatograms of collected fed upon 3 amino acid diet. The chromatograms show the peak intensity differences between 24 h and 48 h feeding. A similar trend was observed with feeding from the other two diets. A blank sample was run for each corresponding sample so that the peaks that were not seen in blank sample should be due to aphid spits. However, when mass spectrometry analysis

performed later, it was found that 24 h feeding samples were less contaminated with bacterial proteins than the 48 h feeding samples.

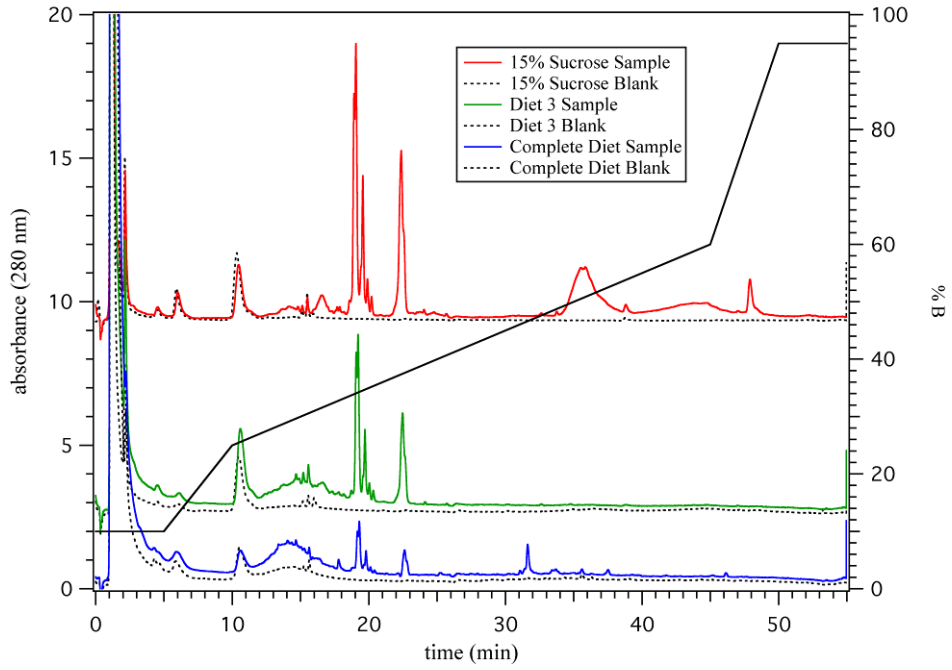


Figure 3.6: RP-HPLC chromatograms at 280 nm illustrating the difference of peak intensities due to different diet fed by 8000 aphids for 48 h. Dotted line spectrums are the blank samples of each corresponding fed samples. The black line plot shows the mobile phase gradient at the right axis. B% indicates the % Acetonitrile[†].

[†] Chromatographic experiments and analysis performed by Dr. Amanda Meyer

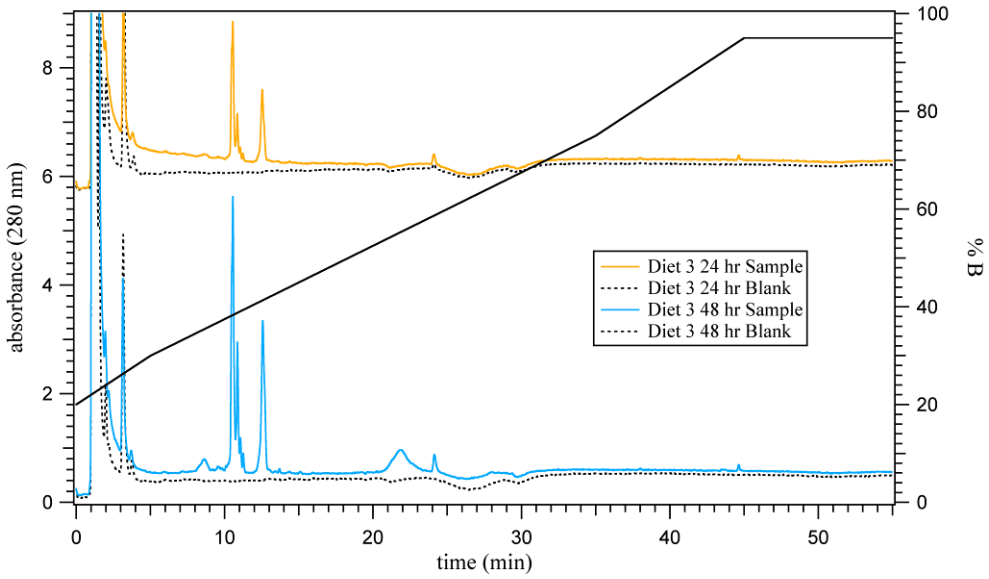


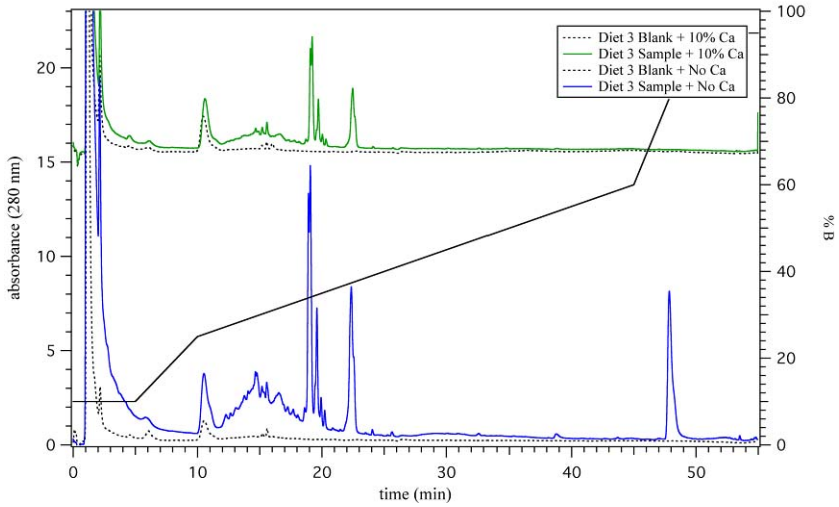
Figure 3.7: RP-HPLC chromatograms at 280 nm illustrating the impact on time duration of feeding, 24 h and 48 h. Peak signal amplitudes became almost doubled in 48 h feeding than 24 h feeding. The mobile phase gradient (black trace) is represented as a function of %B on the right axis. %B indicates the % Acetonitrile[†].

3.3.3.3 The impact of Calcium added to the diet

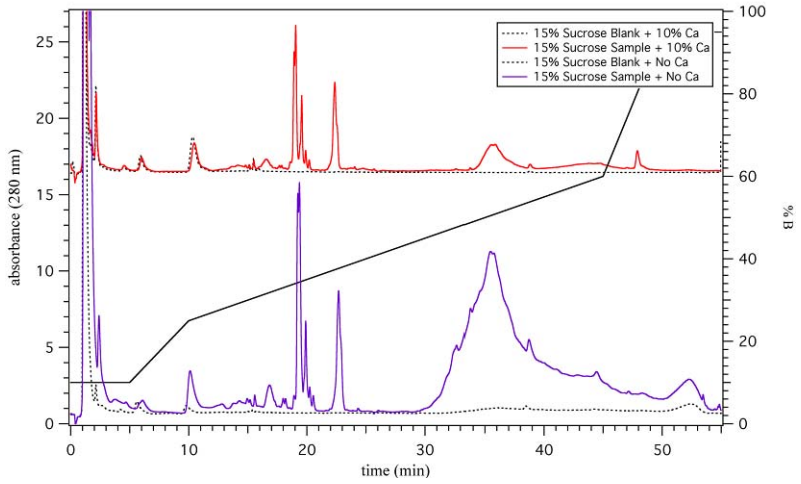
1mM Calcium was added to aphid diets and the salivary secretions were analyzed. For all three different diets calcium negatively affected for salivary secretion. This is illustrated in HPLC chromatograms of samples with diet 3 and 15% sucrose diets shown in Figure 3.8. The chromatograms of the complete diet feeding with and without calcium added were not shown. Figure 3.8 showed fewer peaks with low amplitudes when extra calcium was included diet compared to samples without extra calcium.

[†] Chromatographic experiments and analysis performed by Dr. Amanda Meyer

Figure 3.9 shows the comparison of diet 3 and 15% sucrose with calcium compared to complete diet without calcium when all the other conditions were same. It shows that the complete diet peak amplitudes were lower when compared with the other two diets even after the addition of extra calcium. There was a slight precipitate or turbidity in some samples when extra calcium was added to the diet. This may be one reason for the reduction in peak amplitudes.



