Western Graduate&PostdoctoralStudies

Electronic Thesis and Dissertation Repository

9-5-2024 2:00 PM

Application of High-Resolution Nano-Scale Liquid Chromatography Tandem Mass Spectrometry to the Study of S. endobioticum-induced Infection In Cultivated Potatoes

Daria Baskova,

Supervisor: Sumarah, Mark W., *Agriculture and Agri-Food Canada*
Co-Supervisor: Yeung, Ken K.-C., *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Chemistry
© Daria Baskova 2024

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Analytical Chemistry Commons, and the Biochemistry Commons

Recommended Citation

Thesis Abstract

Potato wart is an infection of cultivated potato plants caused by the soil-borne biotrophic fungus *Synchitrium endobioticum*. The main symptoms of infection are formation of warts on the tuber tissue and the production of biflagellated mobile spores that undergo long-lasting dormancy periods. Genomic studies have significantly contributed to an overall understanding of the fungal life cycle, including the discovery of the first *S. endobioticum* avirulence genes, and the identification of multiple pathotypes. However, the molecular pathways involved in host-pathogen interactions during potato wart infection and the factors contributing to resistance in plants remain unknown. In this work, a combination of proteomic analysis and protein structure analysis resulted in the discovery of protein classes associated with immune response-based changes post-infection. Comparative proteomics of four potato varieties highlighted the differences in resistance between groups. Co-immunoprecipitation experiments led to detection of interaction partners of recombinant AvrSen1 protein, providing insight into its localization and functions.

Keywords: potato wart, proteomics, high resolution mass spectrometry, protein expression, effector molecules, co-immunoprecipitation

Summary for Lay Audience

The fungal kingdom is estimated to contain around 11 million different species that vary in structure, functions, and preferred living environments. [1,2] Many different types of fungi grow on plants and can have either beneficial or harmful effects on the growth and development of certain crops, including potatoes. One of the fungal infections that results in major crop losses in Canada and worldwide is potato wart. The pathogen that causes this infection is not well-understood and the factors that play a role in disease progression remain unknown. Previous research indicated that the fungus responsible for potato wart is capable of producing chemicals that facilitate the infection. Identifying these fungal molecules may aid in understanding the mechanisms the fungus is using to attack potato plants. Mass spectrometry (MS) is an analytical technique that determines the structure and mass of molecules that are present in the sample. The components of the sample mixture are broken down into fragments that are then detected by the instrument. These generated fragments are characteristic of their overall chemical structure, where fragmentation patterns can be compared and matched to libraries of previously identified chemicals to confirm their identity. In this study, mass spectrometry was used to compare the differences in protein levels between uninfected and infected tubers. Overall, around 600 proteins were identified in each variety. Following data analysis, it was determined that different potato varieties respond to infection differently.

In addition to protein analysis, the other goal of this project was to express, purify, and characterize one of the effector proteins that is produced by the fungus. By mixing the purified fungal protein with protein extracts from tuber tissue, we were able to identify protein-protein interactions that occur between the plant and the fungus.

Co-authorship Statement

Chapter 2 – Comparative proteomics analysis of control and potato wart-infected *Solanum tuberosum* tuber samples

The experiments, experimental design, and data analyses were performed by the author. Melissa Antoun (Canadian Food Inspection Agency, Charlottetown, PEI) collected and provided the tuber samples. Dr. Mark Sumarah (Agriculture and Agri-Food Canada, London, ON) and Dr. Christopher Garnham (Agriculture and Agri-Food Canada, London, ON) assisted with sample processing. Dr. Justin Renaud (Agriculture and Agri-Food Canada, London, ON) provided assistance with proteomics data analysis and methodology. Zoltan Richter-Bisson (University of Western Ontario, Department of Chemistry) helped with data analysis. Megan Kelman (Agriculture and Agri-Food Canada, London, ON) assisted with metabolomics data analysis and Dr. Aaron Simkovich (University of Western Ontario) helped with protein annotation. Serena Deelen (Agriculture and Agri-Food Canada, London, ON) helped to process the samples. Dr. Mark Sumarah (Agriculture and Agri-Food Canada, London, ON) and Dr. Ken Yeung (University of Western Ontario, Department of Chemistry and Biochemistry) supervised the project.

Chapter 3 – Expression of AvrSen1 and its characterization

All experimental work and data analyses were conducted by the author. Dr. Mark Sumarah (Agriculture and Agri-Food Canada, London, ON) and Dr. Ken Yeung (University of Western Ontario, Department of Chemistry and Biochemistry) supervised the project. Dr. Christopher Garnham (Agriculture and Agri-Food Canada, London, ON) was a co-supervisor on this project and assisted with ordering of amino acid sequences. Dr. Patrick Telmer and Angelo Kaldis provided assistance with method development, technical assistance and training for the protein expression and purification. Dr. Justin Renaud (Agriculture and Agri-Food Canada, London, ON) assisted with method development for bottom-up analysis.

Chapter 4 – Investigation of protein interaction partners of AvrSen1

All experimental work and data analyses were conducted by the author, with the help of Serena Deelen (Agriculture and Agri-Food Canada, London, ON) with sample processing and

Acknowledgments

Firstly, thank you to Dr. Justin Renaud, who I met during my third year of undergraduate studies and who introduced me to the field of instrumental chemistry. Thank you for teaching me that failure is common and that being busy does not mean being productive. Most of all, thank you for your patience and guidance and for answering millions, if not billions, of my stupid mass spectrometry-related questions. I am constantly fascinated by your enthusiasm for research, and I am so thankful for everything I got to learn from you.

Thank you to all the incredible technicians and both graduate and undergraduate students I got to learn from and work with during my graduate studies at Dr. Sumarah's lab, with a special thanks to Megan Kelman and Natasha DesRochers. Thank you for sharing addiction to caffeine with me and for dealing with so many of my impostor syndrome freakouts. Thank you for reassuring me (a lot) and thank you for proof-reading both my emails, conference abstracts, my first-year report and now this thesis. Meganula, I truly believe you are one of the kindest people I know, and I can't begin to explain how grateful I am to have a friend like you. Natasha, thank you for helping me deal with all the stress that comes with grad school by going to yoga with me and yapping afterwards. I am so grateful for all support you two provided me with in the last two years, I truly couldn't have done it without you. I'd also like to say a special thanks to my undergraduate student, Serena Deelen. Thank you for being patient with me and thank you for all your hard work (and also all the incredible baking you did). I very much appreciate all the help you have provided with my projects.

Thank you to my co-supervisor, Dr. Ken Yeung, for making sure my studies at Western go smoothly. I'd like to thank you for your assistance on my projects and for all your help with organizational moments. I appreciate all the feedback you have given me on my work, and I am very grateful that I was able to go to Trent Conference with your help. I am thankful to have you as my co-supervisor and I am looking forward to continue working with you throughout my Ph.D.

Thank you to everyone who I got a chance to work with in Dr. Christopher Garnham's lab. Chris, thank you so much for your understanding as I worked with AvrSen1. Thank you for believing in this project and for all your help with sample collection and analysis. An especially big thank you to Dr. Patrick Telmer and Angelo Kaldis for teaching me all there is to know about bacterial

transformation and protein expression. I appreciate all of your assistance and help with the biochemistry part of my project and for bearing with me when I asked you a thousand questions over and over again.

I'd also like to thank my family and friends for all love and support you have given me over these past few years. Thank you for letting me vent when I needed it and thank you for comforting me when I needed a shoulder to cry on. Most of all, thank you for being there for me during one of the most amazing and challenging moments of my life over these last two years. I would like to thank my mom, my stepdad and my sister for always believing in me and for letting me follow my heart, even though it means we live on different continents now. I am eternally grateful for your support and encouragement to pursue academic research. Mom, you are the reason I am who I am now, and I am so grateful to have such a great example in life. Alyssa, Kennedy, Laura, John, Sashulkin (aka Alexandra Troitskaya), and Lexie, I love each and every one of you and I wouldn't be where I am without you.

Lastly, I want to thank my supervisor, Dr. Mark Sumarah. Thank you for believing in me when I didn't, thank you for pushing me when you knew I could do more and thank you for all your understanding and patience throughout the years. I am constantly amazed by the work that you do, and I am so grateful to be a part of your lab. I know I might not be the easiest person to work with at times as I like to ask millions of questions, have my fair share of anxiety attacks, and basically require constant reassurance, but I am thankful you were able to look past that. I couldn't have asked for a better supervisor, and I hope I get to achieve at least half of what you have.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Symbols and Abbreviations

AABG = Advanced Active Beam Guide

AC = alternating current

ACN = acetonitrile

AGC = Automatic Gain Control

amu = atomic mass unit

Avr factor = avirulence factor

β-ME = β-Mercaptoethanol

BEH = ethylene-bridge hybrid material

CEM = chain ejection model

CFIA = Canadian Food Inspection Agency

CID = collision Induced Dissociation

Co-IP = co-immunoprecipitation

CRM = charged residue model

DC = direct current

DDA = data-dependent acquisition

DIA = data-independent acquisition

DTT = dithiothreitol

ECD = electron capture dissociation

EDTA = ethylenediaminetetraacetic acid

ESI = electrospray ionization

ETI = effector-triggered immunity

FA = formic acid

FC = fold change

FFT = fast Fourier transform

GST = glutathione S-transferase

GO analysis = gene ontology analysis

HCD = Higher Energy Collision Dissociation

HLB = hydrophilic-lipophilic copolymer

HPLC = high-performance liquid chromatography

HR = hypersensitive response

IAA = iodoacetamide

IDA = iminodiacetic acid

IEM = ion evaporation model

IMAC = immobilized metal affinity chromatography

IPTG = β-thio-galactosidase or isopropyl β-D-1-thiogalactopyranoside

IT = injection time

LB = Luria Broth (LB)

LC-MS = liquid chromatography - mass spectrometry

LC-MS/MS = liquid chromatography–tandem mass spectrometry

LFQ = label-free quantification

LS = light scattering

m/z = mass-to-charge ratio

MBP = maltose-binding protein

MS = mass spectrometry

NCBI = National Center for Biotechnology Information

NGC = Next Generation Chromatography

NTA = nitrilotriacetic acid

NusA = N-utilization substance protein A

OD_{600} = optical density measurement at 600 nm wavelength

PAMP = pathogen-associated molecular pattern

PBS = phosphate-buffered saline

PCR = polymerase chain reaction

PEG = polyethylene glycol

pI = isoelectric point

PTI = PAMP-triggered immunity

PTMS = post-translational modifications

R gene = resistance gene

RBC = Russet Burbank control

RBI = Russet Burbank infected

RF = radio frequency

RI = refractive index

RPLC = reverse-phase liquid chromatography

SDC = sodium deoxycholate

SDS = sodium dodecyl sulfate

SDS-PAGE = sodium dodecyl-sulfate polyacrylamide gel electrophoresis

SEC = size-exclusion chromatography

SPE = solid-phase extraction

ssDNA = single-stranded DNA

SUMO = small ubiquitin related modifier

TC = Targhee Russet control

TCEP = Tris(2-carboxyethyl)phosphine

TEV = tobacco etch virus

TFA = trifluoroacetic acid

TINW = Targhee Russet infected non-warted

TIW = Targhee Russet infected with warts

TLC = thin-layer chromatography

Tris = tris(hydroxymethyl)aminomethane

TRX = thioredoxin

UV = ultraviolet

VOCs = volatile organic compounds

1 Chapter 1 - Introduction

1.1 Fungal infections in plant material

The fungal kingdom is comprised of approximately 11 million species that are extremely diverse in nature and are widely distributed in the ecosystem, with 150,000 species previously characterized. [1-4] Fungi can utilize numerous types of surfaces for growth purposes and are capable of infecting various eukaryotic cells. However, while fungal diseases in both animals and humans are less common, the majority of plant infections are fungi-induced. [4] More specifically, more than 120 fungal genera, or 10,000 species, have been previously identified to cause diseases in plants. [4-7] The first observation of the interaction of plant parasitic fungi with the surface of the host plant was observed in 1866 by Anton De Bary[8], who recorded the change in the hyphae structure post-germination. It is important to highlight that fungi have a limited range of susceptible hosts they can infect, and require specific environmental conditions (e.g., temperature, moist content of the plant surfaces and water availability) to occur for infection to proceed. [4, 5]

Based on the nutrient acquisition strategy, fungi are categorized into biotrophs, necrotrophs and hemitrophs. [8, 9] Biotrophic fungal plant pathogens rely on plant tissue as the main source of nutrients needed for their survival and growth. [6] Necrotrophs promote tissue necrosis in plants to use the resulting necrotic tissue as a food source. [6, 8] Hemitrophs, or hemibiotrophs, establish a biotrophic relationship with the host during the initial stage of infection or until sporulation, followed by transition to necrotrophic lifestyle to acquire nutrients from dead cells that underwent necrosis. [8] While all three types of fungi are common plant pathogens, biotrophic fungal species are widely distributed and cause numerous crop infections, including potato wart. [10] Plant diseases caused by biotrophic fungi are highly destructive, resulting in global food insecurity while causing negative economic impact. [6, 11] The harmful effects of biotrophic pathogenic fungi on the crop yield resulted in the need to study cellular, molecular, and resistance mechanisms involved in plant-pathogen interaction. [7, 9]

The main steps of biotrophic pathogenesis includes attachment to the host cell, penetration inside the cell through the cell wall and the plasma membrane, and proliferation. [8] The formation of infection structures is controlled on both a genetic level and through numerous regulatory fungal pathways. Fungal infection can occur in intracellular (within cells), intercellular (between cells),

extracellular (outside the cell) layers, or in plant epidermis, which is an outer layer of plant organs. [8] Biotrophic fungi are separated into obligate biotrophs, endomycorrhizae, facultative biotrophs, and hemibiotrophs. [12] All four groups establish symbiotic relationships with plant organisms to survive and complete their life cycle. [12] The growth and infection mechanisms of both obligate and non-obligate biotrophic fungi are associated with the development of infection structures, restricted secretory activity (limited activity of cell-wall degrading enzymes in the host cells), continuous suppression of the host immune defense, presence of interfacial layers that are rich in sugars and proteins that separate the plasma membrane of the pathogen and the plant, as well as formation of haustorium. [6, 8, 12] Haustorium represents a specialized fungal organ that plays a role in disease progression and nutrient uptake. [8, 12-15] Other functions of haustoria include water uptake, signalling, and inhibition of recognition by the host. [12, 16-20] Both haustorial membrane and hyphae tissue (filamentous structure of a multicellular fungus) are separated through the intracellular membrane of the host by the interfacial matrix. [12] The fungal infection is considered successful if the pathogen has penetrated both the plant cuticle and the cell wall and has entered the cell. [6, 13, 21] The adhesion of fungal tissue, such as spores or hyphae, to the surface of the plant is facilitated by the establishment of hydrophobic interactions between the cuticle and the pathogen material. [4] Spores are reproductive cell units in fungi that promote fungal dispersion and survival. Fungal spores can remain in the dormancy state until environmental conditions become favourable for their proliferation. [7] Some spores are capable of synthesising mycotoxins and hydrolase or depolymerase enzymes to promote wearing off of the cuticle layer, which enhances the fungal adhesion. [22-25] Spread of sporulated material can occur by wind, soil and water. Spores can also be transferred to the neighboring farm fields by animals. Additionally, certain fungal species are capable of penetrating the host cell wall by increasing the turgor pressure of the hyphae up to 80 bar. [26] During the infection, fungal hyphae pumps excess water into the cell which results in high osmotic pressure. The generated pressure is then used to push fungal hyphae through the cuticle of the plant cell.

The plant-pathogen interaction is accompanied by the production of numerous proteins and small-molecular weight molecules by both organisms. These compounds are synthesized as a result of long and complex co-evolution between the plant and fungus and play an important role in plant defense mechanisms. [4] Plants adapt to inhibit fungal infections though synthesis of pathogenesis-related proteins, cell wall strengthening, induction of necrosis through hypersensitive response, or

oxidative burst. [4, 9] In response, fungal pathogens evolve to overcome those defensive mechanisms and inactivate certain plant defense enzymes, causing their reductancy. Fungi are capable of producing effector molecules with both narrow and broad ranges to facilitate the infection in one or multiple hosts, respectively. As a result, two types of defense responses are initiated in plants, which includes host-specific resistance and non-host resistance, respectively. [4] The host-specific resistance is established through recognition of fungal avirulence gene products by host resistance proteins. [9] The non-host resistance is accomplished in plants through expression of pathogenesis-related proteins and antifungal proteins. [27] While pathogenesis-related proteins are induced in response to numerous biotic stressors and may exhibit antimicrobial activity, antifungal proteins are produced in plants following fungi-triggered infection. These proteins have a wide-ranging action. [27] Some examples include chitinases, thaumatin-like proteins, chitin-binding proteins, defensins, and protease inhibitors. [27] Chitinases and chitin-binding proteins degrade or bind to chitin, respectively, which is a predominant component of fungal cell walls. [28] This results in the increased permeability of the plasma membrane of the pathogen. [27] Similarly, thaumatin-like proteins affects the permeability of the fungal cell membrane due to their binding affinity to β-1,3-glucan, one of the building blocks of fungal cell wall. [29] In contrast, the mechanism of fungal inhibition used by defensin proteins remains understudied. [27] In terms of protease inhibitors, they play an important role in plant defense as they bind to proteases secreted by fungi and block their catalytic sites. [30]

Plant pathogenic fungi can compromise global food security, which has been previously observed during potato late blight outbreak or *Pyricularia oryzae*-induced infection in rice. [7] Additional challenges associated with fungal plant infections are caused by the high persistence and fungicide resistance of the pathogen. A further complication is a lack of conventional detection methods due to the shared similarity of plant disease symptoms between different fungal infections. [7] Current methods for fungal analysis include visual examination of plant material for the presence of fungal growth, strain isolation through microbiological applications (i.e., culturing, plating, and isozyme isolation), and biochemical analysis. [7] Fungal detection can also be accomplished using immunology-based and polymerase chain reaction (PCR)-based techniques, spectroscopic imaging, and mass spectroscopic analysis of volatile organic compounds (VOCs). [7]

1.2 Potato wart

Potato wart, also known as potato canker, is a fungal disease of cultivated potato plants (*Solanum tuberosum*). [31] The causal agent for this infection is *Synchitrium endobioticum*, which is an obligate biotrophic soil-borne fungus that was first described in 1896 by Schilberzsky. [31, 32] Since then, potato wart has been reported across multiple continents and is currently considered one of the most predominant quarantine pathogens that affect potatoes. In Canada, potato wart was first detected in Prince Edward Island in 2000. [33] This resulted in the implementation of Potato Wart Domestic Long Term Management Plan that outlines sampling requirements and restrictions associated with limiting the risk of spread of potato wart. Since the first detection of potato wart in Canada, the export of Canadian potato to U.S. has been suspended twice, resulting in huge economic losses. [33] More specifically, it is estimated that the industry lost $50 million CDN in sales following the 5-month closure of U.S.- Canada border from late 2021 until early spring of 2022. [33, 34]

The main disease symptom of *S. endobioticum*-induced infection is characterized by the growth of wart-like malformations (Figure 1A-D) on plant meristematic tissue. [31] This includes tubers, sprouts, and stolons (Figure 1E); however, the root system usually remains unaffected. Warts have also been detected on aboveground plant organs, such as leaves and shoots. *S. tuberosum* represents the main host for the disease; however, potato wart can also affect other Solanaceous plants, such as tomatoes, cherries, and nightshades, under certain environmental conditions. [31]
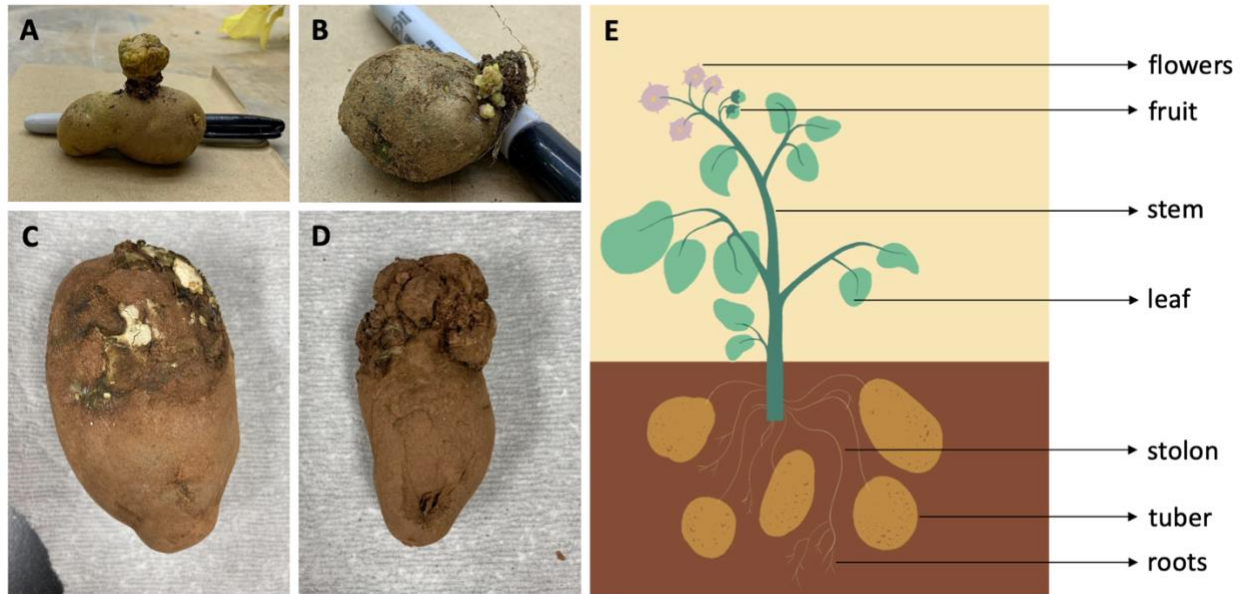
Figure 1. Variations in appearance of warted material in different potato varieties and a schematic diagram of a potato plant. Potato wart may affect different potato varieties differently, causing unique infection symptoms. Figures 1A, 1B and 1C show the formation of warted material on tubers of Russet Burbank potato variety. Depending on the farm field and the stage at which the tuber was collected, severity of the infection varies even among one variety. The *S. endobioticum-*associated warty malformation on Targhee Russet tuber is shown on Figure 1D. Figure 1E shows a schematic diagram of a potato plant, including the main types of meristematic tissue.

S. endobioticum belongs to the *Synchitrium* genus, which is a part of the Chytridiomycota (chytrid) fungal division. [31] Fungal species in this lineage inhabit aquatic and terrestrial environments high in moisture and are categorized into saprotrophic and pathogenic species. [35-37] One of the main characteristics of chytrid fungi is the ability to produce zoospores, which are motile biflagellated spores. [31] Zoospores are short-lived germinated spores that are capable of migrating over short distances to promote dissemination of fungal infection. [31, 38] Zoospores are approximately 1.5-2 μm in radius and contain flagella that enable their mobility through soil. [39-41]

In potato wart, zoospores play an important role in invading the host. The infection of host cells with *S. endobioticum* occurs through encystment on the plasma membrane. [31] This occurs when resting *S. endobioticum* spores undergo germination, resulting in release of haploid zoospores into the soil. Zoospores, which are capable of detecting external effector molecules secreted by potato plants, move towards the host organism through the pores in the soil. [31] Following detection of

host-derived signals, zoospores form a cyst outside the plant cell wall and penetrate through the lipid bilayer inside the cytosol. Penetration of the plasma membrane results in introduction of a fungal thallus into the host cytoplasm, while the cyst wall remains in the extracellular space. [31, 42] Inside the cell, the fungal body undergoes repeated mitosis to generate more spores. Infection-induced hypertrophic growth of host cells results in cellular lysis and release of replicated zoospores. [31] Newly produced zoospores can then undergo asexual replication cycle or go through a diploid phase. As a part of asexual cycle, mitotic reproduction of haploid zoospores continues through re-infection of host tissue. [31] Alternatively, the produced spores can fuse together, forming a diploid zoospore with two flagella. [31, 43] These spores infect the susceptible potato plants to generate resting spores, completing the life cycle. [43] Once favourable conditions occur, the resting spores undergo germination again and the cycle repeats.

Due to the tight relationship between *Synchitrium endobioticum* and *Solanum tuberosum*, the prolonged co-evolution of this fungal pathogen with its potato host resulted in the formation of multiple pathotypes of *S. endobioticum*. Over 40 different pathotypes of S. *endobioticum* are recognized currently. [31] Based on the previously conducted genomic studies, data suggest that these pathotypes could establish different pathogen-host interactions. The variation in the mechanism of host invasion of unrelated fungal strains resulted in the lack of available eradication techniques and, therefore, current regulatory control systems are focused on limiting the spread of *S. endobioticum.* [31] This includes extensive monitoring and sampling of infected material to prevent new infestations. The cultivation of susceptible varieties is usually restricted within the fields where infected material has been previously detected; however, different countries may follow different guidelines for limiting the spread of potato wart. [31]

1.3 Analytical techniques

1.3.1 Liquid Chromatography

Liquid chromatography is an analytical technique that separates analytes present in the mixture based on their physical properties, such as hydrophobicity, molecular size, binding efficiency, or charge. [44, 45] The analyte separation in liquid chromatography is achieved based on the interaction efficiency of sample components with stationary and mobile phases. The compounds that have

higher affinity to solvent travel with the solvent front and elute from the column faster. In contrast, the analytes that exhibit stronger interaction with the stationary phase are collected in the later elution fractions.

The most prevalent technique of liquid chromatography is called high-performance liquid chromatography, or HPLC. [44] The main difference between the classical liquid chromatography and HPLC is the use of additional pressure. In LC, the separation is often gravity-induced which results in the slow flow rate. In comparison, the HPLC is equipped with a pump that pushes the solvent through the column at high pressure. The application of a high-pressure pump in HPLC instruments facilitates faster analyte separation and increases analysis efficiency.

In its simplest form, the HPLC analysis starts with injection of sample solution into the column by the autosampler. The HPLC system can also be equipped with a guard column, which is a miniature version of the analytical column. [46] It is between 1 to 5 centimeters in length and its main goal is to prevent clogging of the column caused by the accumulation of material that gets irreversibly adsorbed to the stationary phase. [46] The solvent is continuously pumped through the system in either isocratic or gradient elution mode. During the isocratic setting, the solvent composition is consistent over the run. In comparison, the ratio of mobile phase solvents is changed during the gradient separation. Gradient elution is used to separate a mixture of analytes that have varying polarities and, therefore, retention times, providing a better resolution. As analyte compounds separate and elute from the column, their signal intensity and retention times are measured by the detector which generates the chromatograph. [44] Common detector types include fixed-wavelength ultraviolet (UV) detector, fluorescence detector and electrical detector. [45, 46] The HPLC system is also commonly used in tandem with mass spectrometry, which is referred to as LC-MS.

The choice of column for HPLC analysis depends on the desired mode of separation; however, the two main components of the column packing are the rigid support and the stationary phase attached to it. [44] Due to high mechanical strength, broad solvent and ligand compatibility, and high stability at various pH and pressure levels, silica-based particles are the most commonly used HPLC supports. [44, 45, 47] Other alternative option for column packing material are graphitic carbon, alumina, zirconia, titania, and porous-polymeric particles. [44, 45] The stationary phase is organic in nature and is covalently bound to the support. It is attached to the support through the covalent

7

interaction between the silanol groups that are located on the silica surface and organosilane supplemented with a functional group. [44-46] The functional groups used in stationary phase are categorized into alkyl, phenyl, fluoro, cyano, and aryl ligands. In reverse-phase liquid chromatography (RPLC), which is a predominant HPLC mode, the stationary phase consists of nonpolar saturated alkyl carbon chain attached to porous silica particles. [44] The alkyl substituent can contain either three, four, five, eight, eighteen, or thirty carbons. [44, 45] Both C8 and C18 are commonly used for separation of compounds with relatively low molecular weight, while C4 phase is used to separate analytes of larger size. [45] The mobile phase in RPLC is represented by a mixture of polar organic solvent, such as acetonitrile, and water. [48] The sample compounds are separated based on their individual polarities. The polar compounds have a higher affinity to a polar water-containing mobile phase and are, therefore, eluted from the column faster. [44, 45] The nonpolar solutes strongly interact with nonpolar stationary phase and are retained by the column longer. [44, 45] However, as the ratio of organic solvent to water increases, the retention of the analytes decreases.

Due to the high compatibility of reverse-phase liquid chromatography with mass spectrometry, RPLC is a commonly used mode for peptide separation. [48] The mobile phase for proteomic analysis is supplemented with low concentrations of acid (such as acetic, formic, or trifluoroacetic) to facilitate the positive charging on peptide sequences, as well as to prevent ionic interactions between the protein fragments and the stationary phase. [46, 48] The HPLC analysis acts as desalting step and results in enhancement of MS signal, reduced ion suppression, and increased resolution. [48]

1.3.2   Mass Spectrometry

Mass Spectrometry is an analytical technique that allows for the analysis of compounds present within a sample and measurement of their abundances. The mass spectrometric analysis consists of fragmentation of gas-phase ions, followed by their separation based on the mass-to-charge ratio (m/z). [49] The mass of each ion is measured in Daltons (Da), which is a unified atomic mass unit (amu). Numerically, one amu is equal to one twelfth of the mass of ^{12}C atom which is equivalent to 1.66×10^{-27} kg. [50] In comparison, the total charge of the analyte ions is represented by the

multiplication of the number of ion charges (z) and the absolute value of elementary charge (1 e = 1.602×10^{-19} C). [49]

In order for analytes to be detected by the mass spectrometer, they first have to be ionized. While protonation and cationization are the most common ionization types, neutral molecules can also be ionized by addition or removal of electrons, protons, or charged groups. Depending on the ionization mechanism, different adducts of one analyte can be detected within a spectrum.

Mass spectrometers are often coupled with chromatographic separation techniques to obtain higher selectivity. The high-resolution proteomics data herein was generated using Q-Exactive Orbitrap combined with HPLC system. The main components and their functions of this type of mass spectrometer are described below.

1.3.3   Electrospray ionization

Electrospray ionization (ESI) is an ionization technique that was first developed by Fenn and Yamashita in 1984. [51] ESI occurs at atmospheric pressure and induces minimal to no in-source fragmentation, which made it a popular method for analysis of large molecules in biological solutions. [51-53] One of the main advantages of ESI over other ionization techniques is the ability to easily couple it with liquid chromatography.[54] Additionally, low fragmentation rates that occur during ESI allow for the detection of high-molecular weight compounds, such as proteins, nucleic acids, polymers, and biopolymers because multiple charging of these compounds results in a relatively low *m/z* ratio of their ion fragments. [49, 54] However, this also makes the ESI spectrum difficult to interpret, which is one of the main disadvantages of this technique.

Depending on the nature of the analyte, ESI can be operated in either positive or negative ionization modes. In both cases, the analyte is usually suspended in a polar solvent which is introduced to the mass spectrometry system via a capillary tube at 1-100 µL/min. [49, 54, 55] To induce the formation of charged droplets at the ES capillary tip, an electric voltage of around 3 kV is applied to the spray capillary, which is located approximately 3 centimeters away from the plate of the mass spectrometer. [49, 54] The applied voltage causes polarization of the solvent, resulting in the charge accumulation near the surface of the meniscus and the accumulation of negative ions away from it. This causes distortion of the meniscus and results in the formation of a Taylor cone.

[49, 54] When the tip of the solvent cone becomes unstable, releasing fine jet from the cone which then breaks into small-charged progeny droplets. This process is Rayleigh limit dependent. Briefly, Rayleigh limit refers to the instability state when the electrostatic forces within the droplet (which is caused by the repulsion of charged ions inside the droplet) is counterbalanced by the surface tension. [56] These droplets undergo continuous shrinkage due to solvent evaporation, either induced by application of heat to the capillary or via interaction with pre-heated inert gas. [49, 54] When the charge repulsion exceeds the droplet surface tension, the droplet undergoes Coulomb explosion, forming more progeny droplets. The first-generation droplets are usually around several (i.e., 1-2) micrometers in size and can carry up to 50,000 elementary charges, while the offspring droplets have a radius of a few nanometers. [49, 55] This process continues until desorption occurs, producing gas-phase analyte ions.

The main difference between the positive and negative ion modes is determined by the source of electrons. [49] In the positive mode, the electrons have to be provided through oxidation of analytes in the capillary tip. For the negatively charged ions, electrons have to be acquired through the process of reduction. Both oxidation and reduction occur in the metallic capillary tube, making ESI an electrochemical process. [49]

Regardless of the ionization mode, the analyte desorption (or ion generation) occurs through one of the three main ESI mechanisms (Figure 2), which includes charged residue model (CRM), ion evaporation model (IEM), and chain ejection model (CEM). [55, 57] In the charged residue model, the analyte acquires the charge from the solvent as it evaporates to dryness. [55, 58] This mechanism is believed to apply to large molecules, such as proteins, that are spherical in shape and might be charged or neutral in nature. The IEM model is characterized by the ejection of solvated analyte ions from the surface of the droplet. [55, 59] This is caused by the high electric field within the Rayleigh-charged nanodroplet. [55] As a result, the analyte ion has some residual solvent molecules attached to it. This model is believed to be applicable for small and pre-charged species. In addition to these two well-established mechanisms, the CEM has been proposed as an alternative mechanism for release of protein ions from the Rayleigh-charged nanodroplet. More specifically, as the proteins unfold in solution, protein structure changes from compact to extended conformation and the unfolded chain migrates towards the outer surface of the droplet. [55] To minimize the repulsion forces, the chain terminus eventually gets ejected from the droplet. As

protein separates from the droplet, the peptide chain gains the charge from the solvent, resulting in the formation of a highly charged analyte ion. [55]
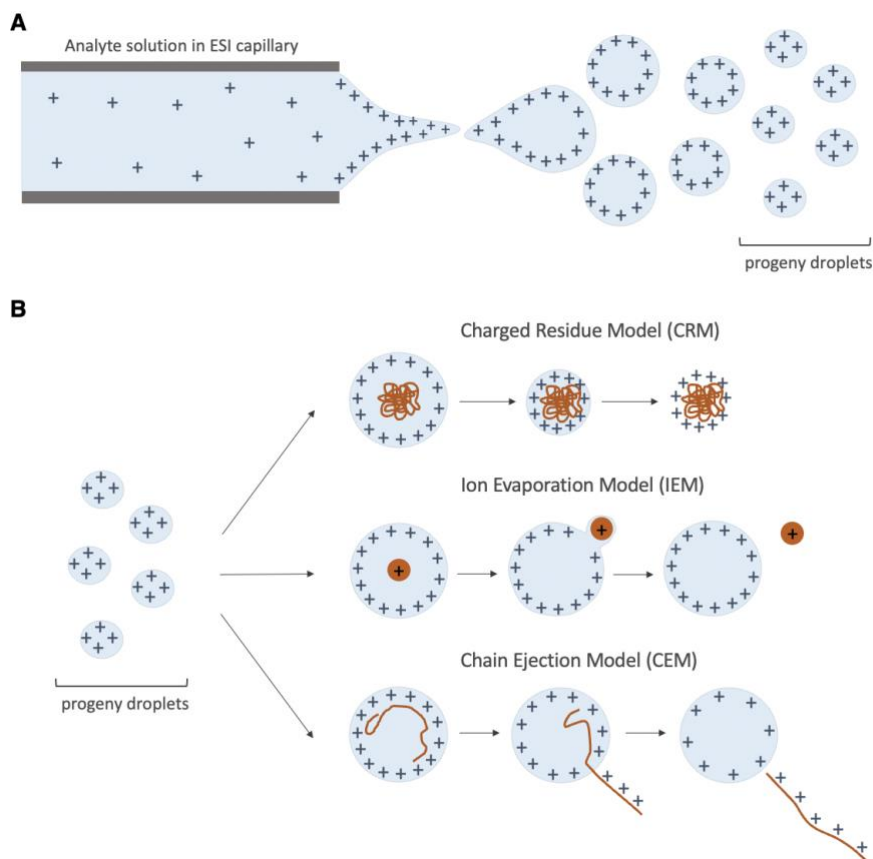
Figure 2. Schematic diagram of three ESI mechanisms. Figure 2A shows the process of solvent evaporation and formation of progeny droplets induced by ESI. Figure 2B shows the three mechanisms of ion release from the droplet. Depending on the nature of analyte, ions can be released from the sample droplet through Charged Residue Model, Ion Evaporation Model, or Chain Ejection Model.

For conventional proteomics analysis, nano-electrospray ionization is utilized. While nano-ESI follows a similar mechanism employed in a standard ESI, it operates at decreased flow rate of approximately 20-50 nL/min and requires small sample volume for analysis. [60, 61] The droplets produced following nano-ESI are smaller in magnitude in comparison to the solution droplets generated in ESI mode and do not exceed 150 nm in diameter. [60, 61] The small size of produced droplets also results in faster evaporation rates of the solvent, which allows to minimize the distance between the spraying capillary and the orifice of the mass spectrometer. [61] This increases

the signal sensitivity and minimizes the sample loss. [60] Additionally, each nano-ESI droplet contains on average one molecule, which prevents spectral clustering. [61]

1.3.4   Quadrupole mass filter

A quadrupole is a type of mass analyzer that is used to filter analyte ions based on their mass-to-charge ratios. [62] A quadrupole consists of four electrodes that have either hyperbolic or cylindrical cross section. [62, 63] The mass-resolving properties of the quadrupole are caused by the application of time-dependent alternating current (AC) and time-independent direct current (DC) to the electrodes. [62] The electrodes located opposite to one another receive the same voltage and, therefore, one pair of electrodes is negatively charged while the other pair is given positive AC and DC voltages. To promote the travel of ions through the quadrupole, the voltage gets successively altered between two pairs of electrodes which guides the beam of ions in x, y, and z axis.

For given voltage settings, only the ions of a specific *m/z* will pass through a quadrupole filter and will be identified by the detector. These ions will have a stable trajectory and undergo finite displacement along x or y axis (Figure 3). In contrast, analyte ions that have unstable trajectories collide with one of the electrodes and get filtered out from the ion beam. The quadrupole mass filter can also be operated in radio frequency (RF)-only mode which can be activated by removal of DC voltage. [62] The RF-only quadrupoles transmit almost all analyte ions in a high-range *m/z* ratios and are operated as ion guides.
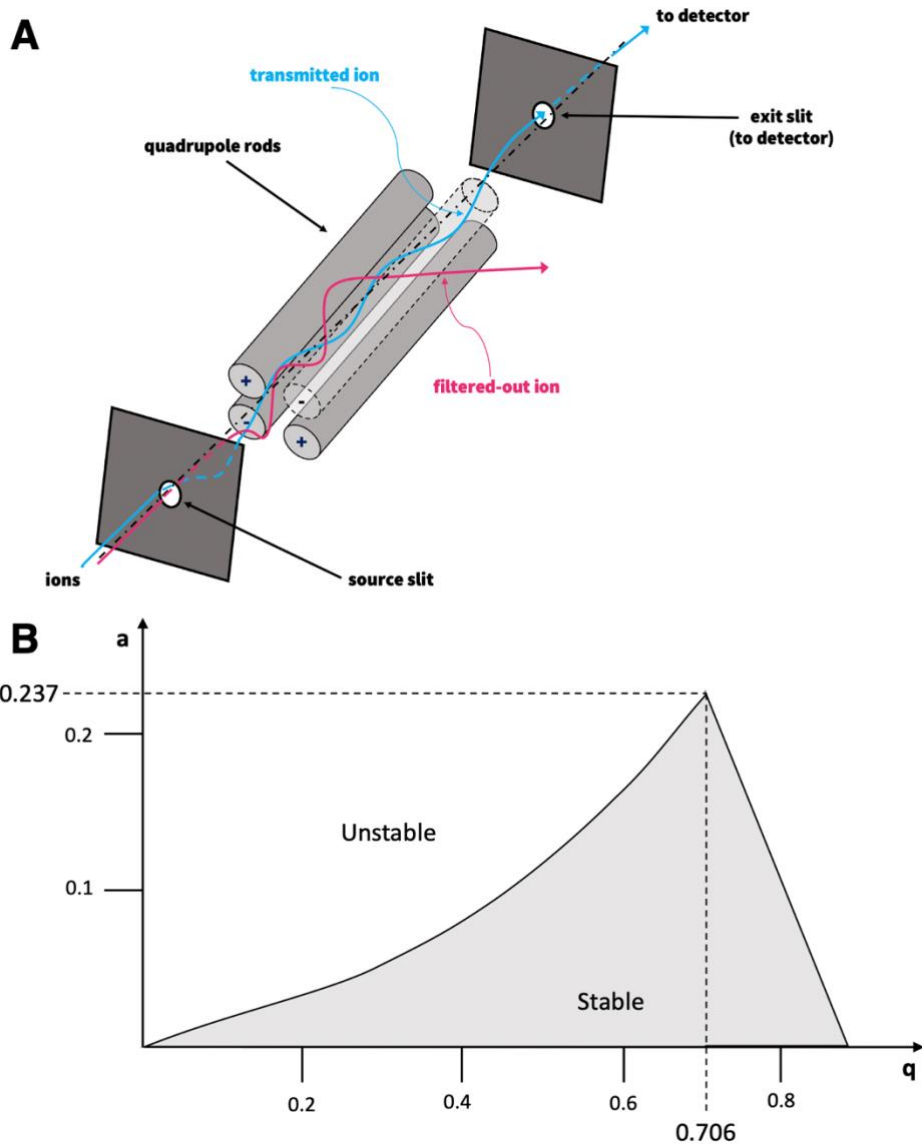
Figure 3. Schematic diagram of ion transmission in quadrupole mass filter. A) The difference between ion with stable trajectory (transmitted ion) and ion with unstable trajectory (filtered-out ion). B) The stability diagram of quadrupole mass filter. Depending on the voltage setting, the ions of different *m/z* ratios will be scanned by the detector. Analyte ions that end up being detected will fall into stable region, which has a tip at a = 0.237 (proportional to DC voltage) and q = 0.706 (proportional to DC voltage).

1.3.5 Orbitrap mass analyzer

The Orbitrap is a type of ion-trapping mass analyzer that was invented by Alexander Makarov. [64] It consists of the central spindle electrode and the outer barrel-like electrode and uses the electrostatic fields to induce stable ion trajectories. [64] After ions get injected into the Orbitrap, they start oscillating simultaneously around the inner electrode and along the z-axis of the analyzer (Figure 4). While the ions entering the Orbitrap have the same amplitude, the ions of different *m/z* values will adapt different rotational and radial motion frequencies. [64] More specifically, different ions will rotate around the central electrode with different magnitude, which will result in their radial dephasing. Analyte ions of the same *m/z* will remain in-phase and will continue to oscillate along the z-axis together. The frequency of coherent axial oscillations is measured through the acquisition of current transients which are then converted to *m/z* ratios using fast Fourier transform (FFT) algorithms. The correlation between *m/z* ratio of the ion and the frequency of its oscillation can be represented by the following formula:

$$\omega = \sqrt{(z/m) \times k}$$

where ω is a specific frequency, z is charge, m is mass, and k is an instrumental constant.
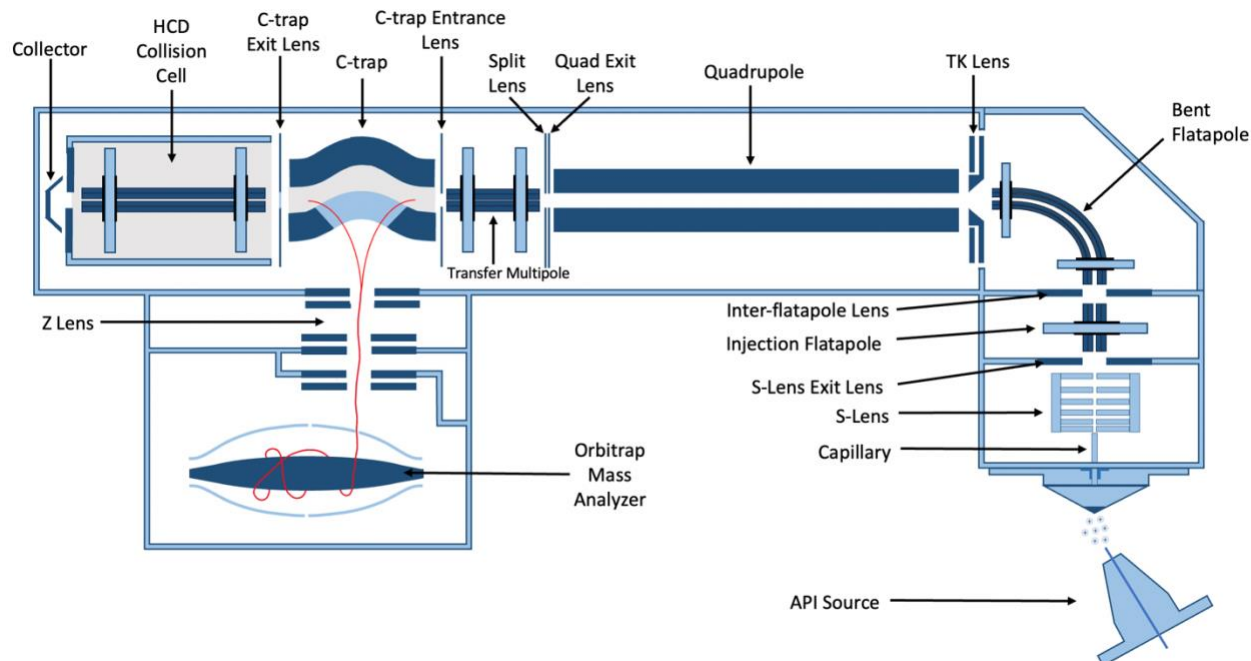
Figure 4. Schematic diagram of ion movement in Q-Exactive Orbitrap mass spectrometer. The ions generated by the ESI source get captured by the RF lens where they get focused in a tight beam. [65] The ions are then transferred to the bent flatapole ion guide called Advanced Active Beam Guide

(AABG) to reduce signal noise, followed by quadrupole mass filter. [65] The quadrupole transmits ions of a specific mass-to-charge ratio based on the precursor settings. [62, 65] The selected ions are then transferred through the RF-only octupole and enter C-trap, which acts as an ion trap. [65] Here, the ions are slowed down by the nitrogen gas molecules and are accumulated into packets. [64] Ions are then injected into Orbitrap mass analyzer to generate mass spectra. [64, 65] To acquire MS/MS data, the ions are pushed into the Higher Energy Collision Dissociation (HCD) cell, where the ions collide with nitrogen and undergo fragmentation. [65] Following fragmentation, the ions are moved back to the C-trap and are injected into the Orbitrap analyzer.