(a)



(b)

Figure 3.8: RP-HPLC chromatogram at 280 nm illustrating the impact on addition of extra calcium to diets of (a) Diet 3 and (b) 15% sucrose compared to the spectrums of same conditions but without addition extra calcium. The respective blank samples under the same conditions are shown in dash black traces. The addition of calcium has decreases the peak amplitude when compared to chromatograms with and without added calcium in both the diets. The HPLC gradient is shown (black trace) as a function of %B in right axis. %B indicates the % Acetonitrile[†].

[†] Chromatographic experiments and analysis performed by Dr. Amanda Meyer

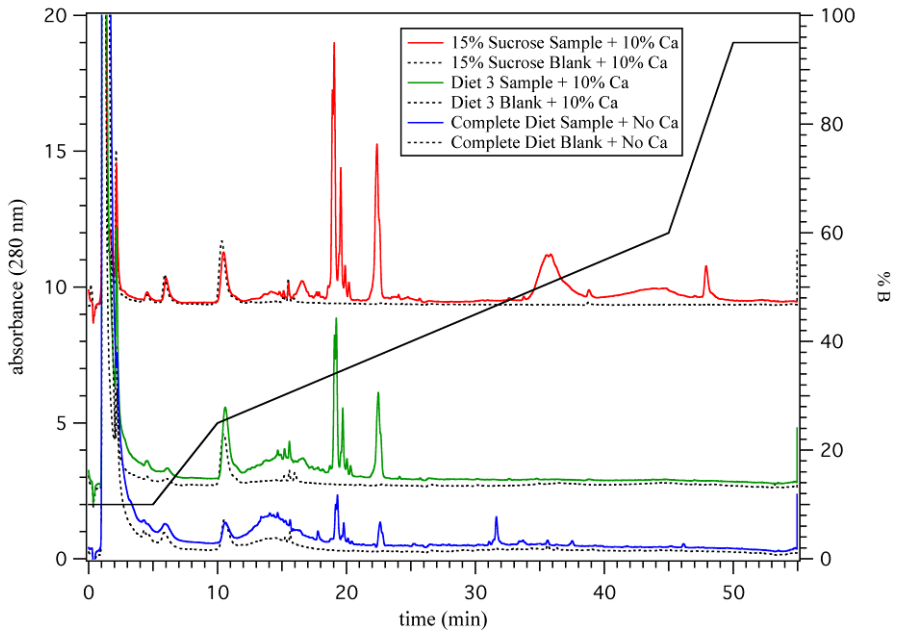


Figure 3.9: RP-HPLC chromatograms at 280 nm illustrating the impact on addition of extra calcium to diet. The 15% sucrose diet (red trace) and diet 3 (green trace) contain 10% of the 10 mM calcium chloride were added but in the complete diet no extra calcium was added. Note that the chromatograms compared to complete diet feeding shown in blue trace has even lower peak amplitudes than the diet 3 and 15% sucrose feeding peak signals. The HPLC gradient is shown (black trace) as a function of %B in right axis. %B indicates the % acetonitrile.[†]

3.3.3.4 The effect of volume of diet added to each sample

There was no distinct trend on effect of the volume added to the diet on the HPLC chromatogram. Different volumes were added (0.75 mL – 3.5 mL) to each dish to optimize the feeding. When the volume was too high relative to the fed volume it took longer time to concentrate. On the other hand if the volume was too low, the aphids had

[†] Chromatographic experiments and analysis performed by Dr. Amanda Meyer

limited diet so that some aphids might have consumed already secreted saliva from their neighboring aphids, which may have caused a sample loss. It was found that adding 1 mL per box generated optimal results for 24 h feeding for all three diets and for 48 h feeding on diet 3 and 15% sucrose. For the complete diet, 48 h samples 3.5 mL/ per box was sufficient.

3.3.3.5 The impact on different wash solutions

When centrifuge filtration is done using 10% acetonitrile in water solution better HPLC peak signals were obtained than with the 10% acetonitrile in 100 mM sodium phosphate solution. Even though Figure 3.10 shows a comparison of complete diet samples similar results were found in other two diets. There were more peaks and around 5-fold higher peak amplitudes in water wash solution than the phosphate buffer wash solution. In theory phosphate buffer should have the ability to dissolve sugar and concentrate the sample better than the water wash. Therefore, there was no direct explanation for this observation. When phosphate buffer solution was used precipitate formation was occurred in some cases. This can be one reason to get lower signal amplitudes.

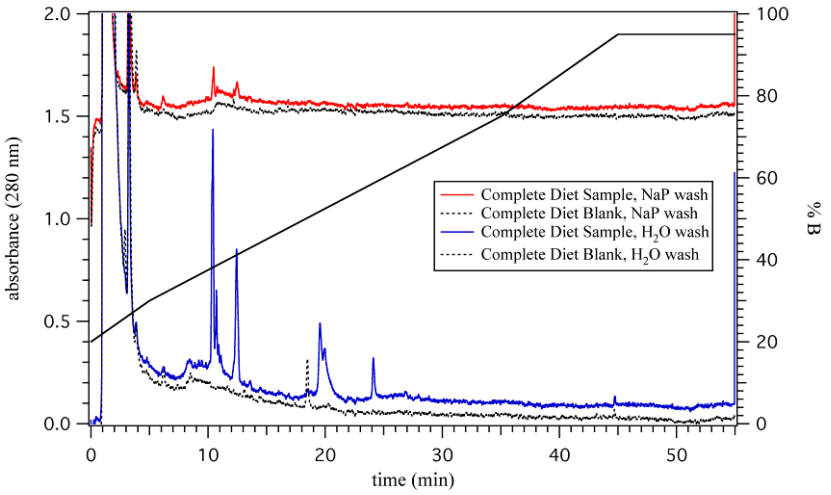


Figure 3.10: RP-HPLC chromatograms at 280 nm illustrating the impact on different wash solution used during the centrifugation. Acetonitrile/water wash gave better HPLC results enhancing the peak amplitude around 5-folds compared to that of acetonitrile/ phosphate buffer wash solution. The HPLC gradient is shown (black trace) as a function of %B in right axis. %B indicates the % acetonitrile[†].

3.3.3.6 The impact on different centrifuge molecular weight cut off filter tubes

According to the previous studies, the aphid proteins have a mass range of 6 kDa-200 kDa.[9] Therefore, both 3000 (3 kDa) and 5000 (5 kDa) were used to filter the samples. When complete samples filtered with 5000 MWCO filters, better HPLC signals were obtained than that of 3000 MWCO filters. This might be due to lower collection volume from 5000 MWCO filtering which generated higher concentrations. However, there were additional peaks that were obtained using the 3000 MWCO filters. It was difficult to determine whether these peaks were due to salivary proteins or other bacterial contamination that could occur due to longer centrifuging time using the 3000 MWCO rather than the 5000 MWCO filters. However, in the diet 3 and 15% sucrose diet samples

[†] Chromatographic experiments and analysis performed by Dr. Amanda Meyer

there were no significant difference between the results generated using the 5000 and 3000 MWCO filter tubes. As a whole, there was not enough evidence to achieve a proper conclusion regarding the type of filtering tube to choose.

3.4 Summary

Salivary samples were collected from 8000 aphids feeding off of an artificial diet for 24 h/48 h. The collected samples were centrifuge filtered to reduce the volume and reduce the sugar components so that the concentrated sample had a sufficiently high protein concentration for HPLC analysis. The HPLC chromatograms confirmed that the separated peak signals were due to aphid saliva when compared to blank samples obtained under the same conditions. Special sterilized conditions were followed to minimize bacterial and human protein contaminations. Better separations with higher peak amplitudes were obtained when sample preparation conditions were optimized. According to the HPLC results, 1 mL per chamber 15% sucrose diet with no extra calcium added, fed for 24 h generated optimum salivary secretions. In addition, better HPLC results were obtained when the samples were washed with acetonitrile in water rather than in acetonitrile in phosphate buffer. There was no significant impact on HPLC results for different molecular weight cut off filters (3000 MWCO or 5000 MWCO) for centrifugal filtration of collected salivary samples.