1.3.6 C-trap and Higher-Energy Collision Dissociation Cell

The C-trap is a C-shaped RF-only quadrupole (flatapole) that accumulates and stores ions prior to their injection into Orbitrap mass analyzer. [66-69] The C-trap is filled with nitrogen gas atoms which promote collisional cooling and capturing of ions. [66-69]

After the ion concentration reaches the pre-set amount, the pulsed high voltage is applied, and the stored ions are orthogonally injected into the analyzer in packets. [66-69] The injected ions undergo steady electrical acceleration which prevents movement of ions back to the trap once they enter the trapping field. [67] This allows simultaneous injection of ions to the analyzer, providing better signal. The C-trap is capable of storing ions that underwent fragmentation at different conditions which allows for parallelization of multiple measurements. [67] Ability to parallelize the analysis allows for MS/MS scans on one packet of ions, while the other set of fragmented ions is trapped in either the C-trap or the higher collision energy dissociation cell (HCD). [68] For MS/MS analysis, the ions can be directed into HCD, which is an RF-only quadrupole type cell. [67, 68] Here, the DC offset voltage is applied to the electrodes of the quadrupole to accelerate the precursor ions. The fragmentation occurs when ions collide with neutral gas molecules, which is usually either helium or nitrogen. [67, 68, 70] The high collision energies used in the HCD cell of the Orbitrap mass spectrometer allows multiple ion fragmentations to occur. [70] The resulting fragment ions are then trapped and cooled inside HCD, followed by their injection back to the C-trap and subsequent analysis by the Orbitrap detector.

1.4   Proteomics

Proteomics is a large-scale study of the proteome in order to identify protein structure and analyze protein expression and interactions. [71, 72] The proteome refers to all proteins that are expressed by a certain organism. [72-74] Proteomic analysis provides complementary information to genomic data and allows for identification of numerous protein modifications. [71] This includes analysis of post-translational modifications (PTMs), proteoforms (or different conformations of a singular protein), and abundances of each protein in a sample. [71, 72] Due to the complex and dynamic nature of proteins, the extensive characterization of the proteome using mass spectrometry-based proteomics can enhance our understanding of numerous biological processes. [73]

Following introduction of Electrospray Ionization (ESI), tandem mass spectrometry (MS) has become a leading method for qualification and quantification of proteomics samples. In mass spectrometry-based proteomic analysis, the mass-to-charge (m/z) ratio of the intact peptide sequence (or full proteoform) is analyzed in the initial MS scan. [73] In the MS2 scan, the analyte ions get fragmented, resulting in generation of the product ions. [73] The m/z values of the product ions are then used to determine the amino acid sequence of each peptide or the full-sized protein present in the sample and are used for the protein identification. [73]

Modern proteomics is categorized into top-down and bottom-up proteomics. [75] In the top-down proteomics, the mass spectrometry analysis is performed on the intact protein. [73, 75, 76] The amino acid sequence obtained from top-down analysis correlates with the specific biologically active proteoform of the protein of interest, which preserves any post-translational modifications. [73] Additionally, top-down analysis provides information regarding the function of the protein in vivo. [75] The main limitations of top-down proteomics are associated with the low abundance and low signal-to-noise ratio of product ions, as well as poor sequence coverage. [73] In addition to that, the solubility of vast majority of proteins in the appropriate for mass spectrometry analysis solvent may be difficult to achieve. [73]

Bottom-up proteomics (also known as shotgun proteomics) involves proteolytic digestion of protein mixture into small peptides that vary from 6 to 50 amino acids in size. [72, 73, 75, 76] The main steps of bottom-up proteomic experiments include extraction of proteins from the sample, concentration measurement of isolated protein mixture, enzymatic digestion, mass spectrometry

analysis, protein annotation and identification. [71] The ionization, solubility, and fragmentation of peptide sequences produced using bottom-up approach is easier to achieve over that of full-sized proteins. [71, 72] The other advantages of shotgun proteomics include improved sample homogeneity and separation efficiency. [75] One of the main limitations of the bottom-up approach is associated with differentiation of analyzed proteins or peptides between different organisms, which presents a challenge when analyzing heterogeneous samples (e.g., infected plant material). [73] The amino acid sequences of some protein groups are highly homologous, resulting in the large number of shared peptides. [73] This makes the peptide annotation process complicated.

Peptide fragmentation in mass spectrometry-based proteomics occurs due to intramolecular transfer of proton and thus the charge transfer to one of the cleavage sites. [76] The six main ion types produced during peptide ionization are called x, y, z and a, b, c ions (Figure 5) when the charge is retained at the C-terminal and N-terminal, respectively. [77] The cleavage at the amide bond results in generation of b and y ions. [76] These types of ions are produced following Collision Induced Dissociation (CID). [76] The loss of a CO group from the b ions results in generation of a ions that form an ion pair with x ions on the C-terminal end. [76] The a and x ions represent the breakage of the peptide bond between the alpha and carbonyl carbons. [78] In comparison, if the peptides are ionized using Electron Capture Dissociation (ECD), c and z ion types are formed. [76] These ions indicate the cleavage of the peptide at the amide nitrogen-alpha carbon site. [78] To indicate which peptide bond is cleaved, the ions are numbered. [77] Moreover, the number of hydrogen atoms that are either lost or transferred to the peptide fragment is indicated by the apostrophes. [77] If the hydrogens were transferred to the N-terminal, the apostrophes are placed to the right of the letter representing the ion type. [77] If the hydrogen atoms were left from the C-terminal, the apostrophes are put to the left of the letter. [77]

17
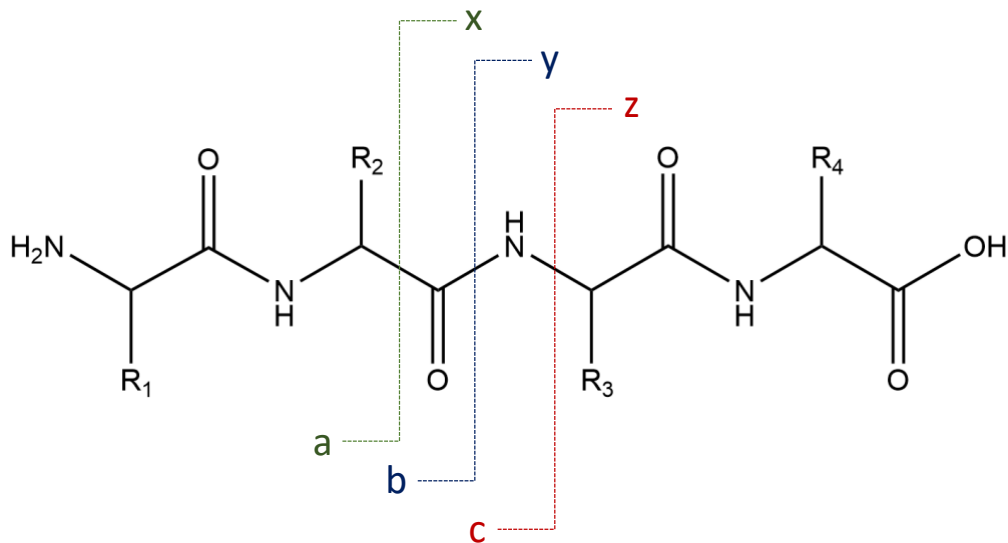
Figure 5. Peptide fragmentation during mass spectrometry analysis. During ionization, peptide sequences get fragmented in one of three sites, producing a and x ions, b and y ions, or c and z ions, respectively.

The workflow of sample preparation for proteomic analysis consists of lysate preparation, protein digestion, and peptide cleanup. Lysate preparation involves cellular lysis to extract proteins from the extracellular space and can be achieved using both mechanical and reagent-based techniques. [79] Lysis is optimized to the cell type of the sample and should be done in the presence of protease inhibitors to prevent protein degradation. [79] To separate proteins from cellular debris and other small molecules present in the sample, numerous purification methods are available. This includes liquid chromatography, dialysis, gel electrophoresis, and protein precipitation with either acetone or a methanol/chloroform mixture. [71] Protein digestion involves denaturation, reduction, alkylation, and proteolytic digestion steps. Denaturation can be accomplished by running sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) or by addition of strong chaotropic agents to the sample solution. [79] The disulfide peptide bridges are then reduced by addition of dithiothreitol (DTT) or Tris(2-carboxyethyl)phosphine (TCEP). [79] To prevent reformation of disulfide bonds, the sulfhydryl groups are alkylated by iodoacetamide (IAA) or iodoacetic acid. [79] Trypsin is the foremost protease used for protein digestion in proteomic analysis due to its affordability and robustness. [72, 73, 80-82] It typically generates small peptide fragments by cleaving the protein after arginine and lysine residues. [72, 73] Chymotrypsin, Lys-C, Lys-N, Glu-C, Asp-N, and Arg-C can also be used as alternative options. [73, 80] Chymotrypsin cleaves at

phenylalanine and tryptophan residues, and Lys-C and Lys-N result in digestion at lysine residue at C-terminal and N-terminal positions, respectively. [73, 80] Glu-C facilitates cleavage at aspartic and glutamic acids, Asp-N cleaves before the residues of aspartic acid, and Arg-C cuts following arginine. [73] The choice of protease depends on the goal of the analysis, as well as digestion time, buffers used for sample processing, and the nature of the protein analyte. Proteolytic digestion requires the unfolded (denatured) structure of the protein due to the easier enzyme access to the digestion sites. [71] To facilitate protein denaturation, the buffer used for cell or tissue lysis is supplemented with ionic detergents (e.g., sodium dodecyl sulfate (SDS) or sodium deoxycholate (SDC)), nonionic zwitterionic detergents (e.g., Triton X-100 or NP-40), or strong denaturing agents, such as urea. [71]

Following digestion and prior to mass spectrometry analysis, resulting peptides undergo cleanup step which can be accomplished using solid-phase extraction and subsequent liquid chromatography step.

1.5 Acquisition modes of mass spectrometric data

Data acquisition, including the selection of precursor ions, during mass spectrometry analysis can be obtained using either a data-dependent or a data-independent approach. In data-dependent acquisition (DDA), all precursor ions are scanned during the MS1 scan. [83-85] Following the generation of the survey scan, only a predetermined number of ions are further fragmented and analyzed to produce MS2 scans. [83-85] The MS/MS data collected provide a detailed information about the precursor ions. In comparison, data-independent acquisition (DIA) involves fragmentation of all precursor ions detected during a specific isolation window of a survey scan. [83, 85, 86] This allows acquisition of fragmentation data for all ions detected in MS scan instead of focusing on a predefined ion set. Depending on the nature of the sample, concurrent fragmentation of co-eluting peptides during DIA approach results in a highly multiplexed ion spectra. [86] However, despite the complexity of the spectra, data-independent acquisition collects unbiased data in a systematic fashion and has higher reproducibility and increased sensitivity than DDA. [83, 86] Additionally, no prior knowledge about mass-to-charge ratio of the precursor peptides is required.

The DDA method is usually offered as a default setting on most mass spectrometers and requires the definition of TopN peptides to be fragmented. [86] Retrospective targeting (using chromatogram extraction) for DDA is possible on MS1 level only, while it can be done for both MS1 and MS2 scans in DIA. [86] In comparison, the DIA requires the definition of the mass range to be analyzed, as well as the width of the precursor window (usually around 25 *m/z* in range) and the number of MS2 scans desired for each isolation window. [86]

The main steps of the sample processing workflow for both DDA and DIA analysis of proteomic samples share similarities and include protein extraction and digestion, fractionation, and data processing and acquisition. [83] However, the fragment ion spectra obtained using DIA analysis are usually complex and require deconvolution. This can be accomplished by interrogating obtained MS data with available spectral libraries. The spectral libraries include information regarding a mass-to-charge ratio of both precursor and fragment ions, as well as retention time observed, and the settings used to generate mass spectrometric data. However, such spectral libraries may not be available due to the lack of experimentally derived data and, therefore, low confidence in identified compounds. [83-85] The other limitations of DIA analysis over DDA include higher cost, time, and larger sample volume required to generate data and, therefore, this technique might not be suitable depending on the sample availability. In contrast, because selection of precursor ion fragments is automated, the data collected using DDA method lacks reproducibility as different precursor ions can be selected for further fragmentation for different samples, resulting in the inconsistency of acquired results. [83, 84]

In the study, I used the data-dependent acquisition approach to analyze acquired proteomics samples. The mass range for the MS1 scan was from 340 *m/z* to 1,800 *m/z* with resolution set to 70,000. The further fragmentation was performed on the 10 most abundant peptides to acquired MS2 scan. More specifically, the MS/MS analysis was conducted on the top ten peptides that had the highest peak intensities. The dynamic exclusion was set to 8 seconds to allow for detection of peaks of lower abundances. Peptide identification was performed using a combination of MaxQuant and Perseus software that aided in quantification and interpretation of large high-resolution proteomic data sets, respectively.

1.6 Recombinant protein expression

Recombinant protein expression using microbial systems is a biochemical technique that utilizes bacterial transformation process for production of recombinant proteins. [87] Bacterial transformation refers to a type of horizontal gene transfer mechanism that results in the acquisition and subsequent incorporation of exogeneous DNA into bacterial genome or plasmid. [88] Bacterial transformation is a parasexual process that naturally occurs in bacterial cells, and it was first discovered in a *Streptococcus pneumoniae* strain. [88, 89] While this type of homologous recombination allows acquisition of new genetic traits by the recipient cell, the protein expression occurs temporarily and requires specific environmental conditions. During transformation, one of the strands of external DNA is incorporated into the cell through a transmembrane channel. Inside the cell, single-stranded DNA (ssDNA) binds to DNA processing protein, which facilitates the attachment of RecA recombinase to the DNA strand. [88] Binding of RecA induces polymerization of ssDNA and its pairing to the bacterial chromosomal DNA. [88]

The main steps of high-throughput protein expression include cloning of the gene of interest into the vector, transformation of the vector into the host cells, induction, purification and characterization. [87] Briefly, the target gene is first cloned into bacterial promoter system and is transformed into bacterial cells. [90] The starter cultures are grown overnight and are scaled up into a larger culture. Depending on the vector used, the cultures that reached the mid-log growth phase can be induced with β-thio-galactosidase (IPTG) to promote protein expression, followed by overnight incubation with vigorous shaking. [90] Finally, the cells are pelleted and resuspended in lysis buffer.

To induce cell competence and promote transformation, chemical transformation or electroporation can be used. [89] The chemically induced transformation can be achieved using various techniques. This includes treating bacteria cells with divalent metal ions (i.e., Ca^{2+}, Ba^{2+}, or Mg^{2+} ions) followed by a brief heat shock incubation or freeze-thaw cycle. [89] The membrane permeability of the cells can also be improved by addition of tris(hydroxymethyl)aminomethane (Tris) buffer supplemented with polyethylene glycol (PEG) to the cells. [89] In contrast, electroporation refers to exposure of bacterial membranes to high-voltage electric fields with the purpose of creating transient holes in the bacterial membrane. [89]

Escherichia coli is often used as a preferred prokaryotic system due to its short generation time, high cell culture density, and fast transformation time. [87, 90] One of the *E. coli* strains that can be used for recombinant protein expression is called BL21, which was first described in 1986 by Studier and Moffatt. [87] The inability to produce Lon protease in levels sufficient for degradation of foreign proteins is one of the important features that makes it a suitable choice as a host organism. [87] Additionally, BL21 cells lack OmpT protease, which is responsible for digestion of T7 RNA polymerase which is needed for expression of extracellular proteins. [87, 90] Despite multiple advantages of using *E. coli* as a host organism for recombinant expression, the rapidity of the cellular growth may result in the expression of misfolded proteins. [90] Furthermore, the expression of eukaryotic proteins in bacterial cells might be accompanied by the absence of necessary post-translational modifications. [90] This may result in an insoluble expression of proteins. One of the techniques that can be used to improve protein solubility is by addition of affinity tags.

Affinity tags are differentiated into peptide tags and fusion partners, respectively, and can be attached to either C-terminal or N-terminal domain; however, the fusion proteins are commonly designed with affinity tag on the C-terminus which is connected to the protein sequence on the N-terminus through a linker region that contains the recognition site for the protease. [87, 91-93] Some of the common affinity tags that can be used to construct fusion proteins include poly-histidine, maltose-binding protein (MBP), N-utilization substance protein A (NusA), ubiquitin, Protein A, thioredoxin (TRX), small ubiquitin related modifier (SUMO), and glutathione S-transferase (GST), or their combination. [87, 93, 94] The designed AvrSen1 sequence used in this study contained both poly-histidine and maltose-binding protein tags to improve its solubility. The 6×His tag has a small size and is uncharged at neutral pH conditions, which minimizes the effects of histidine residues on the structure and, therefore, the function of the tagged protein. [95] In comparison, maltose-binding protein, or MBP, allows the transformation of the protein of interest at *malE* gene in *E. coli.* [92, 93] Some of the advantages of fusing proteins with MBP tag include easy post-transformation purification protocol, high-yield, simple and inexpensive large-scale purification steps, and the lack of cysteine residues that might interfere with expression of the target protein. [92] Incorporation of these affinity tags in tandem results in improved solubility levels and enables the use of multiple methods for efficient purification. [93]

The use of affinity tags during protein expression and purification process has multiple advantages. While the presence of the affinity tags allows development of high-throughput sample processing method that would be suitable for diverse range of proteins, they also promote protein yield and result in increased protein stability. [91, 93] However, affinity tags can potentially affect protein activity or structure and have to be removed as part of the protein purification protocol. [91, 93] This can be achieved through the use of either chemical reagents or enzymes. [87, 91, 93] While the choice of proteolytic reagent depends on the goal of the analysis, the enzymatic digestion is usually preferred. In addition to mild reaction conditions, enzymes usually exhibit greater specificity and prevent irreversible damage to the protein of interest. [91] Some of the proteases that are commonly used for digestion of fusion proteins include factor Xa, enterokinase, thrombin, and tobacco etch virus (TEV) protease. [87, 91, 93] Due to the high specificity and activity over broad pH range and at various ionic strength conditions, TEV protease was selected for the cleavage of MBP tag from AvrSen1 protein. [96] TEV protease recognizes and cleaves between glutamine and serine at ENLYFQS sequence. [91, 93, 96] The serine in P1 position can be replaced with either glycine, alanine, methionine, cysteine, or histidine. [96] While the optimal activity of TEV protease is observed at 37°C, the digestion can also be optimized to low temperatures, which ensures the stability of the protein of interest.

1.7 Immobilized metal affinity chromatography

Immobilized metal affinity chromatography (IMAC) is a single-step protein purification technique that was implemented in 1970s. [97] It allows purification of recombinant proteins or peptides due to the high-affinity binding to chelated divalent metal ions. The metal ions are immobilized in resin and interact with proteins tagged with six consecutive histidine residues (Figure 6). This provides a fast method of purification of the protein of interest from the rest of the contaminants. [95, 97] Other proteins present in the sample solution have either weak or no affinity to the metal ions and end up in the flow-through fractions during the washing steps. [95, 98] Moreover, the addition of NaCl and imidazole to the loading and wash buffers minimizes the electrostatic attraction to metal beads and prevents low affinity binding to resin, respectively. [98] The elution of bound proteins is accomplished by either increasing imidazole concentrations, decreasing pH, or use of a strong chelating agent. [95, 98] While imidazole competes for binding to metal ions and displaces

recombinant proteins from the resin, decreasing pH level results in protonation of histidine residues, which abates binding efficiency between the protein and metal. [95, 98] In comparison, the use of a strong chelator, such as Ethylenediaminetetraacetic acid (EDTA), disrupts metal-resin interaction and causes the elution of metal ions and the proteins attached from the resin matrix. [98] The sample preparation for IMAC purification includes centrifugation of harvested bacterial cells, followed by resuspension of bacterial pellet with loading buffer and its enzymatic or mechanical lysis. [99] The mechanical lysis can be accomplished by sonication, homogenization, or repeated freeze/thaw cycles. Following cellular lysis, the sample is applied to the pre-equilibrated column. In addition to simple sample processing method, other advantages of IMAC over other conventional protein purification systems include scalability, high protein loading capacity, column regeneration, high ligand stability, and low costs. [97]

IMAC columns can be charged with either nickel, copper, zinc, or cobalt. [98] The chelators that are commonly used for metal immobilization include iminodiacetic acid (IDA) and nitrilotriacetic acid (NTA). [95, 98] In this study, Ni Sepharose 6 Fast Flow column, which consists of highly cross-linked beads of agarose pre-charged with Ni^{2+} ions, was used to purify AvrSen1 from the bacterial lysate. [99]

Sepharose beads charged with Ni^{2+} ions

Ni^{2+}

MBP + AvrSen1   6 × histidine tag
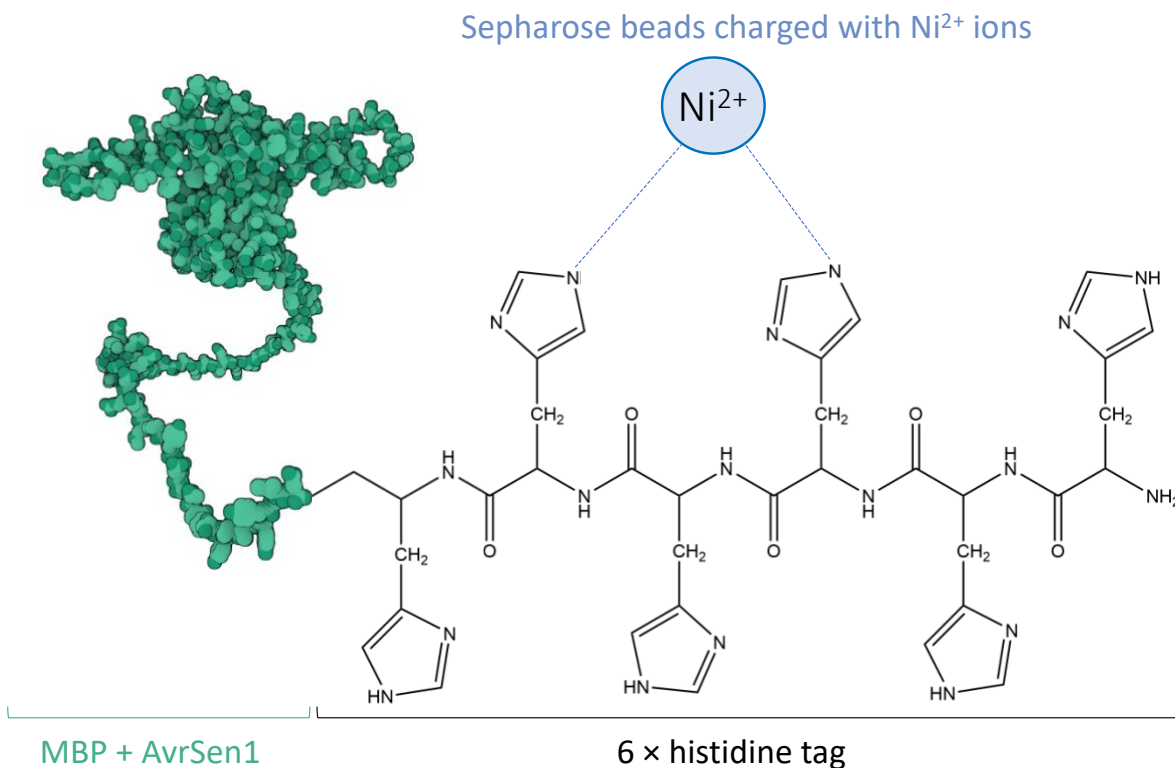
Figure 6. Schematic diagram of the binding between His-tagged protein to Ni^{2+}-charged Sepharose 6 Fast Flow column. The protein structure of AvrSen1 was obtained using AlphaFold prediction. [100, 101] AvrSen1 designed for this study contains both an MBP and a 6×His tag. The His-tag allows to purify the protein using Ni resin. The nickel is bound to the carbonyl groups of the matrix via -N and -OH moieties. Additionally, it binds to two histidine residues of the recombinant protein by forming coordinate bonds.

1.8 Size-exclusion chromatography

Size-exclusion chromatography (SEC) is a non-destructive analytical technique used to separate biomolecules, such as proteins, based on their hydrodynamic radius. [102, 103] In SEC, the separation is facilitated by the diffusion of molecules through a stationary phase and is based on the molecular size of biomolecules. [102, 103] While the stationary phase is composed of porous spherical particles, the mobile phase is usually represented by the aqueous solution. [102] The proteins of larger size are unable to enter and travel through the pores of the stationary phase and, therefore, elute from the column first (Figure 7A and 7B). [104] The elution of subsequent proteins occurs in the order of decreasing molecular weight (Figure 7C). [102-104] The separation time of SEC analysis is directly

correlated to the flow rate of the mobile phase used. [102] Therefore, reduction of run time can be accomplished by either increasing the flow rate or reducing the length of the SEC column to be used.

The SEC column can be packaged with either silica or cross-linked polymeric material. The silica columns most commonly consist of bare silica, 1,2-propanediol-bonded silica, or ethylene-bridge hybrid material (BEH). [102] In comparison, the polymer packings can contain cross-linked agarose or sulfonated polystyrene, polydivinylbenzene, or polyamide. [102] The choice of stationary phase mainly depends on the desired pore size, which is correlated with the size of the protein of interest; however, one of the parameters that should be considered when choosing the stationary phase is non-binding interactions that can occur between silica and proteins. This includes both electrostatic and hydrophobic interactions. [102] More specifically, the electrostatic interactions can occur due to similar or opposite charging on protein and silica. Identical charges can result in electrostatic repulsion that will prevent the protein from entering the pores of the stationary phase particles. [102, 105] In this case, the protein will elute faster than expected, causing elution time shifting. If the protein and silica are oppositely charged, protein adsorption occurs and the elution time increases. [102, 105] The electrostatic interactions can be prevented by increasing the ionic strength of a mobile phase buffer. [102, 103] In addition to non-binding interactions, ionic interactions can also occur during SEC analysis. The ionic binding may alter the protein structure and its three-dimentional conformation, or it can strengthen the affinity of proteins to the silica pores, resulting in elution time shift or peak tailing. Both hydrophobic and ionic interactions can be prevented by supplementation of mobile phase with SEC additives, such as arginine, or by optimizing the pH to ensure it does not exceed the isoelectric point (pI) of the analyte. [102, 103, 105]

Following protein elution from SEC column, the separated fractions are analyzed by the detector. The most common detection mode for SEC analysis is ultraviolet (UV) detector at 210, 214, 220, 270, 275, and 280 nm wavelengths. [102, 103] While the lower UV wavelengths allow detection of amide peptide bonds, aromatic amino acids get detected at higher wavelengths. In addition to UV detector, static and dynamic light scattering (LS), fluorescence, intrinsic velocity, viscometer, and refractive index (RI) detectors, as well as their combination, can also be used depending on the goal of the analysis. [102] In some cases, SEC can be interfaced with a mass spectrometer to determine the accurate molecular weight information of the analyte.
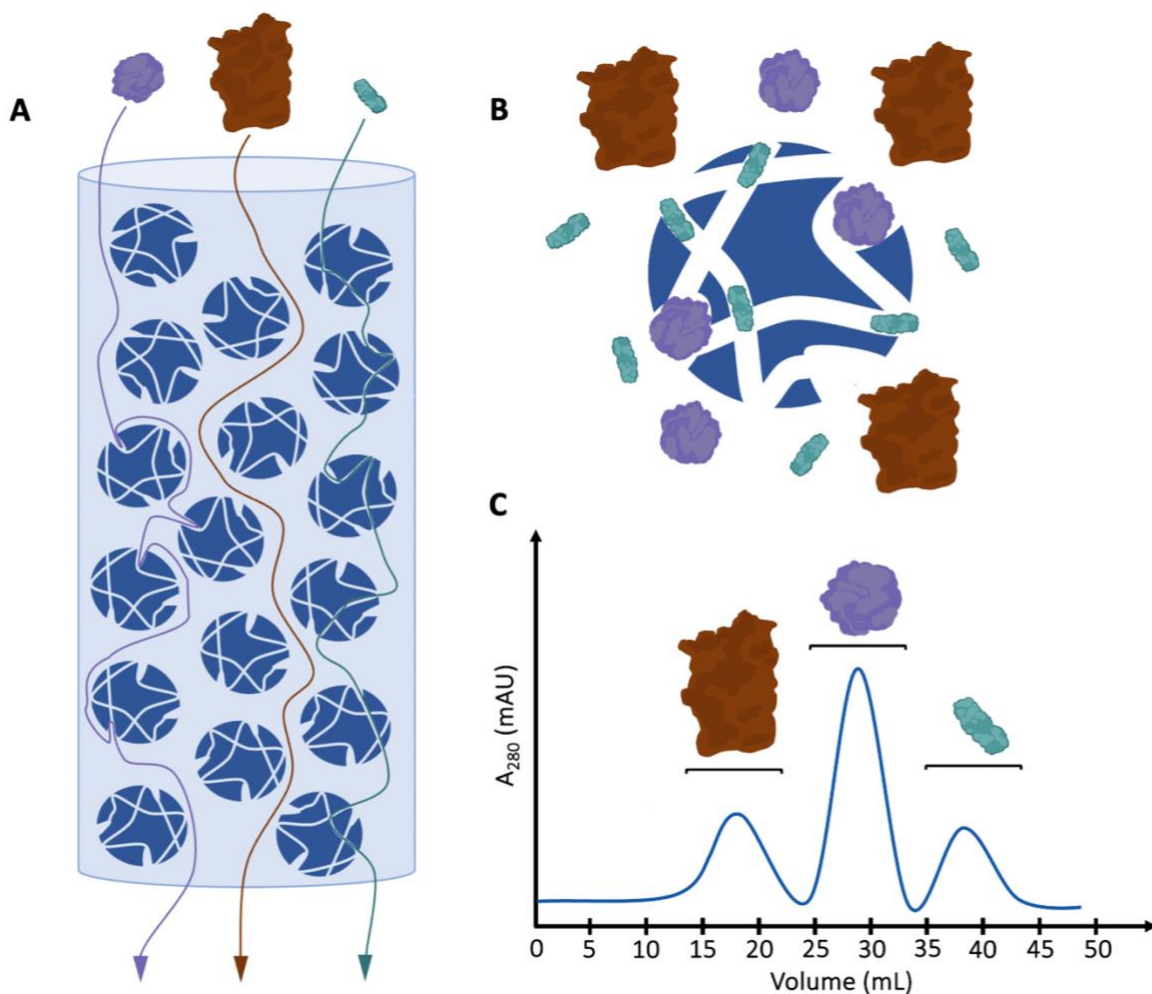
Figure 7. Separation of proteins of different molecular weight on SEC column. Figure 7A shows the difference between the paths that proteins of different size take through the resin during size-exclusion chromatography. Figure 7B shows that the smaller proteins enter the pores of the column and are retained by the stationary phase longer. By comparison, the proteins of large molecular weight are too big in size to enter the porous material and, therefore, elute from the column faster. Figure 7C represents an example of chromatogram obtained following size-exclusion analysis. As indicated before, the larger proteins elute faster and require smaller volume of the mobile phase. The smaller proteins will be represented by the later peaks on the chromatogram.

Even though the data obtained by size-exclusion chromatography are qualitative, the SEC chromatogram can be used to roughly approximate the molecular weight of the analyte. [102] This

can be accomplished by preparing a calibration curve based on proteins of known size. By plotting the logarithm of the molecular weights versus retention volumes obtained during calibration, the third order polynomial curve line can be generated. [102, 105] However, one of the limitations of using a calibration curve to determine the molecular size of the protein of interest is caused by the potential variations in the protein configuration. More specifically, changes in temperature can result in alteration of protein confirmation, viscosity of the mobile phase or the rate of analyte diffusion, which would alter the volume of mobile buffer needed for elution. [105]

In this study, the Bio-Rad NGC Chromatography System equipped with Superdex™ 200 SEC column was used for size-exclusion chromatography purification of AvrSen1. This system consists of the system pump, mixer, sample inject valve with sample pump and the sample loop, column and the column-switching valve, multi-wavelength detector, and the fraction collector. [106] The SEC buffers get pulled through the system by the system pump and are then homogenized in the mixer module. The sample is manually injected into the selected sample loop, where it then gets mixed with the sample buffer. [106] The sample buffer acts as a mobile phase and pushes the sample through the column. Following elution of analyte solution from the column, each of the fractions is analyzed by the conductivity and UV/Vis detectors and is then collected into a 96-well plate by the fraction collector. [106]

2 Chapter 2 - Comparative proteomics analysis of control and potato wart - infected Solanum tuberosum tuber samples

2.1 Chapter 2 Objectives

Potato wart is one of the major quarantine pathogens in the agricultural sector in both Canada and worldwide. Due to high pathogenicity of *S. endobioticum* and inability to grow this fungus in laboratory conditions, the mechanisms of host invasion and molecular host-pathogen interactions involved in disease progression remain underresearched. Studying proteomic and metabolomic changes associated with infection could provide additional information on disease distribution and resistance development and may aid identification of potential biomarkers that can be used for disease diagnosis. Current testing guidelines include harvesting potato tubers following their growth and visually accessing plant material, followed by running PCR or other genomic experiments to characterize fungal pathotype. PCR is a simple and relatively inexpensive process that allows identification of a specific fungal species causing an infection based on the presence of a certain DNA fragment in the genome. However, it does not verify whether the proteins produced by this pathotype are then successfully translated. In comparison, proteomics results in detection of expressed proteins, providing a more accurate representation of the pathways affected by infection. Therefore, development of techniques for plant material screening based on detection of certain upregulated or downregulated biological molecules may result in easier testing approach. In this study, a comparative proteomic analysis of four potato varieties with and without wart infection was conducted to study the immune response mechanisms of S. tuberosum induced following infection with *S. endobioticum*. The analysis was performed using liquid chromatography–tandem mass spectrometry (LC–MS/MS) approach through coupling of an EASY-nLC 1000 system with Q-Exactive Orbitrap mass spectrometer.

2.2 Introduction

Potato wart induced by *Synchitrium endobioticum* is a devastating infection of cultivated potato plants in many geographical areas worldwide. The pathogenicity and virulence profile of *S. endobioticum* pathotypes is assessed based on its ability to promote infection in potato varieties lacking certain resistance genes. Both genome and transcriptome studies provide information on mechanisms involved in plant-pathogen interactions; however, analysis of post-transcriptional processes in proteome and metabolome can advance the understanding of pathogenicity of the

organism and the immune response strategies employed by plants. [107] Conducting proteomic analysis on plants pre- and post-infection allows for quantification of proteins involved in the interaction between the pathogen and its host and to study post-translational modifications and biological pathways that may play roles in the progression of the disease. [107] Qualitative and quantitative proteomics provides insight regarding the physiological and structural changes that occur as a result of infection. [108] Analysis of proteins involved in phytopathogenic fungal disease can result in identification of effector molecules and pathogenic factors that are involved in fungal virulence. Additionally, comparative proteome studies analyze proteins involved in immune response, signal transduction, and metabolic pathways that play an important role in fungi-plant interaction. [108]

2.3   Methods

2.3.1   Plant material acquisition

Samples for this study were provided by the Canadian Food Inspection Agency (CFIA) research centre in Charlottetown, PEI. Overall, four potato cultivars of varying resistance response to *S. endobioticum* infection were obtained for proteomic analysis, which were analyzed in two different rounds. The first round included analysis of Russet Burbank and Targhee Russet varieties, while the second round of sampling consisted of analysis of Russet Burbank, Mountain Gem Russet, and Caribou Russet tubers. The choice of these varieties was based on the availability of sample material collected from the farm fields. For each variety, both non-infected, or control, and infected material were collected. Infected tubers were sampled twice to account for the difference between the warted tissue and visually non-warted tissue. More specifically, tissue fragments were collected from both the tuber end and the wart material, which were then referred to as infected non-warty and infected with visible warts, respectively.

2.3.2   Protein extraction from tuber material

To extract proteins from the tuber samples, a modified version of the protocol developed by Murawska et al. (2017)[109] was used. Briefly, the tuber fragments were cut out from the tuber using a 7-mm corkborer and were then collected in 2.0-mL Eppendorf tubes. To prevent biochemical changes, the tubes were immediately snap-frozen in liquid nitrogen, and the sample material was lyophilized over 48 hours. The freeze-dried samples were manually homogenized using a sterile

micropestle, followed by addition of 15 µL of 4 % sodium deoxycholate and 585 µL of 25 mM NH4HCO3 solution to the ground material. The samples were then sonicated in an ice-cold water bath for 15 minutes, vortexed for 15 seconds, and centrifuged for 10 minutes at $12,000 \times g$ and 4°C. Following centrifugation, 50 µL of supernatant was transferred to a fresh 2.0-mL Eppendorf tube, while the rest of the supernatant was saved and stored at 4°C. To precipitate proteins and separate them from the rest of the cellular debris present in the sample solution, the supernatant was mixed with 1 mL of cold acetone and incubated at -20°C for 1 hour. When the incubation step was complete, the samples were centrifuged for 10 minutes at $12,000 \times g$ and 4°C and the acetone was removed from the tubes, leaving the protein pellet at the bottom of the tube undisrupted. To ensure no more acetone was remaining in the sample, the protein pellets were air-dried and were then stored in -80°C until they were ready to be shipped to Ontario.

2.3.3 Trypsin digestion and SPE clean-up

The dried protein samples were resolubilized in 200 µL of 100 mM digestion buffer (TrisHCl, pH 8.5; 1% SDC). To measure the overall protein concentration, 20 µL of each sample was transferred to a 96-well plate, where it was mixed with the Pierce™ Rapid Gold BCA Protein Assay Kit reagent buffers (Thermo Fisher Scientific, Waltham, MA, United States). The volume corresponding to 25 µg of total protein content in the sample was then transferred into a new 1.5-mL Eppendorf tube. Each sample was mixed with digestion buffer to reach 200 µL overall volume and was incubated for 10 minutes at 60°C and 800 rpm. The proteins were then reduced and alkylated with 2.5 µL of 100 mM dithiothreitol (Roche Diagnostics, Indianapolis, IN, United States) and 3 µL of 200 mM iodoacetamide (MilliporeSigma, Burlington, MA, United States), respectively. The proteolytic digestion was accomplished by adding a working solution of trypsin (Thermo Fisher Scientific, Waltham, MA, United State) in a 1:50 ratio (trypsin:sample). The samples were then incubated at 37°C overnight with intermittent shaking.

Following overnight incubation, the protein digests were mixed with 20 µL of 10% trifluoroacetic acid (TFA) solution and 200 µL of ethyl acetate. The resulting peptide mixture was vortexed and centrifuged at $13,000 \times g$ for 3 minutes. The top organic layer was then discarded, and the bottom aqueous layer was transferred to another 1.5-mL Eppendorf tube, where it was diluted with 200 µL of Milli-Q water. The sample clean-up step involved addition of sample solutions to Oasis

31

HLB 1cc (30 mg) extraction cartridges. More specifically, the cartridges were first conditioned with 2 mL of methanol and equilibrated with 2 mL of 0.1% aqueous solution of formic acid (FA). To load samples into HLB sorbent, the cartridges were filled halfway with 0.1% FA and the samples were added on top of the formic acid solution. The mixture was then allowed to slowly drip through, followed by a washing step with 400 µL of 0.1% FA and elution with 600 µL of 70% acetonitrile (ACN). The final step of the clean-up procedure included drying the samples in the CentriVap Centrifugal Vacuum Concentrator (Labconco Corporation, Kansas City, MO, United States). Once the samples were dried down, the peptide pellets were reconstituted with 200 µL of 5% ACN, 0.1% FA solution and transferred to 250-µL snap cap polypropylene HPLC vials (Agilent Technologies, Santa Clara, CA, United States).

2.3.4   Proteomic analysis on nLC-MS/MS and MaxQuant data processing

All peptide samples were analyzed on an EASY-nLC 1000 system coupled to a Q-Exactive Orbitrap mass spectrometer. Full MS scan was acquired at 70,000 resolution in the mass range of 340-1,800 m/z, with automatic gain control (AGC) target and maximum injection time (IT) set to 1×10^6 and 256 ms, respectively. The high-resolution MS/MS analysis was performed on the top 10 peptides with 17,500 resolution, AGC of 1×10^6 and maximum IT of 110 ms.

The data processing was performed using MaxQuant software v.2.0.3.0. (Max Planck Institute of Biochemistry, Munich, Germany). To quantify identified proteins and generate a protein group library, each DDA-MS raw file was searched against a *S. tuberosum* reference proteome obtained from UniProt (UP000011115), which consisted of 53,106 amino acid sequences, as well as a *S. endobioticum* proteome (UP000320475) that was comprised of 8,518 proteins. Both proteomes were accessed on January 19th, 2024. Most settings were kept as default, with the exception of making sure the Label-free quantification and iBAQ analysis options were selected. Following MaxQuant database search, the protein IDs output file was imported into the Perseus software platform (v. 1.6.13.0.). Protein identifications containing potential contaminants, reverse sequences, and peptides identified by site peptides were filtered out. The peak area values of the remaining proteins were transformed to $\log_2(x)$ values. Lastly, two-sample T-test analysis was performed to determine which proteins were differentially expressed between control and infected plants. The protein identifications with a p-value higher than or equal to 0.05 were ignored.

Significant proteins (proteins with a p-value of less than 0.05) were then annotated using PANTHER classification system. In short, the protein ID numbers were batch uploaded for gene list analysis and *Solanum tuberosum* was selected as an organism in order to view the functional classification of detected proteins.

2.4 Results and Discussion

2.4.1 First round of sample acquisition – Russet Burbank and Targhee Russet

During the first round of sampling, a total of 20 samples were acquired. Eight of the samples belonged to Russet Burbank variety and included four control (RBC) and four infected warted (RBI) samples. The other 12 samples were Targhee Russet tubers with four control (TC), four infected non-warted tissue material (TINW) and four infected with warts (TIW) samples obtained, respectively.

Following MaxQuant analysis, the label-free quantification (LFQ) intensities of identified proteins were transformed to $\log_2(x)$ values and the missing values were normalized. To compare the protein expression levels between the control and infected groups, the two-sample T-test was conducted. The protein groups that had a p-value higher than 0.05 were removed. In Russet Burbank variety, a total of 623 proteins were detected, 102 of which were differently expressed in control and potato wart-infected samples and had a p-value of less than 0.05 (Figure 8). For Targhee Russet samples, all three groups were compared to one another. When comparing TC and TINW samples, out of 415 proteins that had non-zero LFQ values, 414 of them were insignificant ($p > 0.05$). This suggests that there is no significant change on the proteome level between the non-infected tubers and the infected tissue that was sampled at the non-warted end. These observations may indicate that potato wart has a localized nature and low virulence. This has not been described previously; however, considering that *Synchitrium endobioticum* is a biotrophic fungus and heavily relies on its host for nutrient acquisition, the slow disease progression is beneficial for fungal survival. In other words, potato wart seems to spread through tuber tissue slowly. This allows continuation of the normal metabolic processes of the host, which supplies enough nutrients for both fungal and plant growth. The comparison between Targhee Russet control and infected with warts material (TC versus TIW) resulted in detection of 464 proteins with 48 of them having values above zero in one or more samples in both sample group. The rest of the proteins identified in MaxQuant had quantifiable peak intensity values detected for only one sample group and were

excluded from statistical analysis. Out of 48 non-zero proteins, 16 of them did not differ significantly between sample groups, while the other 32 proteins were downregulated in infected plants. Finally, looking at the differences between the infected non-warty (TINW) and infected with warts (TIW), 552 protein groups were identified and 207 of them were significant ($p < 0.05$) (Figure 8). Since the comparison of TIW and TINW yielded the largest number of proteins being identified, these two groups were therefore used for subsequent annotation via gene ontology (GO) analysis.
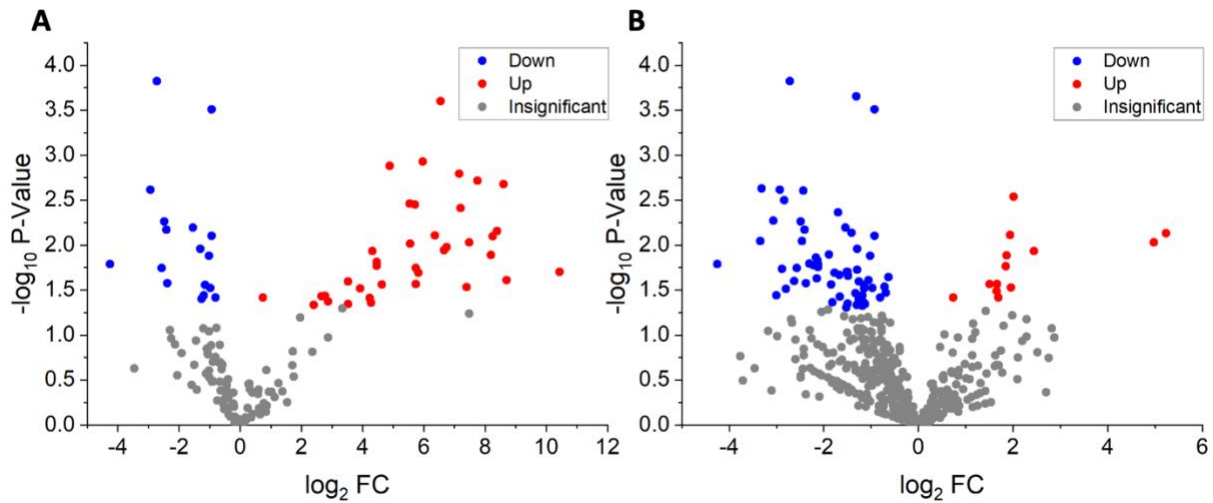
Figure 8. Volcano plot displaying the results of t-test comparisons of A) Targhee Russet infected non-warty (TINW) versus Targhee Russet infected with visible warts tuber samples and B) Russet Burbank control versus Russet Burbank infected tuber material.