CHAPTER 4: Aphid salivary sample analysis

4.1 Introduction

Chapter 4 explains analysis of the collected aphid salivary samples using large scale collection method. Different analysis methods were used such as size exclusion chromatography, electrophoresis and mass spectrometry. Mass spectrometric data was extended to proteomics analysis in order to identify peptide sequences from database searching. All the sample analysis methods were performed by Amanda Meyer.

4.2 Size exclusion chromatography

4.2.1 SEC Experimental conditions

Agilent 1100 series HPLC (Agilent; Santa Clara, CA) was fitted with a Biosep-SEC S2000 column (300 × 7.8; Phenomenex; Torrance, CA). The column temperature was fixed at 40 °C. The mobile phase compositions were 50% water (A) with 0.1% triflouroacetic acid (TFA) and 50% acetonitrile (B) with 0.1% TFA and the flow rate was 0.5 mL/min. Typically, 25 μL aphid sample was injected in each run. However, this was changed in proportion to the collected volume. As for standard proteins, alcohol dehydrogenase, albumin, carbonic anhydrase and α-lactalbumin were chosen in order to have a sufficient molecular weight range for the aphid protein samples. Each of the standard protein samples was injected in 20 μL volumes. Uv detection was done at 200 nm wavelength.

Samples were lyophilized and reconstituted in 20 μL 90%/10% H_2O/acetonitrile and compared with the unwashed samples.

4.2.2 SEC Results

SEC spectra of multiple samples were shown in figure 4.1. All the samples were collected from aphid fed on the complete diet. Each spectrums obtained peaks varying within a closer range. The unwashed sample spectrum consists of multiple peaks unlike the washed samples with fewer broad peaks. The additional peaks in the unwashed sample can be due to additional components such as sucrose that can be eliminated by washing steps.

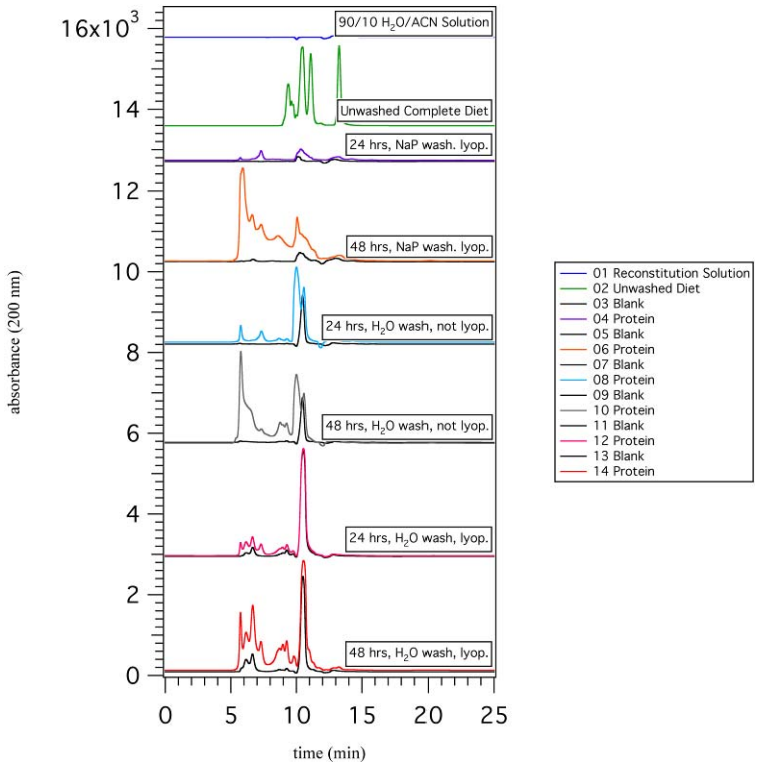


Figure 4.1: Size exclusion chromatograms at 200 nm of sample solution (10% Acetonitrile/ water), unwashed complete diet and multiple aphid salivary samples after washed. Each respective blank result is shown in black trace. The mobile phase compositions were 50% water (A) with 0.1% trifluoroacetic acid (TFA) and 50% acetonitrile (B) with 0.1% TFA. The flow rate is 0.5 mL/min and temperature at 40 °C[†].

4.2.3 Molecular Weight Markers

Size exclusion was carried out with 1 mg/mL molecular marker proteins to determine the corresponding retention times.

[†] Chromatographic experiments and analysis performed by Dr. Amanda Meyer

4.2.3.1 Results

The molecular marker proteins were eluted according to Figure 4.2 with following retention times.

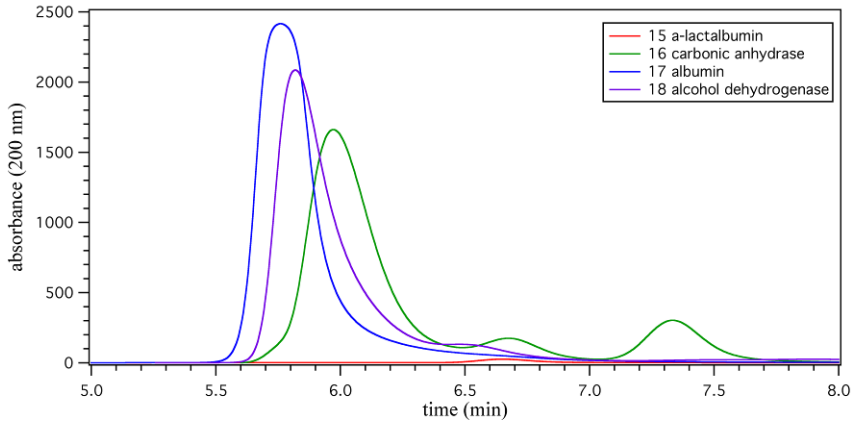


Figure 4.2: Size Exclusion Chromatogram with different molecular marker proteins (1 mg/mL) shown in the right hand corner box. This is to determine the molecular weights of aphid salivary proteins[†].

The retention times of the each marker molecule were,

- α-lactalbumin (14.2 kDa), t_{mig} = 6.65 min
- carbonic anhydrase (29 kDa), t_{mig} = 5.97 min
- albumin (66 kDa), t_{mig} = 5.76 min
- alcohol dehydrogenase (150 kDa), t_{mig} = 5.82 min

[†] Chromatographic experiments and analysis performed by Dr. Amanda Meyer

Even though the alcohol dehydrogenase has the highest molecular weight it eluted after lactalbumin. This may be due to the fact that the alcohol dehydrogenase used in the experiment was more than two years old and it may have decomposed.

Relative to the biomarkers, molecular weight estimation was done for the protein peaks that having high molecular weights and eluting at low retention times. According to the molecular weight markers, the high molecular weight peaks ranged from 14 kDa to 150 kDa from the aphid samples (Figure 4.3). 48 h samples showed more prominent peaks within this region than 24 h samples. On the other hand, the samples washed with acetonitrile in water solution showed more resolved peaks in the corresponded high molecular weights region. Peak resolution was further improved by lyophilizing the samples. The peaks with higher retention times were smaller molecules such as amino acids. In addition there are some overlapping peaks in sample and each corresponding blank spectrum. It should be noted that some of these peaks in the blank samples can be due to bacterial and human skin proteins.

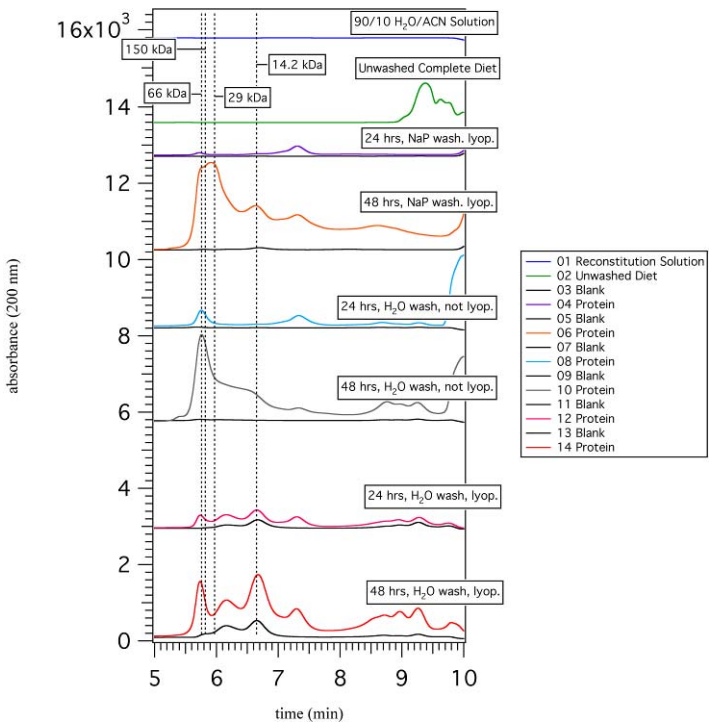


Figure 4.3: SEC results of aphid sample separations at 200 nm with time markers illustrating the corresponding molecular weights compared to the standards. Compared to unwashed aphid diet the washed diet consists of higher molecular weight components. The mobile phase compositions were 50% water (A) with 0.1% trifluoroacetic acid (TFA) and 50% acetonitrile (B) with 0.1% TFA.[†]