By conducting GO analysis using Panther software, 102 and 207 significant proteins in Russet Burbank and Targhee Russet varieties, respectively, were annotated based on their functions. The most abundant protein classes identified in both varieties were metabolite interconversion enzymes, metabolism-related proteins, ribosomal proteins, and protein involved in plant defense (Figure 9). Heat shock (or chaperone) proteins ones involved in plant growth and development, as well as proteases and protease inhibitors were abundantly present in the samples.
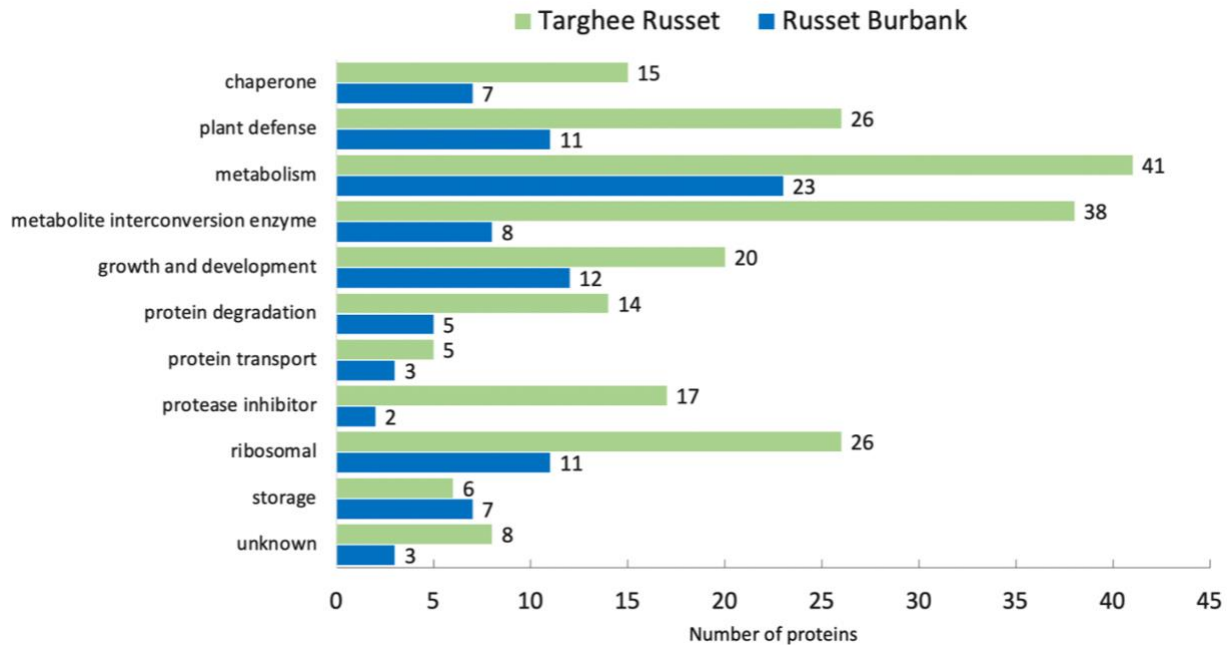
Figure 9. The distribution of significant proteins detected in Targhee Russet and Russet Burbank based on their functions. This bar chart shows the protein groups that had a p-value < 0.05 and were, therefore, identified as significant when comparing control and infected samples (RBC versus RBI and TINW versus TIW, respectively).

Looking at the fold change (FC) of the proteins identified in both varieties, it was observed that there are differences in the way these two varieties respond to infection. These data are visualized in the heat map below (Figure 10). The plot shows the change in protein expression levels post-infection when comparing control and infected tubers. To visualize the difference in immune response mechanisms between varieties, the Log2 fold change was calculated by comparing the average of detected peak areas of control and infected samples of each variety. By plotting the Log2 fold change values for Russet Burbank variety, it was determined that majority of the proteins are upregulated following infection. This trend is observed throughout protein groups of different functions, including metabolism-interconversion proteins. However, this does not apply to the proteins involved in plant defense. More specifically, the expression of some of the proteins that play role in plant defense increased, while other plant defense-related proteins were downregulated. Therefore, future work should concentrate on obtaining more conclusive results on the effects of *S. endobioticum* on the immune system of Russet Burbank potatoes in a larger

sample set. In contrast, Targhee Russet variety seems to respond differently to the infection. All protein groups that were used to plot a heat map were downregulated. Interestingly, three out of four TIW samples did not have any proteins. These results were verified through re-analysis of the same samples, as well as technical replicates of these samples. The lack of proteins in infected potatoes has not been described in literature before, and may suggest that this variety undergoes necrotic decomposition as a result of *S. endobioticum* – triggered disease.
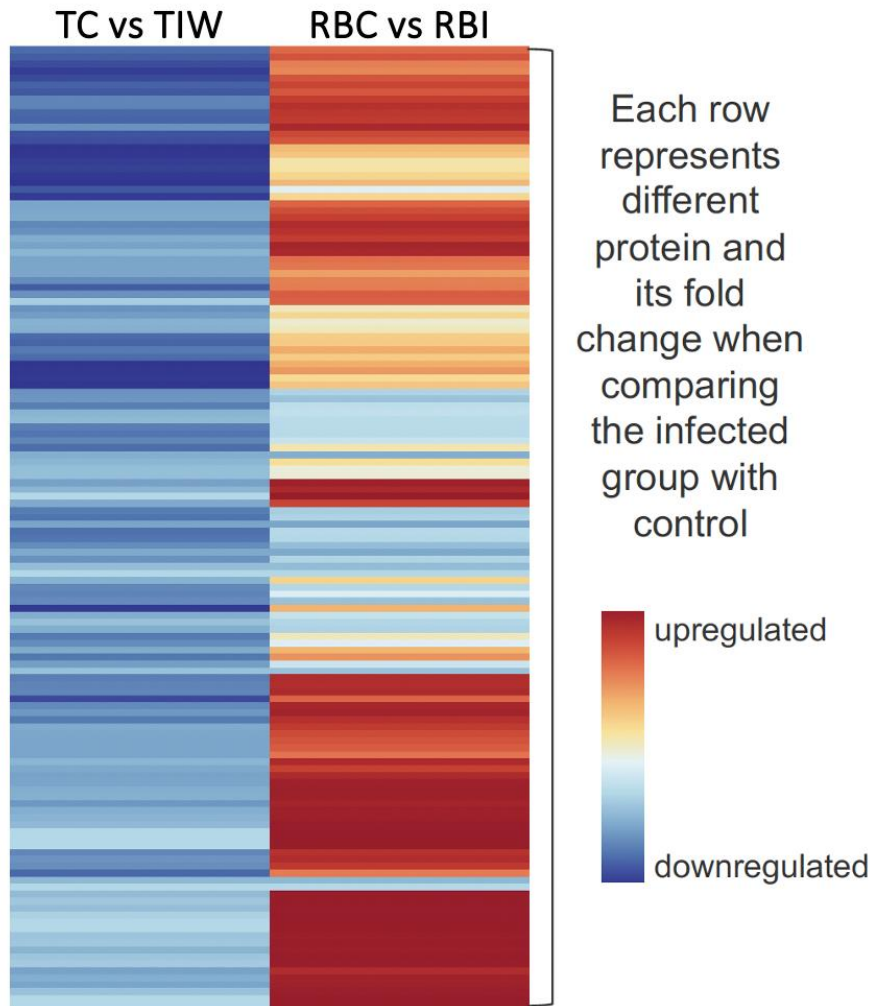
Figure 10. Heat map illustrating the change in expression levels of proteins detected in Targhee Russet and Russet Burbank potato varieties. The two columns of the map show the Log 2 fold change between control and infected samples of Targhee Russet and Russet Burbank varieties. Each row represents a different protein. If the fold change value is positive, that means the level of that protein is higher in infected material. Therefore, this protein is upregulated upon infection and the cell that correspond to that protein is red. Alternatively, if the level of the protein is lower in

infected tubers in comparison to control sample, that protein is downregulated and will be highlighted in the shade of blue on the map. Protein identities can be found in Table S3 in Appendix section.

2.4.2 Second round of sample acquisition – Russet Burbank, Mountain Gem Russet, Caribou Russet, and decomposed warted material

During the second round of sampling, 45 more samples were acquired. Out of these samples, 35 samples belonged to Russet Burbank variety, while Caribou Russet and Mountain Gem Russet groups included two samples each. The decomposed wart material sample group included 6 samples total.

The Russet Burbank variety group included 15 control tubers (RBC), 12 infected without visible warts (RBINW), and 8 infected with warts (RBIW) samples. All three groups contained between 102 and 175 proteins and the comparative analysis was conducted in a same way to the Targhee Russet variety analyzed in round 1. Looking at the differences between RBC and RBINW groups, 105 proteins were considered insignificantly different ($p > 0.05$) between the two sample groups, while the other seven were identified as significant (Figure 11A). Three proteins were upregulated and the levels of the other four decreased following infection. The upregulated proteins consisted of patatin B-2, serine protease inhibitor 5, and monodehydroascorbate reductase. These proteins have been previously identified to play a role in plant defense and/or stress response. In comparison, linoleate 9S-lipoxygenase 8 (putatively identified), patatin, annexin, and miraculin were downregulated following the activation of pathogen-triggered immune response in plants. Using Panther software, it was determined that both linoleate 9S-lipoxygenase 8 and patatin play a role in plant defense, annexin is involved in stress response, and miraculin may play a role in taste modification. These data suggest that while some plant defense proteins are produced at a lower rate, the infection with *S. endobioticum* can cause the overexpression of some proteins involved in defense mechanisms. These results correlate to the data obtained in the first round of sampling.
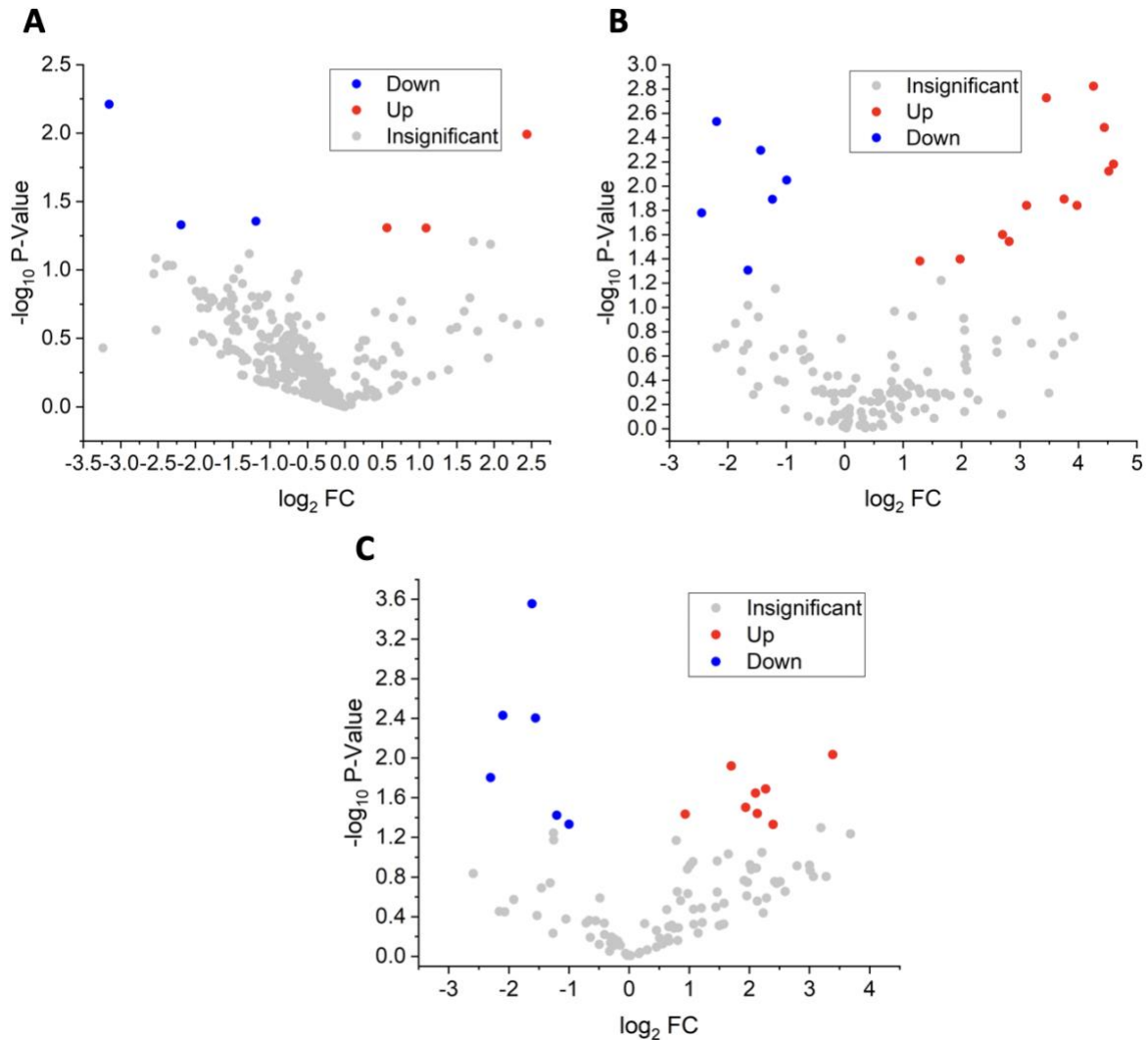
Figure 11. Volcano plot displaying the results of t-test comparisons of A) Russet Burbank control (RBC) versus Russet Burbank infected non-warty (RBINW); B) Russet Burbank control versus Russet Burbank infected with visible warts tuber samples (RBIW); C) Russet Burbank infected non-warty versus Russet Burbank infected warted material.

The comparative proteomic analysis of RBC versus RBIW samples revealed that 18 out of 136 detected proteins were significantly different between the groups and had a p-value less than 0.05 (Figure 11B). Twelve proteins in the significant category were upregulated upon infection, while the levels of six proteins significantly decreased. Five of the downregulated proteins play a role in plant defense and were identified as linoleate 9S-lipoxygenase 8 (putatively identified, based on the annotation provided by UniProt), serine protease inhibitor 2, as well as patatin proteins. One

of the downregulated proteins was α-1,4 glucan phosphorylase and is involved in metabolic processes. The other proteins were upregulated, and their functions varied from protein folding and metabolism to immune response.

Finally, 102 proteins were identified following comparison of RBINW and RBIW groups with 88 of them having a p-value greater than 0.05 (Figure 11C). The eight upregulated proteins were annexin, elongation factor 1-alpha, malate dehydrogenase, CBS domain-containing protein, malic enzyme, nucleoside diphosphate kinase, 5-methyltetrahydropteroyltriglutamate-homocysteine, S-methyltransferase, and ATP synthase subunit beta. Elongation factor 1-alpha is involved in translation, while the other proteins predominantly play a role in metabolism and membrane transport. The downregulated proteins included cysteine protease inhibitor 10, patatin-B2, aspartic protease inhibitor 8, patatin-07, aspartic protease inhibitor 1, and Kunitz-type enzyme inhibitor S9C11. While the cysteine protease inhibitor is involved in regulation of plant growth and development, the rest of downregulated proteins are involved in plant defense. Therefore, the results suggest that *S. endobioticum* interferes with metabolic processes in Russet Burbank plants and causes them to produce metabolism-related enzymes in abundance. This correlates with the results obtained previously. In comparison, this round of proteomic analysis provided clearer data on the effect of potato wart on immune response of the host. The lower levels of plant defense proteins observed in infected tubers suggest that *S. endobioticum* weakens plant immune response, resulting in downregulation of protease inhibitors necessary for inhibition of fungal growth. However, due to the difficulty of determining a specific age (in days or months) of collected tuber material, it is impossible to say whether the inhibition of plant defense proteins occurs during early or late stages of infection.

Both Caribou Russet and Mountain Gem Russet contained one control and one infected tuber sample. Using MaxQuant, only 24 proteins were identified in both groups (Table 1). While some proteins in the infected warted material of Caribou Russet variety had non-zero detection levels, all Mountain Gem Russet infected with warts tubers did not contain any proteins. Despite this difference, all 24 proteins were downregulated upon infection in both Mountain Gem Russet and Caribou Russet. These results correlate with the protein levels observed in the first round of sampling and may suggest that both varieties respond to the infection similarly to Targhee Russet. However, more samples should be obtained to verify these results.

Table 1: Quantified proteins for Caribou Russet and Mountain Gem Russet varieties provided in the second round of sampling.

The table below includes the information about names and functions of the proteins that are present in both varieties of potatoes and shows the peak area values quantified by MaxQuant software.

Protein name	Mountain Gem Russet		Caribou Russet		Protein function
	INW	IW	INW	IW	
Kunitz-type tuber invertase inhibitor	123110000	0	103180000	0	metabolism
Nucleoside diphosphate kinase	105780000	0	42495000	0	metabolism
Metallocarboxypeptidase inhibitor	2174000000	0	1490400000	2958300000	plant defense
Metallocarboxypeptidase inhibitor	36373000	0	0	0	plant defense
catechol oxidase	46903000	0	10336000	0	plant defense
Pathogen-and wound-inducible antifungal protein CBP20	87619000	0	0	0	plant defense
Lipoxygenase	524540000	0	493960000	151400000	plant defense
Aspartic protease inhibitor 11	4130100000	0	7632500000	8772700000	plant defense/plant development
Aspartic protease inhibitor	3583100000	0	623660000	946380000	plant defense/plant development
Patatin	2111600000	0	518320000	798620000	plant defense/storage
Patatin-2-Kuras 3	53891000	0	170720000	0	plant defense/storage
Patatin	43047000	0	357890000	209520000	plant defense/storage
Patatin	357950000	0	713900000	168290000	plant defense/storage
Probable inactive patatin-3-Kuras 1	554180000	0	1488200000	1260300000	plant defense/storage
Patatin group J-1	334660000	0	345120000	174040000	plant defense/storage
Patatin	156090000	0	0	155460000	plant defense/storage
Cysteine protease inhibitor 10 and 8	293110000	0	1539600000	118190000	stress response

Protein name	Mountain Gem Russet		Caribou Russet		Protein function
	INW	IW	INW	IW	
Cysteine protease inhibitor 9	147380000	0	1048200000	1125100000	stress response
Phosphopyruvate hydratase	172620000	0	99807000	0	stress response
Cysteine protease inhibitor 1	413670000	0	91408000	0	stress response
Lactoylglutathione lyase	109520000	0	173780000	133830000	stress response
Probable linoleate 9S-lipoxygenase 3	40502000	0	0	0	stress response
Linoleate 9S-lipoxygenase 1	0	0	43841000	32144000	stress response/plant development
Miraculin	3679600000	0	980900000	1484000000	taste modifier
Miraculin	67197000	0	213560000	45435000	taste modifier

Additionally, CFIA also provided decomposed warted material for sampling. Similar to the other two varieties sampled during this round, only 19 proteins were quantified in six decomposed samples following sample processing and mass spectrometric analysis (Table 2). Four out of six samples (Sample 1, 2, 5, and 6) did not have any detectable proteins, which provided additional confidence levels for the previously acquired data. The other two samples, Sample 3 and 4, had five and nineteen proteins quantified, respectively. Majority of these proteins play a role in metabolism, plant defense, storage, and stress response. As suggested before, the lack of proteins in these samples could be caused as a result of *S. endobioticum* – induced necrosis. The hypersensitive immune response triggered by the plant in response to the fungal infection results in formation of necrotic tissue. The premature death of plant cells causes nutrient deficiency, preventing the metabolic pathways to proceed as normal. Due to the lack of metabolites, the host organism is unable to keep up with nutrient requirements of both the plant and the fungus. The interference of potato wart with plant growth and development may also result in decreased levels of protein synthesis. Alternatively, fungi can also interfere with protein folding, decreasing the activity of produced plant proteins. The other possible reason for low protein number in the decomposed material could be associated with the ability of *S. endobioticum* to synthesize proteases that digest proteins associated with resistance or plant defense; however, more research should be conducted to verify that.

Table 2: Protein groups identified in six decomposed warted samples provided in the second round of sampling.

The table below includes protein names, quantified peak areas, and protein functions of all the proteins identified in all six decomposed material samples.

Protein name	Decomposed material						Protein function
	1	2	3	4	5	6	
Nucleoside diphosphate kinase	0	0	0	275810000	0	0	metabolism
Metallocarboxy-peptidase inhibitor	0	0	2429900000	2278100000	0	0	plant defense
Metallocarboxy-peptidase inhibitor	0	0	0	41292000	0	0	plant defense
Pathogen-and wound-inducible antifungal protein CBP20	0	0	0	39918000	0	0	plant defense
Lipoxygenase	0	0	0	672700000	0	0	plant defense
Aspartic protease inhibitor 11	0	0	4143100000	2052000000	0	0	plant defense and plant development
Aspartic protease inhibitor	0	0	0	4805300000	0	0	plant defense and plant development
Patatin	0	0	0	474720000	0	0	plant defense and storage
Patatin-2-Kuras 3	0	0	38439000	331510000	0	0	plant defense and storage
Patatin	0	0	0	153830000	0	0	plant defense and storage
Probable inactive patatin-3-Kuras 1	0	0	1269300000	3448100000	0	0	plant defense and storage
Patatin group J-1	0	0	0	1380300000	0	0	plant defense and storage
Patatin	0	0	0	119480000	0	0	plant defense and storage
Cysteine protease inhibitor 10 and 8	0	0	0	1594100000	0	0	stress response
Phosphopyruvate hydratase	0	0	0	67462000	0	0	stress response
Lactoylglutathione lyase	0	0	0	134230000	0	0	stress response
Probable linoleate 9S-lipoxygenase 3	0	0	0	27228000	0	0	stress response

Protein name	Decomposed material						Protein function
	1	**2**	**3**	**4**	**5**	**6**	
Miraculin	0	0	0	1020600000	0	0	stress response
Miraculin	0	0	379530000	64683000	0	0	stress response

2.5   Conclusions and Future Work

This study represents the first nLC-MS/MS study conducted to analyze the difference in proteomic profiles between control and potato wart infected tuber samples provided by CFIA. The upregulation of metabolomic proteins obtained for Russet Burbank variety confirms the biotrophic nature of *Synchitrium endobioticum* and its use of plant cells to acquire nutrients. Additionally, the differences observed between the protein levels post- and pre-infection suggest that the varieties analyzed respond differently to potato wart. It is important to point out, however, that those differences could be related to the progression of infection during sampling. Therefore, the future experiments should concentrate on conducting a normalized study where the samples are collected at specific time points. This will confirm whether the proteomic changes observed between varieties in this study are associated with a particular stage of the infection.

While the levels of plant defense proteins reported do not provide a conclusive data on their role in resistance against the fungus, the lack of proteins in infected Targhee Russet, Mountain Gem Russet, and Caribou Russet tubers may suggest facilitation of necrosis in these varieties. Necrotic reaction is induced as a result of hypersensitive immune response, which is believed to occur in varieties resistant to a specific pathotype of *S. endobioticum*. However, more research should be conducted to determine the effects of potato wart on plant defense pathways. This can be accomplished by studying the differences between the protein groups involved in PTI, which is the first line of plant immune response, and proteins that contribute to the effector-triggered immunity.

Analysis of the decomposed material revealed the lack of identifiable proteins in those samples. This might indicate that infection in varieties that trigger activation of hypersensitive response undergo necrosis, resulting in formation of non-viable tissue. This causes death of both host and pathogen cells and the infection is no longer able to proceed.

In addition to identification of differences in immune response between varieties tested, the quantitative proteomic data obtained in this study suggests that there is little to no diversity between control samples and the non-warted tissue. More specifically, the infected tubers tested at the part of the tuber that did not have visual warts present showed high resemblance to the non-infected potato material. This is something that has not been previous documented; however, it confirms that *Synchitrium endobioticum* is a biotroph. Biotrophic fungi heavily rely on its host for survival. Therefore, the localization of infection to a certain area limits the damage to the plant, allowing it to sustain normal metabolic processes. The synthesized nutrients are then used by both plant and fungus, resulting in slow disease progression but production of more zoospores.

Overall, this study provides insight into which protein groups play role in plant-pathogen interactions during potato wart and their potential use as biomarkers for diagnostic screening. The future steps for this work should concentrate on analysis of metabolic changes in the same set of samples to validate collected proteomics data and get a further insight into factors involved in plant resistance pathways. Additionally, microscopy experiments, such as fluorescence imaging, can be conducted to determine the location of proteins downregulated as a result of infection.

3 Chapter 3 - Expression of AvrSen1 and its characterization

3.1 Chapter 3 objectives

S. endobioticum avirulence genes encode for diverse types of secreted effector molecules that can potentially suppress host defense responses and promote host colonization during infection. Similar to many other fungal effector molecules, the intrinsic function of AvrSen1 remains unknown. Previous work done on AvrSen1 consisted of genome mining analysis to identify chromosomal location of a gene encoding for AvrSen1; however, this protein has not been previously expressed or purified using recombinant approach. Understanding its structure and function would aid in determining its role in host resistance and pathogenicity of *S. endobioticum*. Due to the host specificity of *S. endobioticum*, localization and primary structure characterization of AvrSen1 might assist in identification of corresponding resistance genes encoded in potato plants, which could result in successful generation of plants bred to resist infection. To produce recombinant AvrSen1, the gene encoding the protein was synthesized, expressed in *E. coli*, and the recombinant protein was enriched from the bacterial cell lysate via metal affinity chromatography. To ensure AvrSen1 was made correctly in *E. coli*, the protein was immobilized on a resin and analyzed via a bottom-up proteomics approach on a Q-Exactive mass spectrometer.

3.2 Introduction

Fungal pathogens cause numerous infections in plants and account for approximately 10% loss in agronomic production annually. [110] To suppress fungal growth, plants have developed multiple defense strategies. These include pathogen-associated molecular pattern (PAMP)-triggered immunity, or PTI, which is triggered in response to production of certain damage-associated molecules, such as chitin and β-glucan, by the pathogen. [31, 110-114] The PTI is considered the first line of plant defence and leads to changes in hormone biosynthesis, initiation of reactive oxygen species production, and deposition of callose in between the plasma membrane and the cell wall. Another line of host defense includes recognition of fungal effectors, namely avirulence (Avr) genes. This defense mechanism involves recognition of effector molecules by the intracellular receptors, which results in the activation of the hypersensitive response (HR) by the host cells and leads to cell death. [31, 110-112, 114] This type of immunity is referred to as effector-triggered immunity (ETI).

Fungal effector molecules can be both proteinaceous and non-proteinaceous (i.e., secondary metabolites, RNA molecules) in nature and are classified as cytoplasmic and apoplastic molecules. [110, 112-115] Apoplastic effectors inhibit plant immunity mechanisms in the cell wall and surrounding extracellular space through degradation of cell wall constituents, inhibition of chitin-triggered immunity, as well as papain-like cysteine proteases and pathogenesis-related proteins. [115] In contrast, intracellular effectors target protein stability, gene transcription, phytohormone signaling, and receptor mediated defense signaling. [115] Effectors that alter cell structure or the function of the host are considered virulence factors or toxins, while the effector molecules that trigger the immune response in plants are called avirulence factors. [114] Despite having different virulence and avirulence functions, fungal effectors are usually non-enzymatic, rich in cysteine residues and do not exceed 300 amino acid residues in length. [110, 111, 113, 115]

Avirulence genes are extremely diverse and vary between species or even strains. In most cases, their expression is induced by specific plant-pathogen interaction. [111] While the function of many fungal avirulence genes remains unknown, their maintenance in pathogen population might suggest that their expression is beneficial for the pathogen. It is thought that Avr genes encode molecules that are specifically recognized by the host plant, which produce matching resistance (R) genes. [111, 116] According to the gene-for-gene concept, the interaction of an Avr gene product with the corresponding R gene product may occur either directly or through a third component, such as co-receptor. [114, 117] Therefore, if the infected plant contains a resistance gene, the binding of its protein product to the Avr protein would initiate a signal transduction pathway, resulting in necrosis of plant tissue. [114, 116, 117] Alternatively, if the R gene is missing, host cells are considered susceptible and infection progresses. [116] In addition to binding to host targets, fungal effectors can also manipulate host machinery through enzymatic activity, modulation of plant signaling cascades, and altering gene expression through DNA binding. [112] Therefore, effector molecules play an important role in disease development.

AvrSen1 is the only effector molecule that has been previously described for *S. endobioticum.* [31] It is an avirulence protein that consists of 376 amino acids in length and encodes for a putative non-enzymatic protein, which is thought to be recognized by the Sen1 protein produced by resistant potato varieties. [31, 118] The mechanism by which *S. endobioticum* delivers its effector proteins into the host plant cell cytoplasm remains unknown. However, once the fungus has

penetrated the host, it can secrete these effector proteins into the host cell to manipulate the metabolism and physiology of the plant. [119] This is what causes the increase in individual cell size (hypertrophy) and increase in cell production (hyperplasia), which is characteristic of the wart-like growths that develop on the potato meristematic tissue. As described above, AvrSen1-*Sen1* interaction initiates effector-triggered immunity and, therefore, the intracellular recognition of AvrSen1 triggers hypersensitive immune response, resulting in tissue necrosis. Identification of AvrSen1 followed the discovery of *Sen1* gene in plants resistant to pathotype 1(D1); however, the sequencing of additional *S. endobioticum* pathotypes capable of evading plant immune system resulted in detection of multiple variants of *AvrSen1* gene, suggesting its role in fungal virulence. [31, 118] Cloning various potato wart fungus pathotypes showed that five different AvrSen1 variants were found in higher pathotypes that could not trigger a hypersensitive response in Sen1 plants. [119] The main difference between the variants is associated with the presence of single-nucleotide substitutions/insertions. [31, 118] This suggests that AvrSen1 mutated several times during evolution with the purpose of evading detection by the plant's ETI response.

There is a critical lack of structural and functional information for AvrSen1. Biochemical characterization of the protein will aid in better understanding its biological role during potato wart infection. In addition, identification of AvrSen1 structure and its difference between pathotypes could give better insight into the evolution of *S. endobioticum* species, while explaining phenotypic differences between pathotypes.

3.3 Methods

3.3.1 Transformation of designed sequence into competent E. coli cells

AvrSen1 was expressed recombinantly in *E. coli* BL21(DE3) competent cells. A series of AvrSen1 constructs were designed and contained increasing amounts of the open reading frame of the protein. The goal was to identify a construct of AvrSen1 that expressed at high levels and produced functional and stable protein. N- and C-terminal truncations were designed based on predicted secondary structure of the protein. Overall, five different AvrSen1 clones were synthesized: CPG102 (151 a.a. long), CPG103 (181 a.a. long), CPG104 (256 a.a. long), CPG105 (305 a.a. long), and CPG106 (375 a.a. long), respectively (Figure 12). While the size of AvrSen1 has been

previously reported to be 376 amino acids long, the only FASTA sequence for this protein deposited to National Center for Biotechnology Information (NCBI) database contains 375 amino acids. All five sequences were synthesized as 6X-His-tagged N-terminal MBP fusion proteins (Twist Biosciences, CA, USA).
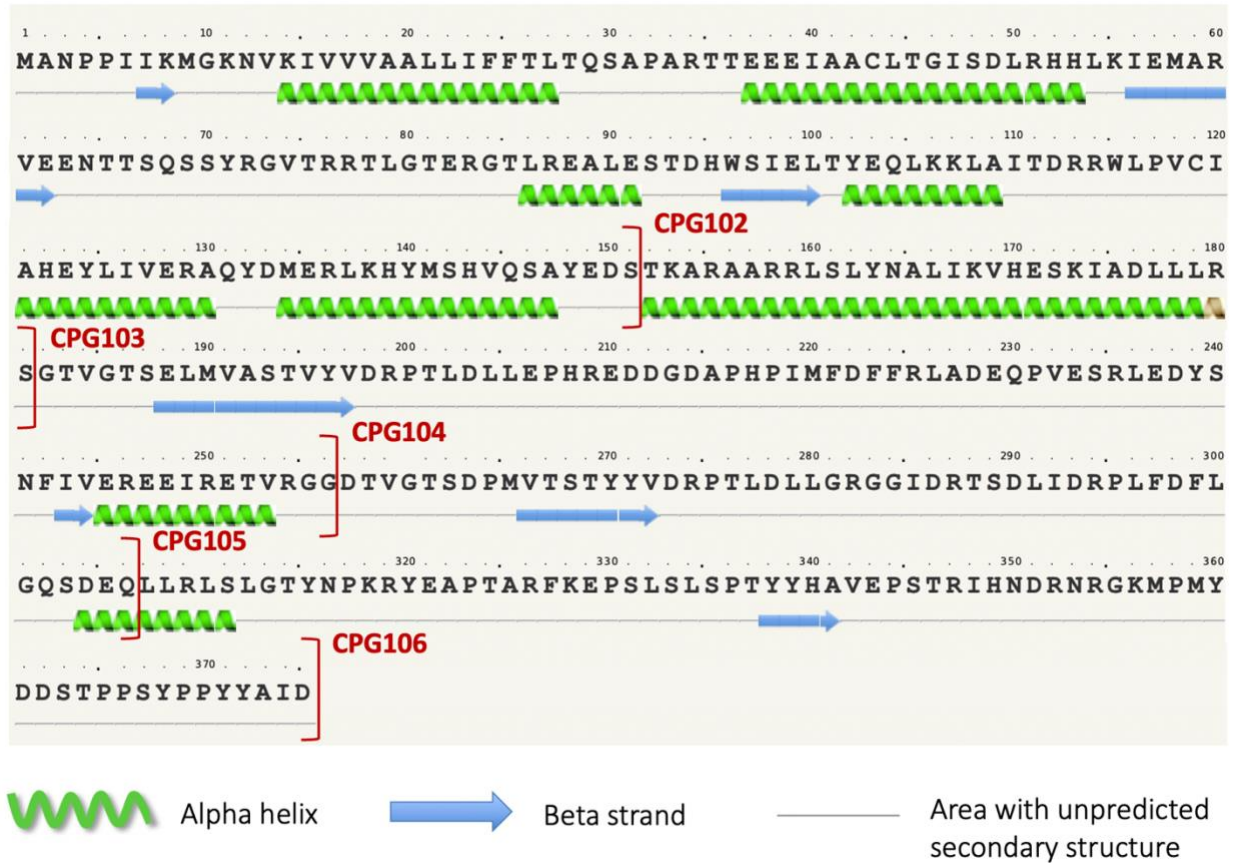
Figure 12. Schematic diagram of designed AvrSen1 homologs. The location of alpha helices and beta sheets was predicted by Phyre2 web portal based on the amino acid sequence of a full-length AvrSen1. [120] While CPG106 sequence represents the full-length AvrSen1, CPG102 was designed to encode for amino acids 1-151, CPG103 encodes for amino acids 1-181, CPG104 consists of 256 amino acids, and CPG105 consists of 305 amino acids.

Since the recombinant AvrSen1 has not been expressed and purified before, choosing the sequence that encodes for functional, stable and soluble protein is critical. CPG102 and CPG103 (151 a.a. long and 181 a.a. long, respectively) contain multiple α-helices and β-sheets and were predicted to have a relatively folded structure. CPG104 (256 a.a. long) and CPG105 (305 a.a. long) represent longer sequences that have regions without predicted secondary structures; however, they have

been previously identified as the predominant AvrSen1 variants in *S. endobioticum* pathotypes in Europe and Canada, respectively, and were, therefore, selected for this study. In comparison, CPG106 represents the full-length protein and was chosen for this study as it is the most accurate representation of the native structure of this effector molecule.

3.3.2 Protein expression in liquid E. coli cultures and cellular lysis

Each construct of AvrSen1 was transformed into the *E. coli* BL21(DE3) expression cell line. Individual colonies of each clone were grown overnight (O/N) at 37°C in 20 mL LB medium + kanamycin (50 μg/mL). Overnight liquid cultures were transferred into 1 L of freshly prepared LB medium supplemented with kanamycin at the same concentration as before. Cells were grown at 37°C with shaking until an OD_{600} value of 0.8-1.0. Protein expression was then induced via the addition of IPTG to a final concentration of 0.05 mM, and cells were grown O/N at 16°C with shaking.

Cells were pelleted via centrifugation, and the pellets were resuspended in 35 mL of resuspension buffer (50 mM sodium phosphate (pH 8), 500 mM NaCl, 10 mM imidazole, 5mM β-Mercaptoethanol (β-ME), 0.5 mg/mL lysozyme; and cOmplete™ protease inhibitor cocktail at one tablet per 50 mL of buffer concentration [MilliporeSigma, Burlington, MA, United States]). Resuspended cells were sonicated for a total of 5 minutes at 30 second on/30 seconds off cycle at 30% amplitude. To clear the lysate from cellular debris, the lysate was centrifuged at 15,000 × g at 4°C for 15 minutes. Following centrifugation, the bacterial pellet was disposed, and the remaining supernatant was used in the steps described below.

3.3.3 Purification of AvrSen1 using metal affinity and size-exclusion chromatography

Recombinant AvrSen1 proteins were enriched from clarified lysate supernatants via immobilized metal affinity chromatography using an Econo-Column® (Bio-Rad Laboratories, Inc., Hercules, CA, United States) packed with Ni Sepharose™ 6 Fast Flow resin (Cytiva, Washington, DC, United States). The Ni was equilibrated with resuspension buffer (50 mM sodium phosphate (pH 8), 500 mM NaCl, 10 mM imidazole, 5mM β-Mercaptoethanol (β-ME), 0.5 mg/mL lysozyme; and cOmplete™ protease inhibitor cocktail at one tablet per 50 mL of buffer concentration [MilliporeSigma, Burlington, MA, United States]). Clarified lysate was added to the resin and

incubated at 4°C for 60 minutes. The resin was then washed with 15 column volumes of wash buffer (50 mM sodium phosphate buffer, 500 mM NaCl, 25 mM Imidazole, 5 mM β-ME). Protein was eluted in elution buffer (50 mM sodium phosphate buffer, 150 mM NaCl, 500 mM Imidazole, 5 mM β-ME). The collected fractions were then analyzed via SDS-PAGE (Bio-Rad Laboratories, Inc., Hercules, CA, United States) to verify successful expression of AvrSen1.

Following Ni column purification, MBP-AvrSen1 fusion proteins were further purified via gel permeation chromatography using a Next Generation Chromatography (NGC) system (Bio-Rad Laboratories, Inc., Hercules, CA, United States). The sample was loaded onto a pre-equilibrated Superdex™ 200 size-exclusion chromatography (SEC) column equilibrated in 50 mM sodium phosphate buffer (pH 8) and 150 mM NaCl. Fractions were analyzed via SDS-PAGE gel to monitor for the presence of AvrSen1 fusion protein.

Fractions containing AvrSen1 protein were pooled and analyzed on a Q-Exactive mass spectrometer (refer to section 3.3.4.) to ensure the correct protein was produced. The remaining sample was treated with Tobacco Etch Virus (TEV) protease to cleave the MBP-tag from AvrSen1 at a 1:50 (TEV:protein) ratio. Digestion occured at room temperature for 6 hours, followed by overnight incubation at 4°C. The digest was then run on SDS-PAGE gel to determine the digestion efficiency.

3.3.4   nLC-MS/MS analysis of fusion AvrSen1-MBP protein

To verify the production and purification of recombinant AvrSen1 proteins, the collected lysate fractions were analyzed using bottom-up proteomic approach. First, 350 µL of the lysate collected from the Ni column in Section 3.3.3 was added to a 2.0-mL Eppendorf tube containing 40 µL of pre-equilibrated magnetic beads (New England Biolabs, Ipswich, MA, United States). These magnetic beads contain Anti-MBP antibodies that bind to MBP tag of the designed AvrSen1 protein, allowing to purify it from other impurities present in the bacterial lysate. The bead-lysate mixture was incubated for a total of 60 minutes in a thermomixer (Eppendorf, Hamburg, Germany) set to 550 rpm. To remove all the proteins that did not bind to the beads, the microfuge tubes were placed in the magnetic separation rack (Thermo Fisher Scientific, Waltham, MA, United States) to pull the beads to the side of the tube. The supernatant was decanted, and the beads were washed thrice with 0.1 M sodium phosphate buffer (pH 8.0). To elute AvrSen1 protein from the beads, the

bead pellet was resuspended in 40 μL of Laemmli sample buffer (Bio-Rad Laboratories, Inc., Hercules, CA, United States). The tubes were then heated at 70°C for 5 minutes and each sample was loaded onto an SDS-PAGE gel (Bio-Rad Laboratories, Inc., Hercules, CA, United States). The gel was run at 200 V for 35 minutes.

Following SDS-PAGE electrophoresis, the gel was stained with Coomassie Brilliant Blue R-250 staining solution to visualize the bands. The bands corresponding to the size of CPG103 and CPG106 proteins were excised from the gel. Each band was divided into 1 mm × 1 mm pieces and placed together in a 1.5-mL Eppendorf tube, where they were destained using ammonium bicarbonate solution. Briefly, the gel pieces were incubated in 300 μL of 2 mg/mL NH_4HCO_3 solution in ultrapure water, followed by a 10-minute incubation step in a thermomixer set to 37°C and 700 rpm. The solution was then replaced with 300 μL of 2 mg/mL NH_4HCO_3 solution in 50:50 $ACN:H_2O$ mixture, and the incubation was repeated. The washing with ammonium bicarbonate solutions was repeated until all of the sample loading dye was removed. Reduction of proteins was performed by adding 30 μL of 10 mM DTT in 25 mM NH_4HCO_3 solution. The tubes containing the sample were incubated for 10 minutes at 60°C. The samples were cooled to room temperature, the reducing buffer was removed, and the gel pieces were washed with 25 mM ammonium bicarbonate solution to wash away the remaining DTT. Protein alkylation was achieved by incubating samples with 30 μL of 100 mM aqueous iodoacetamide buffer in the dark for a total of one hour. Following reduction of cysteine residues, IAA buffer was decanted, and the samples were incubated with 25 mM NH_4HCO_3 solution for 30 minutes at 37°C and with continuous shaking. The trypsin digestion was done by first shrinking the gel pieces with 50 μL of acetonitrile, airdrying them for 10-15 minutes and adding activated trypsin solution (0.01 μg/μL trypsin in 25 mM NH_4HCO_3). The resulting mixture was incubated for 15 minutes at room temperature prior to adding 45 μL of 25 mM NH_4HCO_3 buffer. The microfuge tubes were then left to incubate overnight at 32°C with constant shaking at 300 rpm. The next morning, the digestion mixture was placed in a new 1.5-mL Eppendorf tube, while the remaining gel pieces were mixed with 10 μL of 1% FA solution. Following 5-minute incubation at 32°C and 300 rpm, the extraction solution was added to the digestion mixture. Lastly, 75 μL of 5% ACN solution was added to the peptide samples.

The nLC-MS/MS analysis was performed using EASY-nLC 1000 system connected to Q-Exactive mass spectrometer. The same parameters described in Section 2.3.4 were applied. The qualitative analysis was performed using SearchGUI software (v. 3.3.3.) in combination with Peptide Shaker (v. 1.16.26.) to identify proteins, as well as the number of peptides, detected.

3.4   Results and Discussion

Following transformation of selected AvrSen1 sequences into competent *E. coli* cells and their purification via Ni-NTA chromatography, the expression of all five designed sequences was assessed by running their fractions on the SDS-PAGE gel. The CPG103 encodes for a 181 amino acid long or 20.7 kDa AvrSen1 and had the best expression out of all five sequences, and CPG106 represents the full-length AvrSen1 protein, which is why only these two sequences were selected for MS analysis. The other three sequences either showed inconsistent expression post-transformation, were produced in low quantities or were insoluble in Na_3PO_4 buffer used in the protocol.

The incubation of CPG103 and CPG106 with Anti-MBP magnetic beads allowed further purification of each protein from other contaminants present in the bacterial lysate. Following the gel staining step, bands corresponding to the approximate size of the protein (62.8 kDa and 84.8 kDa for CPG103 and CPG106, respectively) were cut out from the gel (Figure 13).
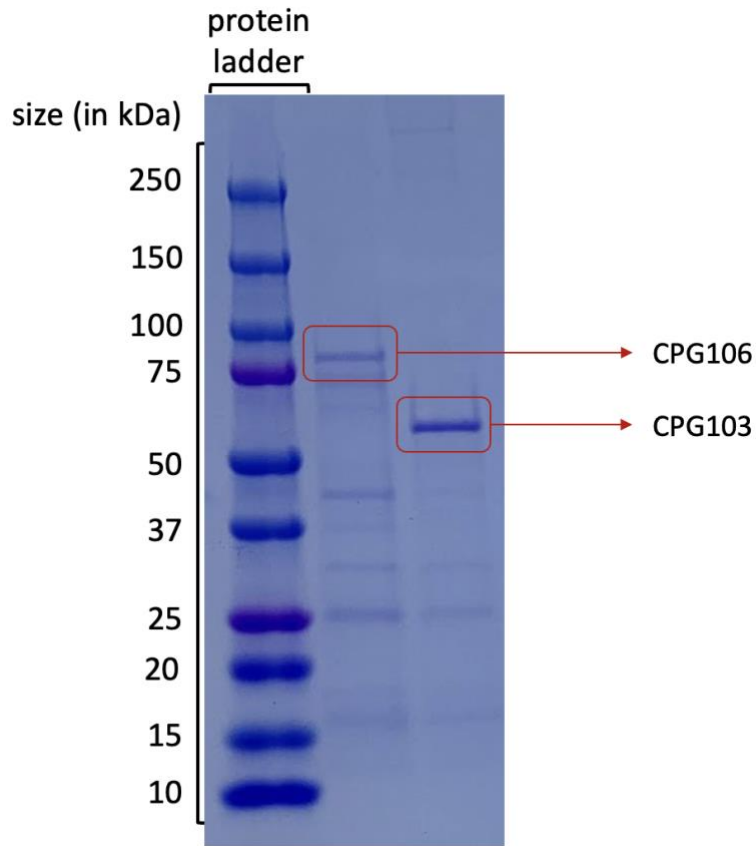
Figure 13. SDS-PAGE analysis of CPG103 and CPG106 after MBP-affinity enrichment. Running bacterial lysate samples containing CPG103 and CPG106 proteins resulted in acquisition of two main bands. The CPG106 band (approximately 84.8 kDa in size) was observed on the gel in between 75 kDa and 100 kDa protein ladder bands. In comparison, CPG103 band (predicted size of 62.8 kDa) was observed in between 50 kDa and 75 kDa ladder bands.