4.3 Electrophoresis

4.3.1 Electrophoresis experimental

In electrophoretic separation was carried out on the aphid salivary samples using 2100 Bioanalyzer System (Agilent Technologies; Santa Clara, CA). Standard preparation protocols were followed for the experiments.[49]

[†] Chromatographic experiments and analysis performed by Dr. Amanda Meyer

Standard Labeling Buffer (SLB) was used to mix the samples. SLB consists of 30mM Tris/HCl at pH 8.5. 5 μL of collected aphid sample was mixed with SLB in 1:10 ratio while maintaining the pH in 8-9 range. Then it was mixed with the dye solution and incubated in ice for 30 min. After that, ethanolamine was mixed to the sample and incubated in ice for 10 min. The dye solution was used for labeling in fluorescence detection. The above mixing procedure was repeated to the protein ladder standard. Subsequently, to both of the sample and protein ladder, denaturation buffer was mixed and heated to prime the chip and was loaded with the gel. Then, the sample and ladder standard were pipetted into the chip well and inserted into the 2100 Bioanalyzer tray. The samples was run for separation and obtained the gel chromatographic images with corresponding masses of ladder standard.

4.3.2 Electrophoresis results

According to the mass values of the protein ladder sample, the relative migration times of the aphid sample peaks were assigned in the electropherogram (Figure 4.4). Most of the proteins that were found in the spectrum were with molecular weights from 7.7 to 43 kDa. There was only one peak with higher molecular weight as 95 kDa. These results were in similar agreement with that of SEC data. However, according to the previous studies, more peaks were expected in the electropherogram. Having less number of peaks may be due to the poor sensitivity of the Agilent Bioanalyzer system as the detection was equivalent only to that of the dye staining. In addition, precipitation reactions can be occurred with the fluorescent medium that may suppress detectable proteins during the separation.

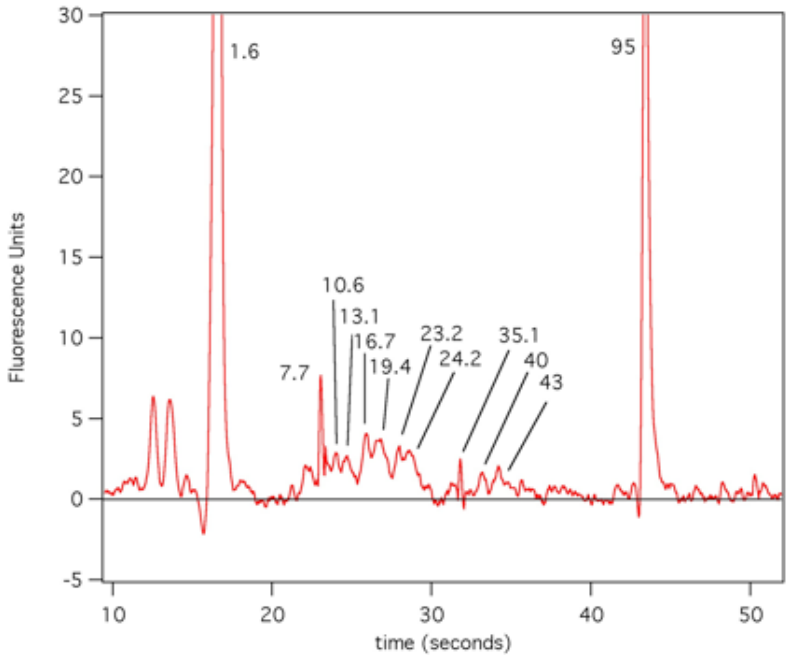


Figure 4.4: Electropherogram of aphid salivary samples obtained from the 2100 Bioanalyzer system. The electropherogram was assigned with representative peak masses according to fluorescently labeled standard protein ladder[†].

4.4 HPLC fraction collection

10 HPLC fractions were collected during the HPLC separations described in Chapter 3, using automatic fraction collector. These fractions are based on defined time intervals. This was performed by pooling three of the 8000 aphid samples that were collected in same optimized conditions. The sample conditions were 24 h feeding of 15% sucrose diet, with no extra calcium added, using screened chambers covered with yellow theater

gel. The collected samples were washed with water/10% acetonitrile for both sample and blank. All three samples and respective blanks were pooled, lyophilized and reconstituted

[†] Electrophoresis experiment and analysis performed by Dr. Amanda Meyer

in 66 µL with same wash solution. The reconstituted sample was injected to HPLC system and collected the fractions in 1-10 numbered time intervals as shown in Figure 4.5. The fractions were lyophilized again for mass spectrometry analysis that was carried out at the Biotechnology Core Facility (Kansas State University; Manhattan, KS).

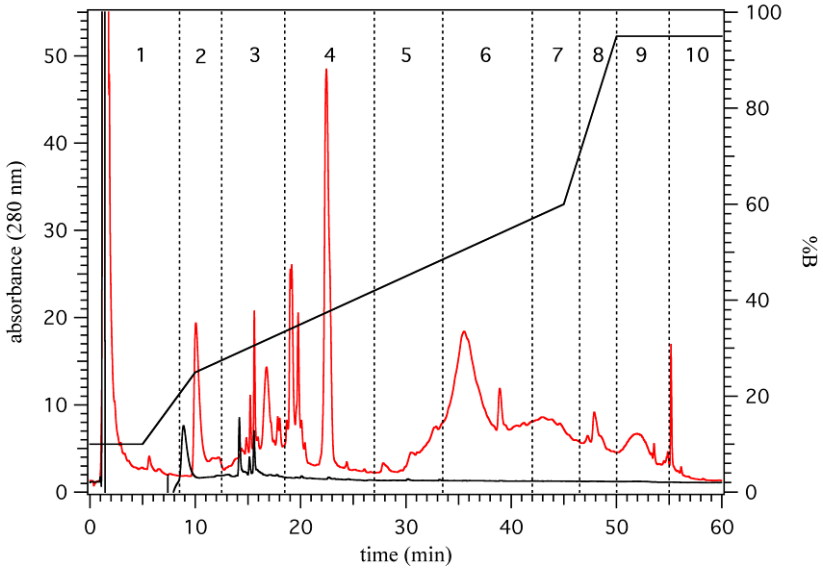


Figure 4.5: RP-HPLC chromatogram collected at 280 nm of ~8000 aphid samples pooled together (red trace) and corresponding blank samples (black trace). The dotted vertical line shows the 1-10 fractions collected. The HPLC gradient is shown as a function of %B in right axis. %B indicates the % Acetonitrile.[†]

[†] Chromatographic experiments and analysis performed by Dr. Amanda Meyer

4.5 Mass spectrometric analysis

4.5.1 Mass spectrometry - sample preparation

Protein digestion was carried out for all the fractions collected from HPLC system. Since the protein concentration was unknown in the samples, trypsin digestion was performed using 10 µL of 10 ng/µL trypsin and 10 µL of ammonium bicarbonate per eppendorf tube fraction. This mixture was kept overnight for digestion. Then the solvent was evaporated and 10 µL of 2,5-dihydroxybenzoic acid (DHB) was added.

4.5.2. Mass spectrometry - experimental

Three types of mass spectrometry were performed from the multiple RP-HPLC fractions taken from the aphid salivary samples. Those are MALDI-TOF, ESI-MS and MALDI-TOF-TOF. The experiments were performed using Biotechnology Core Facility (Kansas State University; Manhattan, KS).

For MALDI-TOF and MALDI-TOF-TOF (MS/MS) analysis (Ultraflex II TOF/TOF; Bruker Daltonics), 2 µL of each fraction was dried on the analysis plate. The remaining fraction of volume was used to perform Electrospray Ionization mass spectrometry (ESI-MS), if necessary, by injecting into nano-HPLC system and sprayed into ESI system (HCT Ultra; Bruker Daltonics).

MALDI-TOF-TOF analysis was used to confirm the peptide sequences from ‘b’ and ‘y’ values. Higher score of b, y values illustrated the correct peptide sequence.

4.5.3 Mass spectrometry - results

MS/MS analysis gave the most promising results out of all of the mass spectrometry methods attempted. As shown in Figure 4.5, the peptide fragment in fraction 5 was selected and performed MS/MS analysis. The corresponding peptide in the fraction 5 was ‘ VDYSAVER’. This was confirmed with ‘b’, ‘y’ values as shown in Figure 4.7.

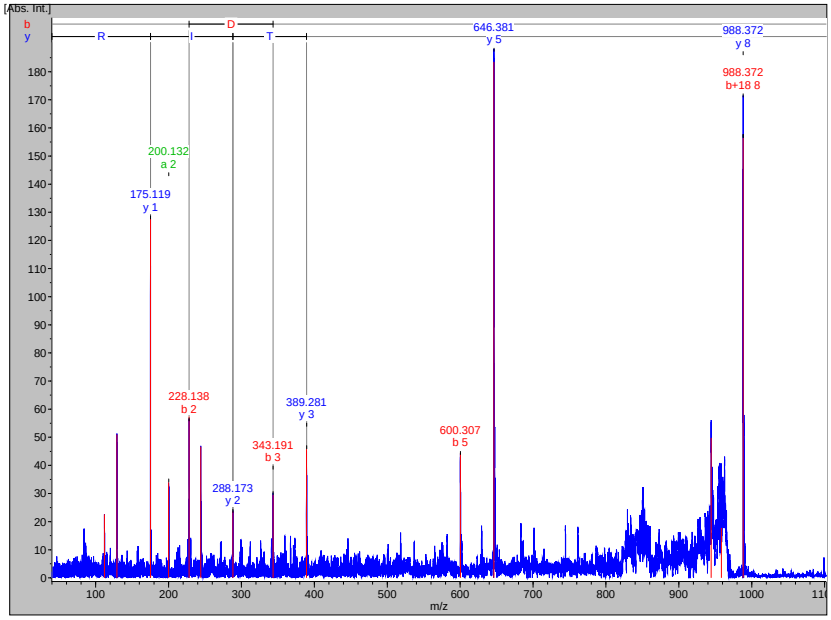


Figure 4.6: MS/MS results of aphid protein sample in fraction 5 of HPLC chromatogram with a peptide sequence ‘VDYSAVER’[†].

[†] MS/MS experiments and analysis performed by Dr. Amanda Meyer

	V	D	Y	S	A	V	E	R	Val	Asp	Tyr	Ser	Ala	Val	Glu	Arg
Ion	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
a	V	D	Y	S	A	V	E	R	72.081	187.108	350.171	437.203	508.240	607.309	736.351	892.452
a-17	V	D	Y	S	A	V	E	R	55.054	170.081	333.144	420.177	491.214	590.282	719.325	875.426
a-18	V	D	Y	S	A	V	E	R	54.070	169.097	332.160	419.193	490.230	589.298	718.341	874.442
b	V	D	Y	S	A	V	E	R	100.076	215.103	378.166	465.198	536.235	635.304	764.346	920.447
b-17	V	D	Y	S	A	V	E	R	83.049	198.076	361.139	448.171	519.209	618.277	747.320	903.421
b-18	V	D	Y	S	A	V	E	R	82.065	197.092	360.155	447.187	518.225	617.293	746.336	902.437
b+18	V	D	Y	S	A	V	E	R	118.086	233.113	396.177	483.209	554.246	653.314	782.357	938.458
c	V	D	Y	S	A	V	E	R	117.102	232.129	395.193	482.225	553.262	652.330	781.373	937.474
x	V	D	Y	S	A	V	E	R	201.098	330.141	429.209	500.246	587.278	750.342	865.369	964.437
y	V	D	Y	S	A	V	E	R	175.119	304.162	403.230	474.267	561.299	724.362	839.389	938.458
z	V	D	Y	S	A	V	E	R	158.092	287.135	386.203	457.241	544.273	707.336	822.363	921.431
i	V	D	Y	S	A	V	E	R	72.081	88.039	136.075	60.044	44.049	72.080	102.054	129.113
	8	7	6	5	4	3	2	1	Arg	Glu	Val	Ala	Ser	Tyr	Asp	Val

Figure 4.7: Peptide fragment tags with notations (a,b,c,x,y,z,i) and corresponding mass fragments of amino acid in the peptide sequence of 'VDYSAVER'. The available MS/MS peaks were bold in red. 'b' and 'y' values indicate individual amino acid fragments. 'b' and 'y' values are covering most of the peptide sequence indicating the reliability of the results[†].

4.6 Proteomics analysis

The mass spectrometric data was interpreted using 'Bruker Daltonics Biotools' software. Four databases searching have been performed for the tryptic digested protein fragments. Those are,

(1) Dr. Gerald Reeck's (KSU) expressed sequence tag (EST) for pea aphids

NCBI databases for

(2) Metazoa

(3) Other Metazoa

(4) Bacteria

[†] Chromatographic experiments and analysis performed by Dr. Amanda Meyer

From the database searches, the peptide matches highest scores were obtained. By integrating these peptide matches, the possible peptide sequences were obtained in different scores. The highest score peptide sequence were selected by a probability test with a threshold value $p > 0.05$. The peptide hits that fall out of the probability box were chosen as acceptable. Then the corresponding peptide database was imported with all the details including molar mass of protein, pI values amino acid sequence etc. With the database amino acid sequence, the experimental peptide fragments were matched and highlighted in red color. The percentage coverage was determined according to the number of peptide matches. From the Basic Local Alignment Search Tool (BLAST) search of the particular protein or a similar protein with structural and functional details were determined.

4.6.1 Results

From EST database, few protein hits were recognized for the aphid samples that consist of similar functionalization to other insects such as *Apis mellifera* and *Ades aegypti*. This result was similar to the results obtained in previous green peach aphid experiments. There were other protein hits such as exopeptidase, angiotensin converting enzyme and few other proteins from *Buchnera*, a bacterial symbiont in the aphid gut.

Out of four databases Other Metazoa NCBI database gave the most promising results from MALDI-TOF-TOF mass fragments. The peptide fragment of fraction 5 was analyzed in MS spectrum as shown in Figure 4.6. The peptide fragment sequence

('VDYSAVER') was confirmed by corresponding MS/MS spectrum as shown in figure 4.7.

Two protein hits were obtained from the Mascot results for fraction 5 peptides as shown in Appendix B and C. The first hit was gi/19358006 (*Acyrthosiphon pisum*) with a nominal mass of 129 851 Da, pI of 4.62 and Mascot score of 102. According to the sequence string it has only 4% percent coverage. This protein was similar to a known protein of *Apis mellifera* (CG1204-PA) according to the BLAST search. The second hit was gi/193598873 (*Acyrthosiphon pisum*), nominal mass 17,570 Da, pI 4.68 and Mascot score 59. It was with a high percentage coverage as 11%. However there was no similar known protein found in BLAST search.

CHAPTER 5: Discussion and Conclusion

5.1 Aphid salivary sampling - large scale vs small scale

Small scale aphid sampling experiments were helpful in preliminary studies. It was illustrated that the aphids could survive in laboratory environment feeding on an artificial diet for 2-3 days. The adult aphids were able to produce nymphs and the nymphs had better survival in the aphid chamber. The small scale experiments showed that the aphids were able to puncture the parafilm and feed on the artificial diet on the other side of the parafilm. In addition they showed the influence of some factors on aphid feeding, such as attraction to yellow color and yellow light, proper ventilation, and impact on different diet compositions on their survival. The small scale sample experiments also demonstrated that aphids have the ability to find the etched channel filled with the diet. Since the collected diet/saliva sample volume was very low (~ 100 μL from 100 aphids) pre concentration was less time consuming.

Large scale aphids salivary sample collection gave more promising results than that of small scale in HPLC. The aphid salivary secretion can be influenced by many other factors that have not been considered throughout the experiments such as age of the aphids, day light time, season of the year and feeding habits etc. Therefore, the total amount of saliva was substantially different in each sample. These factors might be more prominent in the case of small scale sample collection with few aphids so that getting reproducible results was difficult. However in large scale sample collection these factors might have average effect so that overall better sample were able to collect.