To confirm the identity of the putative AvrSen1 bands, CPG103 and CPG106 were digested with trypsin and run on the Q-Exactive mass spectrometer. The mass spectrometric analysis of the 181 a.a. long AvrSen1 protein bound to MBP resulted in detection of six proteins (Table 3).

Table 3: Protein detected following proteomics analysis of CPG103 protein enriched via anti-MBP magnetic beads.

The number of PSMs refer to the number of peptide-spectrum matches. It shows the total number of all peptides detected for the protein.

	Protein name	Protein accession number	MW [kDa]	Coverage [%]	#Validated Peptides	#Validated PSMs	Confidence [%]
1	CPG103-MBP+AvrSen1	-	62.72	89.01	117	562	100
2	Cytochrome P450	M1B865	48.94	2.08	1	0	88
3	Heat shock protein	M0ZM32	110.25	1.53	1	1	100
4	Calcium uniporter protein	M0ZL04	24.90	5.86	1	1	86
5	Elongation factor Tu	M1B640	35.69	5.21	1	1	100
6	50S ribosomal protein L15	M1B5R1	29.22	3.99	1	1	100

CPG103 was the main protein detected in the sample fraction. The 117 peptides identified for CPG103 provide 89% coverage of the whole protein sequence (Figure 14). Since the expressed protein of interest underwent two purification steps, which included IMAC and SEC, prior to its analysis on the mass spectrometer, it was assumed that majority of contaminants were removed from the lysate. Therefore, the proteomic data correlates with the expected results and verifies that the expressed protein was in fact AvrSen1. The other five proteins present in the mixture had only one peptide detected each (Table 3). This might suggest that other proteins present in the cleared lysate co-migrated with CPG103. Alternatively, identification of those proteins in the sample shows a potential contamination of either the beads, buffers or microfuge tubes used.

A

GSSMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGL
LAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYF
TWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAW
SNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAK
DPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNGIEEENLYFQSNAMANPPIIKMG
KNVKIVVVAALLIFFTLTQSAPARTTEEEIAACLTGISDLRHHLKIEMARVEENTTSQSSYRGVTRRTLGTERGTLREALEST
DHWSIELTYEQLKKLAITDRRWLPVCIAHEYLIVERAQYDMERLKHYMSHVQSAYEDSTKARAARRLSLYNALIKVHESKI
ADLLLRS

B

GSSMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGL
LAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYF
TWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAW
SNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAK
DPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNGIEEENLYFQSNAMANPPIIKMG
KNVKIVVVAALLIFFTLTQSAPARTTEEEIAACLTGISDLRHHLKIEMARVEENTTSQSSYRGVTRRTLGTERGTLREALEST
DHWSIELTYEQLKKLAITDRRWLPVCIAHEYLIVERAQYDMERLKHYMSHVQSAYEDSTKARAARRLSLYNALIKVHESKI
ADLLLRS

　　　 AvrSen1 portion of the sequence

　　　 peptide sequences detected

Figure 14. Bottom-up nLC-MS/MS analysis of CPG103. Figure 14A shows an amino acid sequence of CPG103 protein, with AvrSen1 portion of the protein highlighted in pink. Figure 14B shows the sequence coverage of CPG103 based on the peptides detected using Q-Exactive mass spectrometer.

The results of bottom-up analysis of CPG106-containing sample looked similar to the CPG103 protein. CPG106 was the predominant protein present in the sample, with a total of 156 peptides detected by mass spectrometry (Table 4). Similar to CPG103, the other proteins present in the sample had no more than one peptide identified with overall coverage varying from 2.08% to 15.29%. The peptide coverage for CPG106 was determined to be approximately 86% and is visualized on Figure 15.

Table 4: Protein detected following incubation of CPG106 protein with Anti-MBP magnetic beads.

	Protein name	Protein accession number	MW [kDa]	Coverage [%]	#Validated Peptides	#Validated PSMs	Confidence [%]
1	Cytochrome P450	M1B865	48.94	2.08	1	3	100
2	Probable inactive patatin-3-Kuras 1	Q3YJS9	41.17	4.28	1	1	100
3	Integrase core domain containing protein	M1DG18	14.66	9.92	1	1	100
4	Serine protease inhibitor 7	P30941	24.01	7.69	1	3	100
5	ATP-dependent clp protease ATP-binding subunit clpx	M1AIM3	61.99	2.28	1	2	100
7	ATP synthase subunit alpha, mitochondrial	M1B325	9.43	15.29	1	4	100
9	Replicase	M1C030	45.66	2.86	1	1	100
10	CPG106-MBP+AvrSen	-	84.78	85.75	156	1128	100
12	Heat shock protein 70-3	M0ZHT9	71.16	2.47	1	2	100
13	50S ribosomal protein L15	M1B5R1	29.22	3.99	1	4	100

A

GSSMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGL
LAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYF
TWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAW
SNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAK
DPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNGIEENLYFQSNA**MANPPIIKMG
KNVKIVVVAALLIFFTLTQSAPARTTEEEIAACLTGISDLRHHLKIEMARVEENTTSQSSYRGVTRRTLGTERGTLREALEST
DHWSIELTYEQLKKLAITDRRWLPVCIAHEYLIVERAQYDMERLKHYMSHVQSAYEDSTKARAARRLSLYNALIKVHESKI
ADLLLRSGTVGTSELMVASTVYVDRPTLDLLEPHREDDGDAPHPIMFDFFRLADEQPVESRLEDYSNFIVEREEIRETVRG
GDTVGTSDPMVTSTYYVDRPTLDLLGRGGIDRTSDLIDRPLFDFLGQSDEQLLRLSLGTYNPKRYEAPTARFKEPSLSLSPT
YYHAVEPSTRIHNDRNRGKMPMYDDSTPPSYPPYYAID**

B

GSSMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGL
LAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYF
TWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAW
SNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAK
DPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNGIEENLYFQSNAMANPPIIKMG
KNVKIVVVAALLIFFTLTQSAPARTTEEEIAACLTGISDLRHHLKIEMARVEENTTSQSSYRGVTRRTLGTERGTLREALEST
DHWSIELTYEQLKKLAITDRRWLPVCIAHEYLIVERAQYDMERLKHYMSHVQSAYEDSTKARAARRLSLYNALIKVHESKI
ADLLLRSGTVGTSELMVASTVYVDRPTLDLLEPHREDDGDAPHPIMFDFFRLADEQPVESRLEDYSNFIVEREEIRETVRG
GDTVGTSDPMVTSTYYVDRPTLDLLGRGGIDRTSDLIDRPLFDFLGQSDEQLLRLSLGTYNPKRYEAPTARFKEPSLSLSPT
YYHAVEPSTRIHNDRNRGKMPMYDDSTPPSYPPYYAID

▇ AvrSen1 portion of the sequence

▇ peptide sequences detected

Figure 15. Bottom-up nLC-MS/MS analysis of CPG106. Figure 15A shows an amino acid sequence of CPG106 protein, with AvrSen1 portion of the protein highlighted in pink. Figure 15B shows the sequence coverage of CPG106 based on the peptides detected using Q-Exactive mass spectrometer.

Overall, the nLC-MS/MS analysis of AvrSen1-containing bacterial lysates confirmed the successful expression of both CPG103 and CPG106. Since no one has worked with recombinant AvrSen1 before, it was critical to verify that *E. coli* was able to produce this protein correctly. Therefore, this data provides additional confidence level that the ordered AvrSen1 sequences were transformed into bacteria successfully. However, despite successful expression of CPG106, it was determined that CPG103, which has a length of a 181 amino acids, had the best expression. Due to its expression efficiency and high solubility levels, this homolog of AvrSen1 has been chosen for initial protein analysis and purification experiments.

Following extraction of CPG103 from bacterial cells, it was purified via Ni Sepharose™ 6 Fast Flow and Superdex™ 200 columns to remove other bacterial proteins from the sample. The size-exclusion analysis resulted in detection of two main peaks which included MBP-tagged AvrSen1 and MBP alone, respectively (Figure 16).
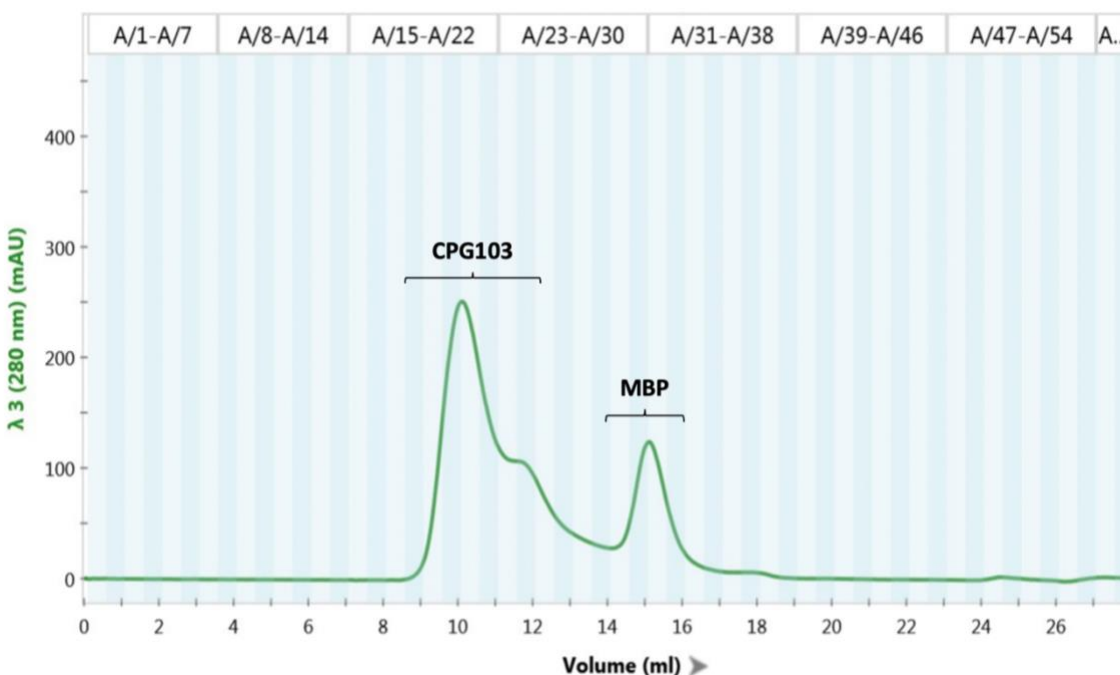
Figure 16. Chromatogram obtained following size-exclusion analysis on NGC system. By running bacterial lysate containing fused AvrSen1 protein on Superdex™ 200 column, two main peaks get detected at 280 nm wavelength. Running multiple fractions of each peak on SDS-PAGE gel allowed to identify that the first peak contains CPG103 (or MBP-bound) AvrSen1protein, while the second peak consists of MBP alone.

To verify the size and purity of the protein, the fractions collected as a part of the CPG103 peak were run on SDS-PAGE gel (Figure 17). All fractions had one band that appeared on the gel in between 50 kDa and 75 kDa bands of the protein ladder, which is consistent with the predicted size of the protein. The presence of one thick band also confirms the efficiency of the protein purification.

Once the protein expression and size was confirmed, all CPG103 fractions were combined together and the protein concentration was measured using NanoDrop Microvolume Spectrometer (Thermo Fisher Scientific, Waltham, MA, United States).
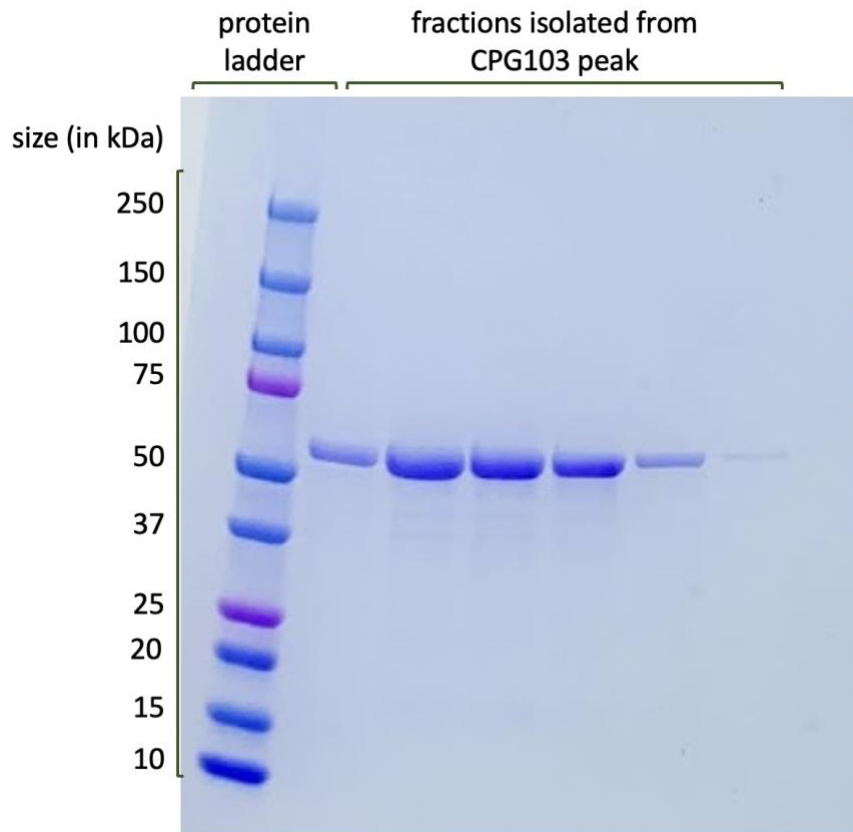
Figure 17. Qualitative analysis of CPG103 fractions isolated from SEC column. Fractions that were collected as a part of the first peak on Superdex™ 200 column formed one predominant band on the gel. The size of this band is in between 50 kDa and 75 kDa, which is consistent with the estimated size of 62.3 kDa for CPG103 protein.

The next steps involved enzymatic digestion of the protein with the purpose of removing MBP tag from the AvrSen1 portion of the protein. The TEV enzyme was added at the 1:50 ratio of enzyme to protein. The resulting mixture was left to incubate for 6 hours at room temperature. To check the efficiency of digestion, the sample was run on SDS-PAGE gel (Figure 18).
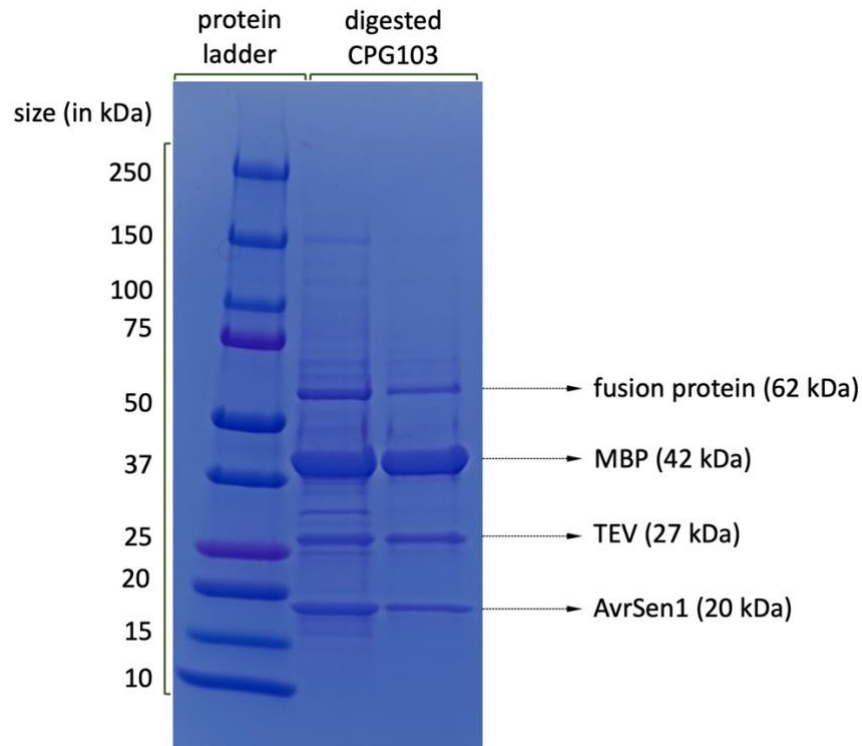
Figure 18. Analysis of TEV-digested CPG103 protein on an SDS-PAGE gel. Following digestion of obtained CPG103 protein with TEV enzyme, the resulting sample digest was run on SDS-PAGE gel. Apart from the presence of minor contaminants, four main bands were observed which corresponded to the sizes of CPG103 protein (62 kDa), MBP alone (42 kDa), TEV enzyme (27 kDa) and AvrSen1 alone (20 kDa), respectively.

Looking at the gel, it can be seen that digestion of the fusion protein was successful as it resulted in the appearance of four main bands corresponding to fusion protein, MBP, TEV, AvrSen1, respectively. The next steps focused on isolation of AvrSen1 from the sample digest; however, all the protocols that I've tried to date to accomplish this were unsuccessful. The future steps for this project are further described in the Conclusions and Future Work section of this chapter.

3.5   Conclusions and Future Work

Preliminary results collected for this project verified the successful expression and purification of two AvrSen1 proteins in *E. coli* cells. The immobilization of AvrSen1 on magnetic beads and its further analysis using a bottom-up approach allowed for identification of numerous peptides that were expected to be present based within the protein sequence used for expression. For the 181 amino acids long and the full-length AvrSen1 sequence, a total of 117 and 156 peptides were identified, respectively. Considering that the overall coverage obtained was 89.01 % and 85.75%, this study provides an additional confidence level of the obtained results and confirms successful AvrSen1 expression in *E. coli* system. However, while the data represent the first mass spectrometric analysis conducted on the effector molecule of *Synchitrium endobioticum,* the native structure of AvrSen1 still remains unknown.

Future work should concentrate on identification of AvrSen1 protein that behaves well in solution and can be produced at biochemical levels (in milligram quantities). CPG103 expressed here may behave differently to the full-length AvrSen1 due to it having different physical and chemical properties. The future experiments also ought to focus on separation of AvrSen1 from MBP tag and mass spectrometric analysis of the purified AvrSen1 using top-down proteomics. Performing the top-down analysis of a purified protein should provide a better understanding regarding the structure of this fungal protein, as well as reveal any potential PTMs or metal interactions. Once homogenous recombinant AvrSen1 is produced, a major goal of this project includes determining the X-ray crystal structure of the protein, which will help in understanding the function of the protein.

Additionally, current work has shown that the use of *Escherichia coli* as a heterologous host for AvrSen1 expression results in production of a unstable protein, suggesting the low efficiency of this system. Therefore, the future work will also concentrate on utilizing *Saccharomyces cerevisiae* or *Pichia pastoris* yeast cells and *Nicotiana benthamiana* plants to identify a suitable heterologous host for production of functional protein. Successful purification and acquisition of AvrSen1 crystals will result in identification of its structure, which will allow functional assays to better understand the type of proteins AvrSen1 interacts with upon penetration of the host membrane. Additionally, conducting a pull-down experiment of AvrSen1 with protein mixture extracted from resistant potato varieties might identify the Sen1 protein. The specificity of AvrSen1-Sen1 interaction and its role in induction of hypersensitive immune response in host cells suggest the

importance of Sen1 in resistance to infection. Therefore, identification of this protein could promote the identification of the resistance loci regions in *Solanum tuberosum*. This would allow an improved understanding of the infection mechanisms at the molecular level.

4 Chapter 4 - Investigation of protein interaction partners of AvrSen1

4.1 Chapter 4 objectives

AvrSen1 is the only effector molecule that has been previously described to be associated with potato wart; however, its role in suppression of pathogen-triggered immunity in *Solanum tuberosum* plants remains unknown. [31] Therefore, understanding which potato proteins interact with AvrSen1 might help to expand our knowledge on the function of this protein, as well as its localization. This can be accomplished by isolating protein groups from various tissue material of potato plants and establishing co-immunoprecipitation reaction with the AvrSen1-containing fungal lysate. Hence, the objective of this study is to determine the interaction partners of AvrSen1 using immunoprecipitation approach coupled with mass spectrometry analysis.

4.2 Introduction

Phytopathogens, like *S. endobioticum*, secrete effector molecules that act to suppress plant immune defenses and reprogram the host cell's metabolism for its benefit. Investigating protein-protein interactions between the fungal protein of interest and plant proteins can yield crucial information about the function of fungal effector molecules. [121, 122]

One of the techniques that can be used to study protein-protein interactions following infection is called co-immunoprecipitation (Co-IP). This approach involves expressing and harvesting the protein of interest in the cell culture extract, followed by its binding to a solid support, such as magnetic beads. [121-123] To differentiate between the protein of interest and the rest of bacterial proteins present in the lysate, the magnetic beads are usually coated with a tag or an antibody that is specific to the target protein, resulting in its immobilization. The rest of the proteins remain unbound in solution and are removed during the washing steps. [123] The beads are then enriched with the protein solution that contains potential interaction partners. Unbound proteins are washed away, leaving only the target protein and its interaction partners attached to the beads. Depending on the goal of the analysis, the bound proteins can either be eluted and analyzed qualitatively on the SDS-PAGE gel or they can undergo on-bead trypsin digestion and identification using analytical techniques, such as mass spectrometry. The combination of co-immunoprecipitation technique with proteomics is referred to as Co-IP-MS or IP-MS. [121, 123]

This study focused on the detection of AvrSen1 interaction partners in the leaf and tuber material of Russet Kennebec potato variety to better understand the function of AvrSen1 during infection. The interaction partners from both types of tissue were isolated using Co-IP technique, followed by their analysis on a nano-LC-MS/MS instrument. More specifically, MBP-AvrSen1 fusion protein was attached to the magnetic beads covalently coupled with anti-MBP antibodies. The MBP-tag of AvrSen1 acts as an antigen, resulting in strong antigen-antibody interaction between the fusion protein and the antibody molecule attached to the beads.

4.3 Methods

4.3.1 Transformation of AvrSen1 into competent E. coli cells

To transform AvrSen1 protein into competent bacterial cells, the protocol described in "3.3.1 - Transformation of designed sequence into competent *E. coli* cells" section was followed. Overall, two different sizes of AvrSen1 were used, the full-length protein (CPG106) which consists of 375 amino acids, and a shorter AvrSen1 homolog (CPG103) which is 181 amino acids long. Despite poor expression of CPG106 in competent bacterial cells, it should have a structure similar to the native protein. It would, therefore, provide a better understanding of what type of host-pathogen protein interactions occur during infection. In comparison, CPG103 has shown the best solubility and expression levels out of all the ordered AvrSen1 sequences. Therefore, in this study, both CPG103 and CPG106 were used. The MBP sequence was used as a negative control for this experiment and its expression was accomplished using the same steps that were used for AvrSen1 sequences.