5.2 Large scale aphid saliva sample analysis

Compared to the blank samples, the HPLC peak signals obtained for the aphid fed diet indicated salivary secretions into the diet. Aphids salivary sample collection was improved by changing feeding conditions. Peak amplitudes and resolution of HPLC spectra were enhanced when the aphids were fed with good ventilation and light intensities and proper diet compositions. 15% sucrose diet provided the best results out of the three diets compared, while the complete diet gave the worst results. It is interesting that the lower the nutritional quality the higher the salivary secretion. However, the lowest aphid survival rate shown in the 3 amino acid diet was difficult to explain as it is the second best nutritious diet out of three diets. According to these results it can be concluded that the factors that affect the aphid survival and feeding are not necessarily to be the same factors to have for the high salivary secretions.

On the other hand, extra calcium on diet gave negative results on the HPLC signals. This may be due to two reasons. Since a precipitate was obtained some proteins might have been trapped inside the precipitate so that detectable proteins were reduced. Calcium can also cause to suppress the salivary secretion of aphids. It also can be an indication of presence of calcium binding proteins.[15]

During the pre-concentration process there were number of steps that affected for the HPLC peak signals. By following all the steps under sterilized conditions bacterial and human protein contaminations could be minimized. This was illustrated by obtaining the

reproducible peaks over set of samples collected in different times under optimized conditions. Samples were pre concentrated by centrifuge filtration using molecular weight cut off filters. This helped to wash and eliminate most of the sugar from the samples so that high protein concentration can be obtained and avoid clogging of HPLC column due to sugar molecules. Acetonitrile in water was used in the wash solution to increase the solubility of sugar and mix the sugar and amino acids into the aquase. Phosphate buffer was added to try to increase the solubility of sugar but it resulted in more precipitation. This may be due to calcium in the diet that can produce calcium phosphate precipitate.

One disadvantage of using the centrifuge filter tubes was that protein can irreversibly bind to the filters causing sample losses. This was confirmed by comparing the HPLC peak intensities with the number of filtering tubes used to divide the collected sample. If the number of tubes were higher, peak intensity was low and vice versa. Therefore the minimum two tubes were used to centrifuge a single collected salivary sample. It was difficult to use only one centrifuge tube because it took longer time to spin down the washed solution.

5.3 Aphid salivary protein analysis

The sizes of the salivary proteins were estimated using electrophoresis and size exclusion chromatography. The molecular weights obtained from these experiments were ranged between 7 kDa to 150 kDa. These results were similar to previous reports on aphid proteins (6 kDa to 300 kDa).[9]

Mass spectrometric analysis of protein fragments were performed in number of fractions obtained in HPLC spectrums. Some peaks identified were due to bacterial proteins and human skin proteins, mainly keratin. When sterilized conditions were used this was improved with minimum contaminations (Data not shown). However, it was difficult to eliminate contamination completely. Because there were several sources of bacterial contaminations that cannot be avoided such as the bacterial from aphid gut mixing with the collected saliva. In addition, it was found that 24 h feeding was better than 48 h feeding in terms of bacterial contamination because in 48 h feeding the diet was kept longer time at room temperature.

There were number of protein hits obtained from MS/MS data (data not shown). Few of them were with high scores and percentage coverage that matched with proteins of *Acyrthosiphon pisum*. Even though the protein hit for gi/19358006 in fraction 5 came with only 4% coverage, the matched peptide fragments were scattered over the sequence string and therefore it had a high probability to consider as a correct match. Since the total pea aphid genome sequence is not identified it is difficult to characterize these

protein hits with putative domains (e.g. gi/193598873 protein hit for fraction 5) Therefore, further analysis is necessary before any final conclusion is reached.

5.4 Conclusion

Analyzing the aphid salivary secretome is a challenge because of difficulty of sample collection and small scale collection volume. Two approaches were investigated to collect salivary samples. The first method was to use miniaturized devices to collect saliva from a few aphids (10-25) into a small volume of diet solution. The second method was to use large scale sample collection from a large number of aphids (8000) feeding on a larger volume of diet. The small scale device was useful to accomplish preliminary tests on aphid feeding on an artificial diet. However, these experiments did not give reproducible results on normal HPLC separations. The large scale sample method, on the other hand, included lengthy pre-concentration steps but it generated more promising results. It was found that aphid salivary secretions can be increased with the composition of diet and maintaining the aphid chamber with proper ventilation and light intensities. Pre-concentration was accomplished by using molecular weight cut off filters with proper wash solutions. Under sterilized conditions and other optimized conditions reproducible HPLC spectra were obtained. Few of the mass spectrometric analysis revealed salivary protein hits that related to insect proteins. Further analysis is needed to confirm the results combined with more advanced techniques.

5.5 Future work

Further experiments need to be carried out to improve the miniaturized sample collection device with optimizing the aphid feeding in order to achieve real time detection. Investigations of other pre concentration methods may lead to provide better results and shorten the time spend on the experimental procedure. Further studies of aphid feeding behavior may be useful in order to obtain optimized conditions for the aphid salivary secretions. Finally, detail analysis of mass spectrometry and proteomic analysis are needed to obtain reproducible results on aphid salivary proteins.

References

1. Harmel, N.; Letocart, A.; Cherqui, A.; Giordanengo, P.; Mazzucchelli, G.; Guillonneau, E.; Pauw, E. D.; Haubruge, E.; Francis, F.; Identification of aphid salivary proteins : a proteomic investigation of *Myzus persicae*, **2008**, *Insect Mol. Biol.*; 17(2), 165-174.

2. http://www.discoverlife.org/nh/tx/Insecta/Homoptera/Aphididae/Acyrthosiphon/pisum (03/31/2009).

3. http://www.ext.vt.edu/departments/entomology/factsheets/gaphids.html (03/31/2009).

4.http://ipm.ncsu.edu/AG271/forages/pea_aphid.html (03/31/2009).

5. http://w3.rennes.inra.fr/aphidbase/images/cyclePisum3.gif (03/30/2009).

6. Agrios, G.N.; **2005**; *Plant pathology*, 5th ed.; Academic Press.; p275.

7. Miles,P.W.; *Aphid saliva*; **1999**, Biol.Rev; pp 41-85.

8. http://www.york.ac.uk/depts/biol/units/symbiosis/phloem.html (03/30/2009).

9. Will, T., van Bel., J.E.; *Physical and chemical interactions between aphids and plants* **2006**; *J. of Experimental Botany*; 57(4); 729-737.

10. Gaupels, F.; Buhtz A; Knauer T., Deshmukh, S., Waller, F., Bel, A.J.E.V., Kogel, K.H., Kehr, J.; *Adaptation of aphid stylectomy for analyses of proteins and mRNAs in barley phloem sap* **2008**; *J. Experimental Botany*; 59 (12); 3297-3306.

11. http://en.wikipedia.org/wiki/Aphid (04/01/2009).

12. Tjallingii, W.F.; *Phloem sap proteins: their identities and potential roles in the interaction between plants and phloem-feeding insects;* **2006**; *J. Experimental Botany*, 574); 739-745.

13.Golaska, S.; *Deterrence and Toxicity of Plant Saponins for the Pea Aphid Acyrthosiphon Pisum Harris*; **2007**; J.Chem.Ecol.; 33; 1598-1606.

14.Ni,X.,Quisenberry,S.S.,Pornkulwat,S.,Figarola,J.L.,Skoda,S.R.,Foster,J.E., *Possible roles of esterase, glutathione S-transferase, and superoxide dismutase activities in understanding aphid–cereal interactions*; **2000**; *Annals of the Entomological Society of America*; 93(3); 595-601.

15. Mutti,N,.Louis,J.,Pappan,L.,K.,Pappan,K.,Begum,K.,Chen,M.Park,Y.,Dittmer,N., Marshall,J.,Reese,J.,C.,Reeck,G.R.; *A protein from the salivary glands of the pea aphid, Acyrthosiphon pisum, is essential in feeding on a host plant* **2007**;*PNAS*,105(29),9965-9969.

16. Will,T.,Tjallingii,W.F.,Thonnessen,A.,Bel,A.J.E.V.; *Molecular sabotage of plant defense by aphid saliva*; **2007**; *PNAS*; 104(25); 10536-10541.

17. (http://sciencekit.com/parafilm (04/02/2009).

18. http://www.earthlife.net/insects/aphids.html (04/02/2009).

19. Dixon, A.F.D., **1998**, *Aphid Ecology, An optimization Approach*,2nd ed.;Springer; pp8-13.

20. http://books.google.com/books?id=Qfg028GkovEC&pg=PA342&dq (04/03/2009).

21.http://books.google.com/books?id=LpiuJVxJS_AC&pg=PA89&dq=molecular+weigh t+cut+off+filters#PPP1,M1(04/03/2009).

22.http://books.google.com/books?id=tDGYh3x_C4C&pg=PA77&dq (04/03/2009).

23. http://books.google.com/books?id=_kL5nRCN96EC&pg=PA235&dq (04/05/2009).

24. Asenjo, J. A.; *Separation Processes in Biotechnology*; Marcel Dekker, Inc., **1990**; pp.226- 240.

25.http://bdml.stanford.edu/twiki/pub/Haptics/MaterialSelection/Polyethersulfone_PES.p df (04/05/2009).

26. http://www.sartorius-stedim.com/index.php?id=457 (04/05/2009).

27. Nagendran, A.; Mohan,D.; *Protein separation by cellulose acetate/sulfonated poly(ether imide) blend ultrafiltration membranes*; **2008**; *J. of Applied Polymer Science*, 110, 2047-2057.

28. http://en.wikipedia.org/wiki/Electrophoresis (04/05/2009).

29. Allen, R.C., Budowle, B., *Gel Electrophoresis of Proteins and Nucleic Acids*, Walter de Gruyter & Co., Berline, 1994; pp 26-30.

30. Miller, J.M.; (2005); *Chromatography Concepts and Contrasts*; 2nd ed. ; John Wiley & Sons, Inc. NJ.; pp 41-83.

31. http://www.rsc.org/ej/DT/2001/b104502b/b104502b-f3.gif (04/05/2009).

32.http://books.google.com/books?id=3MHPcr7iQBEC&pg=PR5&dq=protein+analysis+HPLC&lr=#PPA14,M1(04/05/2009).

33.http://chromatographyonline.findanalytichem.com/lcgc/data/articlestandard//lcgceurope/232003/59042/article.pdf (04/05/2009).

34. Kirkland, J.J. *Superficially porous silica microspheres for the fast high-performance liquid chromatography of macromolecules;* (1992); *Anal.Chem.*; 1882(64); 1239-1245.

35. http://en.wikipedia.org/wiki/Mass_spectrometry (04/05/2009).

36.http://books.google.com/books?id=7IUVi06u0TQC&printsec=frontcover&dq=LC/M# (04/05/2009).

37. http://images.the-scientist.com/content/figures/0890-3670-050117-32-1-1.jpg (04/05/2009).

38.http://www.astbury.leeds.ac.uk/facil/MStut/mstutorial_files/image020.jpg (04/05/2009).

39. http://books.google.com/books?id=Eb0gfuHhriAC&pg=PA120&dq (04/05/2009).
40. http://en.wikipedia.org/wiki/MALDI-TOF (04/05/2009).

41. http://books.google.com/books?id=cmPlgm7SnnMC&pg=PA40&dq (04/06/2009).

42. http://books.google.com/books?id=GvuvA3EMIHAC&pg=PA149&dq=LC/MS (04/06/2009).

43. http://en.wikipedia.org/wiki/Protein_mass_spectrometry (04/06/2009).

44. http://en.wikipedia.org/wiki/MS/MS (04/06/2009).

45.http://books.google.com/books?id=ADjA1jAJm9IC&pg=PA64&dq=mascot+peptide#PPA64,M1(04/06/2009).

46. http://www.nature.com/nmeth/journal/v4/n10/fig_tab/nmeth1088_F1.html (04/06/2009).

47.http://www.matrixscience.com/search_intro.html (04/06/2009).

48.http://books.google.com/books?id=qSGK11GH6fcC&printsec=frontcover&dq=Mass+spectrometry+proteomics#PPA22,M1(04/06/2009).

49. http://www.chem.agilent.com/Library/usermanuals/Public/G2938-90009_HiSenseProtein_Labeling.pdf (04/06/2009)

Appendix A

Modified Akey/Douglas & Prosser Complete Diet (100mL)

(Ref.: Akey, D.,H., Beck,S.D,(1971),Annals of the Entomological Society of America, 64,(2),353-356)

(1) Dissolve Sucrose in ~2/3 of total volume (~60mL)
(2) Add in the order of amino acids, vitamins, salts/buffer/sterol (crystalline)
(3) Add trace metals from pre-made solutions (100x dilution)
(4) Weigh and dissolve Magnesium Chloride and Choline Chloride
(5) Add ~ 1.25mL of 1.75M KOH and stir for 1-3 h
(6) Slowly, over the period of one half hour to an hour, raise the pH to 7.5 by adding 1.75M KOH.

Sugars		
Sucrose	0.5M	17.1g
Amino acids and amides		
Essential amino acids		
L-Arginine HCl	12.5mM	263.34 mg
L-Histidine(Free Base)	7.5 mM	116.4 mg
L-Isoleucine (Allo Free)	7.5mM	98.4 mg
L-Leucine	7.5mM	98.4 mg
L-Lysine HCl	7.5mM	137 mg
L-Methionine	2.5mM	37.3 mg
L-Phenylalanine	2.5mM	41.3 mg
L-Threonine (Allo Free)	7.5mM	89.3 mg
L-Tryptphane	2.5mM	51.1 mg
L-Valine	7.5mM	87.8 mg
Total essencial amino acids	65 mM	
Nonessential amino acids		
L-Alanine	5mM	44.5mg
L-Asparagine (Anhydrous)	12.5mM	165.15mg
L-Aspartic acid C.P.	12.5mM	166.4mg
L- Cystein HCl monohydrate	2.5mM	39.4mg
L-Cystine	0.2mM	5mg
Gamma amino butyric acid	2mM	20mg
L-Glutamic acid	7.5mM	110.33mg
L-Glutamine(Commercial)	15mM	219.2mg

Glycine	1mM	7.5mg
L-Proline	5mM	57.6mg
L-Serine	5mM	52.5mg
L-Tyrosine	0.5mM	9mg
total nonessential amino acids	68.7mM	896.58mg
Total amino acids	133.7mM	

Vitamins

p-Aminobenzoic acid	0.73mM	10mg
Ascorbic acid	5.68mM	100mg
Biotin	6uM	0.1mg
D-Calcium pantothenate	0.21mM	5mg
Folic acid	22uM	1mg
i-Inositol (meso) dihydrate	1.39mM	50mg
Nicotinic acid	0.812mM	10mg
Pyridoxine HCl	0.121mM	2.5mg
Thiamine HCl	74uM	2.5mg

Salts, Buffers and Sterol

Calcium citrate	0.175mM	10mg
Cholesterol benzoate	50uM	2.5mg
Potassium phosphate monobasic	18.37mM	250mg
Sodium chloride	0.217 mM	12.7mg

Trace Metals (prepared and added from 100x stocks of individual metals)

Cupric chloride	14uM	0.254mg
Ferric chloride	49uM	1.336mg
Manganese (II) chloride	40uM	1.271mg
Zinc sulfate	30uM	0.86mg

Special (these reagents should be stored in dessicator and used immediately

Magnesium chloride.6H20	9.837mM	200mg
Choline chloride	3.579mM	50mg

Appendix B

Mascot Search Results

Protein View

Match to: **gi|193580006** Score: **102**
PREDICTED: hypothetical protein [Acyrthosiphon pisum]
Found in search of DATA.TXT

Nominal mass (M_r): **129851**; Calculated pI value: **4.62**
NCBI BLAST search of gi|193580006 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Acyrthosiphon pisum

Variable modifications: Oxidation (M)
Cleavage by Trypsin/P: cuts C-term side of KR
Sequence Coverage: **4%**

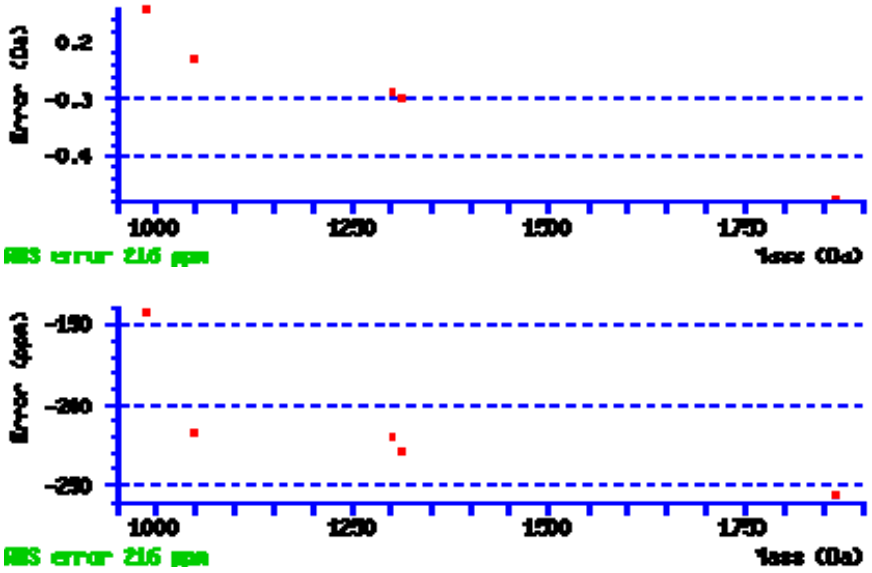
Matched peptides shown in **Bold Red**

```
     1 MRSVLILCVI IAAAVACPVS KTKDCSCGLP KICPSTWKIK TFDSQCETLA
    51 FQGTWFLQMA TPTYIDQQNP LKTGLFCNSY PCTNNQLIFK DTTPCDDTDY
   101 NTEYEVIDSS YNIYTQCTET QIALLCPAYG SSSKYAIFNV GSEKYYTNPI
   151 AEKYPFVDKD LFRHKIGQGS CEREYTVAVI GADDCWKEYM VLVVINQYDN
   201 FFGGDEYITW VLTRDVNPDW STYDKAYNDI KGSGLCPNYL VSVDHSFESM
   251 TGPSMAVPSM APSVAVPSMA PTMPGDVDSM VQKTSVSTTS ATKSISTDCG
   301 STVTSSSTST TTTSTVIIDK SSDFSSIYDI GPCDLYSPYE GLQIYKNLDK
   351 ETIRRAFSGN YYMTQATPCS FYDTPKSKVG LLNTCFPACG MQLCFDDASI
   401 DDWDCNTPRM VMDRGYNMRT GEVHMTRSYI SSAYSDDHPF GTVTYAFHSE
   451 GYYDVPIDET DCLPLDGLIC KPPSDIYKNQ IIANIIGYKD NDYLMFCIAN
   501 RYKNPLFPKK QVPLVYCYTR ERIPSQDTMN SITQEMLRCG LNPNYLMKID
   551 QSKTIEEVFT FDKTYYESTV TSTSTTKSVV STRGLSIGSS PLQGLPSICP
   601 PDWNVKTFDN QYATIAFQGI WNVQMTTPTY INGNNPLKTG LFCNSYPCTN
   651 NQLIFKDTTP CDDTDYNTEY EMTDSSYNLY TQCTETQKAL LCPAYGSCSQ
   701 YAIFNVGSEK YYNTAIIEKY PFIDKDLFRH KIVQGTCERE YTVAVIGADD
   751 CWEEYMVLAV INQYDNVFGG NKYIIWVVTR DANPKWSTYK KAYEDIERSG
   801 LCPNYLVSVD HSLESMTGPS MAPSVAVPSI APTMPGDVDS MVQKTSVSTT
   851 SATKSISTDC GSTVTSSSTS TTTTSTVIID KSSDFSSLYE LGQCDLYKDI
   901 QIYKNLDKET IRRALSGKYY MTQATPCSYY NSPNSRVGIL NTCFPACGMQ
   951 LCFDDTSIDD WDCNTPRMVM DRGYDMRTGE VQLTRSYISS VYSDDHPFGT
  1001 VTYAFHSEGY YEGPIDEMDC LPLDGMICKP PSDIYKNQII ASIIGYKDND
  1051 YLIFCIANKY KNPLFSSTPV NQVIAYTRER VPSKETIKSM TQELLQCGYN
  1101 PNYLIKIDQT MYMDDDYVFE SSYYESQTSC WSSSSSSCSS STSSSTTISS
  1151 SSSSSVSISC D
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
154 - 163	1299.3884	1298.3811	1298.6659	-0.2848	1	K.YPFVDKDLFR.H (Ions score 18)
347 - 354	988.4006	987.3934	987.5349	-0.1415	1	K.NLDKETIR.R (Ions score 33)
523 - 538	1863.4184	1862.4112	1862.8866	-0.4754	0	R.IPSQDTMNSITQEMLR.C (Ions score 20)
720 - 729	1313.3889	1312.3816	1312.6816	-0.3000	1	K.YPFIDKDLFR.H (Ions score 20)
773 - 780	1049.3863	1048.3791	1048.6069	-0.2279	0	K.YIIWVVTR.D (Ions score 12)



LOCUS XP_001943898 1161 aa linear INV 02-JUL-2008
DEFINITION PREDICTED: hypothetical protein [Acyrthosiphon pisum].
ACCESSION XP_001943898
VERSION XP_001943898.1 GI:193580006
DBSOURCE REFSEQ: accession XM_001943863.1
KEYWORDS .
SOURCE Acyrthosiphon pisum (pea aphid)
ORGANISM Acyrthosiphon pisum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;
Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.
COMMENT MODEL REFSEQ: This record is predicted by automated computational
analysis. This record is derived from an annotated genomic sequence
(NW_001933671) using gene prediction method: GNOMON, supported by

```
            EST evidence.
            Also see:
                Documentation of NCBI's Annotation Process

            COMPLETENESS: full length.
FEATURES             Location/Qualifiers
     source          1..1161
                     /organism="Acyrthosiphon pisum"
                     /strain="LSR1"
                     /db_xref="taxon:7029"
                     /note="from the clonal line LSR1.AC.F1, produced
from a
                     single generation of inbreeding of strain LSR1"
     Protein         1..1161
                     /product="hypothetical protein"
                     /calculated_mol_wt=129805
     CDS             1..1161
                     /gene="LOC100169243"
                     /coded_by="XM_001943863.1:47..3532"
                     /db_xref="GeneID:100169243"
```

Mascot: http://www.matrixscience.com/

Appendix C

Mascot Search Results

Protein View

Match to: **gi|193598873** Score: **59**
PREDICTED: hypothetical protein [Acyrthosiphon pisum]
Found in search of DATA.TXT

Nominal mass (M_r): **17570**; Calculated pI value: **4.68**
NCBI BLAST search of gi|193598873 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Acyrthosiphon pisum

Variable modifications: Oxidation (M)
Cleavage by Trypsin/P: cuts C-term side of KR
Sequence Coverage: **11%**

Matched peptides shown in **Bold Red**

```
     1 MAFKEIAMFS SLVVVTIISY NVLEGNAQSI KPLIDQDYCK VKASLYDLGE
    51 IGMNLMDDSQ TLNDMQREYF AGKVDYSAVE RARNELNQTK NKLFLKLIKY
   101 IWATNEFEPT VNYQTADPQK LYKAMDDLEN YKDEMHADLL NSMSPTLQPT
   151 VVGA
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
74 - 81	938.3313	937.3241	937.4505	-0.1264	0	K.VDYSAVER.A (Ions score 39)
82 - 90	1073.3870	1072.3797	1072.5625	-0.1828	1	R.ARNELNQTK.N (Ions score 20)

Error: try setting browser cache to automatic.

Error: try setting browser cache to automatic.

```
LOCUS       XP_001951871             154 aa            linear   INV 02-
JUL-2008
DEFINITION  PREDICTED: hypothetical protein [Acyrthosiphon pisum].
ACCESSION   XP_001951871
VERSION     XP_001951871.1  GI:193598873
DBSOURCE    REFSEQ: accession XM_001951836.1
KEYWORDS    .
SOURCE      Acyrthosiphon pisum (pea aphid)
  ORGANISM  Acyrthosiphon pisum
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta;
Pterygota;
            Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha;
Aphidiformes;
            Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.
COMMENT     MODEL REFSEQ:  This record is predicted by automated
computational
            analysis. This record is derived from an annotated genomic
sequence
            (NW_001930867) using gene prediction method: GNOMON,
supported by
            EST evidence.
            Also see:
                Documentation of NCBI's Annotation Process

            COMPLETENESS: full length.
FEATURES             Location/Qualifiers
     source          1..154
                     /organism="Acyrthosiphon pisum"
                     /strain="LSR1"
                     /db_xref="taxon:7029"
                     /note="from the clonal line LSR1.AC.F1, produced
from a
                     single generation of inbreeding of strain LSR1"
     Protein         1..154
```

```
                     /product="hypothetical protein"
                     /calculated_mol_wt=17450
     CDS             1..154
                     /gene="LOC100167427"
                     /coded_by="XM_001951836.1:59..523"
                     /db_xref="GeneID:100167427"
```

Mascot: http://www.matrixscience.com/