4.3.2 Expression of AvrSen1 and its isolation from E. coli liquid cultures

The procedure described in "3.3.2 – Protein expression in liquid *E. coli* cells" section was used, without any modifications. Before proceeding to the next step, the obtained supernatant was run on SDS-PAGE gel to ensure that AvrSen1 was expressed and to verify the sizes of AvrSen1 sequences. The parameters used for gel electrophoresis remained unchanged.

4.3.3 Extraction of leaf and tuber proteins

The leaf and tuber material used for this experiment was extracted from one-month old Kennebec potato plants. More specifically, the leaf samples were obtained by cutting one leaf at a time using

tissue scissors. To extract the tuber material from plants, a 7 mm corkborer was used to collect small tissue fragments. Following extraction, both leaf and tuber fragments were placed into separate 50-mL Falcon tubes and were immediately snap-frozen. Both tuber and leaf samples were then placed at -80°C overnight and were then freeze-dried using a two-day lyophilization cycle. After the lyophilization step was complete, the mass of each sample was recorded. The freeze-dried material was then homogenized at 1000 rpm for 60 seconds using stainless steel beads.

The protein extraction protocol was adapted from Singh et al.[121], with minor modifications. The homogenized material was resuspended with Co-IP buffer (100 mM TRIS-HCl, pH 7.5; 150 mM NaCl; 1 mM EDTA; 10 mM $MgCl_2$; 10% glycerol; 0.2% Triton; 5 mM DTT; and 1X protease inhibitor cocktail tablet) at 1:10 and 1:1 for leaf and tuber material, respectively (sample:buffer). The samples were vortexed, incubated in the thermomixer for 30 minutes at 4°C and 600 rpm, and centrifuged for 60 minutes at 4°C and 13,000 rpm. The concentration of protein in the supernatant was quantified using Pierce™ Rapid Gold BCA Protein Assay Kit (Thermo Fisher Scientific, Waltham, MA, United States). The proteins were also analyzed qualitatively by running 10 µL of the sample supernatant on the SDS-PAGE gel at 200 V for 35 minutes.

4.3.4   Incubation of proteins with bead-attached AvrSen1

To isolate interaction partners of AvrSen1 from the rest of soluble proteins present in potato lysate solution, the Anti-MBP magnetic beads were used. First, the MBP-bound AvrSen1 protein was attached to the beads following the protocol supplied by New England Biolabs[124]. Briefly, the beads were first thoroughly vortexed for 15 seconds, followed by the transfer of 40 µL of bead suspension into a new 1.5-mL Eppendorf tube. The conditioning step included washing the beads thrice with 500 µL of 0.1 M sodium phosphate buffer (pH 8.0) with 30-second incubations during each wash. To attach MBP-fusion AvrSen1 protein to the beads, 350 µL of cell culture extract containing the protein of interest was thoroughly mixed with Anti-MBP magnetic beads and incubated for 60 minutes at 4°C and 550 rpm. Following the incubation step, the supernatant was decanted, and the beads were washed with 0.1 M sodium phosphate buffer (pH 8.0) to prevent non-specific binding of other proteins present in the bacterial lysate.

To facilitate binding of leaf and tuber proteins, 350 µL of protein solution was added to the beads, followed by three-day incubation at 600 rpm and 4°C. The flow-through was then removed from

each sample and the beads were washed with 0.1 M sodium phosphate buffer (pH 8.0) again. All of the supernatant fractions were collected to identify proteins that did not bind to the beads.

4.3.5   Trypsin digestion of interaction proteins

The protocol used to enzymatically digest bound proteins was adapted from The Proteomics Unit at University of Bergen[125], with some modifications. The beads were first resuspended in 200 µL of phosphate-buffered saline (PBS) buffer (pH 7.4). The supernatant was discarded, and the PBS wash was repeated twice. To facilitate reduction of peptide bonds, the beads were mixed with 40 µL of 50 mM Tris-HCl buffer (pH 7.9), followed by addition 4 µL of 0.1M DTT to the tube and a 5-minute incubation at 95°C. Once the samples cooled down to room temperature, the proteins were alkylated with 5 µL of 0.2 M IAA. Following the incubation step at room temperature and in the dark for a total of 60 minutes, 0.8 µL of 0.1 M DTT was added and the sample solutions were incubated for 10 minutes at room temperature. The enzymatic digestion was performed by adding 5 µL of trypsin working solution to the bead, which was prepared by mixing 2.5 µL of 2 µg/µL trypsin stock with 47.5 µL of 50 mM NH_4HCO_3. The samples were incubated at 37°C overnight with slight agitation. Following incubation with trypsin, 1.1 µL of 5% formic acid (FA) was added to each sample to reach the final concentration of 0.1% FA.

4.3.6   Sample clean-up using solid-phase extraction

To clean up the samples prior to analysis on the mass spectrometer, the solid-phase extraction (SPE) protocol was used. To do that, Oasis® HLB 1cc (30 mg) Extraction Cartridges were activated with three millilitres of methanol and equilibrated twice with 1 mL of 0.1% aqueous solution of formic acid. To load samples to the cartridges, the peptide solutions were carefully added to the SPE cartridges pre-filled with 500 µL of 0.1% FA. This mixture was then allowed to slowly drip through the cartridge, followed by the washing step with 400 µL of 0.1% FA. The cartridges were then dried, and the peptides were eluted with 400 µL of 70% acetonitrile (ACN) solution. To concentrate digested proteins on the bottom of the tube, the samples were dried using CentriVap Centrifugal Vacuum Concentrator (Labconco Corporation, Kansas City, MO, United States). Once all the solvent evaporated, the dried peptide pellet was reconstituted in 200 µL of

5% ACN, 0.1% FA solution. Prior to MS analysis, all samples were transferred into 250-µL snap cap polypropylene HPLC vials (Agilent Technologies, Santa Clara, CA, United States).

4.3.7 Analysis on Thermo Q-Exactive Orbitrap mass spectrometer

All proteomics data were acquired using Thermo Q-Exactive Orbitrap mass spectrometer fitted with Nanospray Flex™ ion source and coupled to an EASY-nLC 1000 system (Thermo Fisher Scientific, Waltham, MA, United States). Each sample (10 µL) was injected onto Acclaim PepMap™ RSLC C18 column. The peptides were resolved with 0.1% FA in Optima™ LC/MS grade water and 0.1% in ACN mobile phases. The MS analysis was conducted using data-dependent acquisition (DDA) mode comprised of full MS scan in the mass range from 340 m/z to 1,800 m/z. The scans were obtained at 70,000 resolution, with maximum injection time (IT) set to 256 ms and automatic gain control (AGC) target of 1×10^6. The 17,500-resolution MS/MS scans were performed on the top 10 peptides in every window. The AGC target was kept at 1×10^6, while the maximum IT was increased to 340 ms.

4.3.8 Data analysis

To annotate identified proteins, the qualitative analysis was performed using SearchGUI software. More specifically, the obtained DDA-MS raw files were first converted into mgf format using MS Convert (version: 3.0.22166-28b 1b 7b). The sample files were then searched against *Solanum tuberosum* reference proteome obtained from UniProt (UP000011115) with the addition of the FASTA sequences of CPG103, CPG106, and MBP. The majority of settings were kept as default; however, the carbamidomethylation of C was selected under fixed modifications and oxidation of M was chosen for variable modifications. The carbamidomethylation of cysteine is a type of post-translational modification that gets introduced following interaction of the protein with IAA. [126] Similarly, oxidation of methionine is one of the predominant post-translational oxidations in proteins and occurs during sample processing. [127] In addition, the fragment m/z tolerance was changed from 0.5 Da to 0.02 Da. Following SearchGUI analysis, the identified protein accession numbers, as well as the columns with the molecular weight, percent coverage, peptide number, and validation level, were transferred into a new Excel document. All proteins with less than five peptides detected were removed from the data.

The protein groups identified in both half-sized and full-length AvrSen1 were compared to the proteins detected in MBP sample. The proteins that were present in both the sample groups and the negative control were filtered out. The rest of proteins were sorted based on their function in *Solanum tuberosum* plants.

4.4   Results and Discussion

Prior to attachment of MBP-fused AvrSen1 proteins to the magnetic beads, the expression of all three protein sequences was assessed by running SDS-PAGE gel (Figure 19). The samples were run in replicates to ensure the accuracy of their size. The MBP is 42.5 kDa protein and its band appeared on the gel between 37 kDa and 50 kDa when compared to the bands of the standard ladder, which correlates with its expected size. The CPG103 sequence, which consists of half-sized AvrSen1 protein and MBP, encodes for a fusion protein of 63.2 kDa in size. Both lanes containing CPG103 had a thick band that migrated the similar distance as 50 kDa ladder band. Similarly, CPG106 sequence consists of a full AvrSen1 protein and MBP and should be 85.3 kDa in size. Both CPG106 samples had a band eluted close to 75 kDa band of the standard used; however, the size and thickness of the band suggests that the protein was not well expressed or that it might be insoluble.

After the bacterial lysates containing the fungal protein and negative control were incubated with both leaf and tuber mixture, the bound proteins were eluted from the beads to test and visualize the difference between the interaction affinity between maltose-binding protein, half-sized AvrSen1 and the full-length AvrSen1 protein.
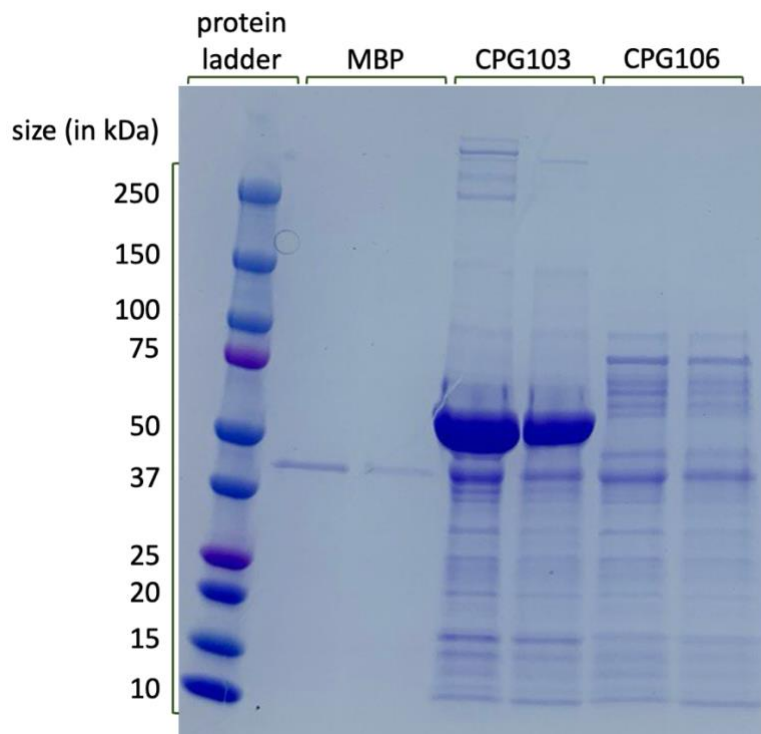
Figure 19. Expression of extracted MBP, CPG103 and CPG106 sequences. Following expression and extraction of MBP and both half-sized and full-length AvrSen1 proteins, their size and expression efficiency were confirmed by running SDS-PAGE electrophoresis. Based on the bands obtained, all three proteins expressed successfully and were of the right size.

The proteomic analysis of CPG103 and CPG106 fractions eluted from the beads resulted in identification of 116 leaf and 203 tuber interaction partners, respectively (Table S1 and Table S2 in Appendix). When comparing the proteins detected in leaf material, 49 proteins interacted with half-sized AvrSen1 protein, while full-length AvrSen1 protein bound 96 potato proteins. Each identified protein was annotated using gene ontology analysis. Proteins that play a role in photosynthesis, plant defense and metabolism, as well as ribosomal proteins were the predominant protein groups identified for both fungal protein sequences (Figure 20). Other classes with fewer proteins included plant growth, chaperone, translation, ATP synthesis, redox signalling, and stress response proteins were detected in both samples.

Number of Proteins Identified in Potato Leaves with CPG103/CPG106
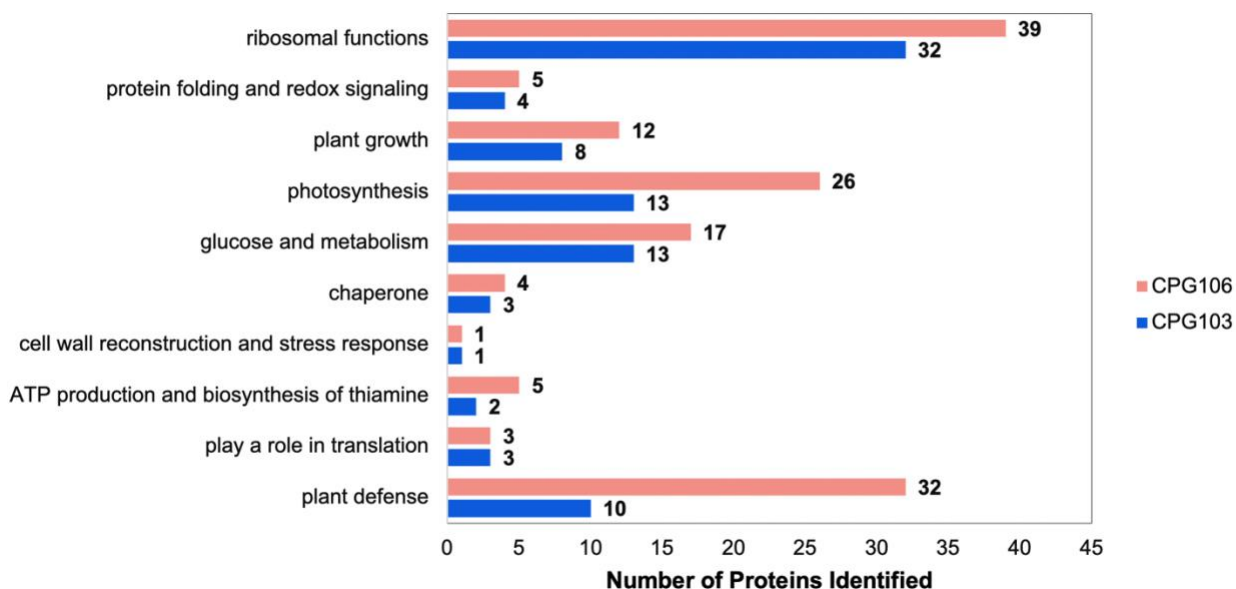
Figure 20. The distribution of the function of identified interaction partners of AvrSen1 in the Russet Kennebec leaf material. Following incubation of protein mixture extracted from Russet Kennebec potato plants with AvrSen1 protein, majority of identified proteins were involved in photosynthesis, metabolism, plant defense, or ribosomal processes. While the number of proteins identified for CPG103-containing proteins was lower, the overall number of proteins and their function-based distribution followed the similar pattern for both CPG103 and CPG106 sequences.

The distribution of interaction partners detected in tuber material was similar to leaf proteins with the main groups being ribosomal proteins, metabolism-related proteins, proteins involved in plant defense and photosynthesis (Figure 21). Majority of the proteins detected in CPG103 sample had multiple functions. These interaction partners either do not have published data providing information about their specific functions or have been previously reported to participate in various cellular processes. In comparison, CPG106 bound only four proteins, which were all involved in plant defense. The lack of interaction partners of CPG106 could be caused by the fact that the AvrSen1 proteins used for this experiment are bound to the MBP tag, which could affect the binding efficiency or activity levels of AvrSen1. However, the qualitative analysis of the proteins that were collected in the flow-through fractions resulted in identification of large number of unbound proteins. This suggest high specificity of the AvrSen1 interaction with potato proteins.

Another reason for detection of low number of tuber proteins for CPG106 protein could be due to the low concentration of AvrSen1 in the sample. Despite it having the similar sequence to the native AvrSen1, the conditions used for this experiment could differ significantly from that of infection. First of all, the poor expression of CPG106 in *E. coli* might have resulted in production of protein in low concentrations, resulting in lower binding efficiency of the protein. Additionally, the buffer used for extraction of proteins from plant tissue contained proteases which could negatively impact the AvrSen1 activity.

Number of Proteins Identified in Potato Tubers with CPG103/CPG106

Function	CPG106	CPG103
unknown functions	0	7
RNA metabolism & processing	0	6
ribosomal functions	0	48
plant browning	0	1
plant growth	0	18
nucleocytoplasmic transport	0	4
metabolism	0	48
membrane targeting and transport	0	6
chaperones	0	10
ATP production and energy transport	0	4
protein biosynthesis or degradation	0	16
plant development & stress response	0	12
cell detoxification	0	12
multiple functions	0	39
stress response	0	6
plant defense & photosynthesis	4	5

Number of Proteins Identified

Figure 21. The distribution of the function of identified interaction partners of AvrSen1 in the tuberous tissue of Russet Kennebec plants. Based on the observed mass spectra, the samples containing CPG106 protein, which is a full-length AvrSen1, did not interact with large number of proteins extracted from the tuber tissue of Russet Kennebec plants. In comparison, the half-sized fungal protein bound tuber proteins that varied in their function.

Overall, the data collected in this experiment resulted in the identification of potential AvrSen1 interaction partners that can play a role in infection. Interestingly, CPG103 had a higher affinity for proteins isolated from tuber material, while CPG106 primarily bound leaf proteins. Based on the number of identified proteins and their distribution based on function annotation, both CPG103

and CPG106 bind the proteins of similar functions when mixed with protein mixture extracted from leaves. However, this trend was not observed with tuber proteins.

The most common functional categories for AvrSen1 interactors were ribosomal, plant defence and metabolic proteins. Considering the biotrophic nature of *S. endobioticum*, the binding of ribosomal proteins and proteins involved in metabolism is not surprising. Similarly, the identification of plant defense protein was also expected as potato wart majorly impacts the growth and development of *S. tuberosum* plants. These data correlate with the protein classes observed when sampling PEI samples as well.

Out of 116 interaction partners from leaf material identified for both sequences, 30 proteins bound to both AvrSen1 sequences, which represents approximately 26% similarity. Comparatively, out of four CPG106 tuber interactors, two of the proteins also had affinity to CPG103, resulting in 50% resemblance. Altogether, the differences between the proteins identified for half-sized and full-length AvrSen1 may indicate that they have distinct functions. The specific proteins and the number of proteins in each group varied when comparing the list of interactors generated for CPG103 and CPG106 proteins.

4.5 Conclusions and Future Work

The preliminary data collected in this study provide a further insight into the function of AvrSen1 during potato wart infection. Majority of the proteins that were detected in the AvrSen1-containing samples were ribosomal proteins or proteins involved in plant defense, metabolism, or photosynthesis. These fundings are not surprising and correlate with the proteomics data collected for control plants and tubers infected with *S. endobioticum*.

While the data obtained in this experiment provides a better understanding regarding the type of proteins participating in immune response mechanisms induced by *S. endobioticum* infection, the acquired data suggests that AvrSen1 interacts with the higher number of leaf proteins in comparison to the proteins extracted from tuber material. This contradicts with the nature of the infection due to the fact that the wart malformations associated with potato wart almost exclusively appear on the tuber tissue alone. Therefore, understanding the affinity of binding between the leaf proteins and AvrSen1 might enhance our knowledge of metabolites and proteins involved in the

host-pathogen interactions during infection, as well as the effects of disease progression on the overall plant health.

The future work should focus on validating the identified AvrSen1 interactors. While numerous detected proteins have been previously associated with plant defense, their specific role in potato wart remains unknown. In addition to that, a lot of identified interaction partners were involved in metabolomic pathways and might lack experimental evidence to support their function in plant immunity. Therefore, the interaction efficiency of those proteins should be validated through transformation of their gene sequences into bacterial cells and their subsequent expression, extraction, and purification. Testing the isolated proteins instead of the protein lysate mixture might provide an addition support for binding affinity with AvrSen1.

Furthermore, the future experiments should concentrate on testing different potato varieties. By incubating AvrSen1 with proteins from resistant and susceptible varieties, some of the less essential interaction partners can be eliminated. This would also provide a better understanding regarding the different immune pathways that get activated following infection.

Finally, the use of maltose-binding protein as a negative control allowed elimination of false positive proteins that were detected in samples due to nonspecific binding; however, the future work should be focused on using purified AvrSen1 protein without an MBP-tag to replicate the endogenous structure of a protein during infection. The isolated AvrSen1 can be acquired using the transformation and expression method described in this study. Alternatively, it can also be extracted from the *S. endobioticum*-inoculated plants to simulate the endogenous levels of AvrSen1 that occur during potato wart infection. Similar to the suggestion mentioned above, testing the pure and isolated AvrSen1 will provide an additional confidence level and validate the obtained results.

Chapter 5 - Conclusions

The general goal of this thesis was to understand the effects of potato wart on plant health and immunity and to study the difference between control and infected plants. The work summarized here includes three main experiments that were conducted as a part of my master's project and includes proteomics study of collected samples, structure analysis of AvrSen1 and identification of interaction partners via co-immunoprecipitation of AvrSen1 with potato proteins.

Comparative proteomic study consisted of analysis of the provided control and infected tuber material. The infected tubers were sampled twice, but each sample was taken from a different area of the tuber to account for potential changes in disease progression. Over the two-year period, a total of 65 samples of four different potato varieties were collected and processed. Looking at the acquired data, it was observed that Russet Burbank and Targhee Russet use different immune pathways in response to disease progression. Following potato wart, majority of protein groups, including metabolite interconversion enzymes and proteins related to plant growth and development, are overexpressed in the infected Russet Burbank group in comparison to control samples. In contrast, the infected Targhee Russet tubers do not contain many proteins. This could indicate that this variety may undergo necrosis as disease progresses. Similarly, Caribou Russet and Mountain Gem Russet varieties either had no protein groups detected in the infected tuber material or the proteins were significantly downregulated. Based on the data published previously, these results may suggest that Targhee Russet, Mountain Gem Russet and Caribou Russet undergo hypersensitive immune response once the antigens produced by *S. endobioticum* are detected by plant cells. Therefore, all three of these varieties are considered resistant to this specific *S. endobioticum* pathotype. In contrast, Russet Burbank shows susceptibility to infection. Overexpression of metabolism-related proteins could indicate that fungus is potentially interfering with the host metabolic processes to utilize nutrients synthesized by the plant. Furthermore, the downregulation of proteins involved in plant defense may indicate that progression of infection results in the arrest of defense mechanisms of this variety.

In addition to observed differences in resistance between varieties, the proteomic analysis of non-warted material of infected tubers shows no significant variation on protein level between control tissue and the material collected from visually unaffected area of infected sample. This was observed in both Russet Burbank and Targhee Russet varieties throughout both runs of the study.

Such observation confirms biotrophic nature of *S. endobioticum* which is associated with heavy dependence on its host for nutrient acquisition, as well as low virulence of this pathogen and slow progression of infection.

Overall, the proteomic nLC-MS/MS analysis conducted provides a better insight regarding the mechanisms of potato wart infection and its effects on plant growth, immune response and metabolic processes. However, some limitations of the obtained results should be addressed in the future. First of all, the future work should include PCR analysis of collected infected material to determine the specific pathotype of *S. endobioticum* responsible for infection in each variety. Since *S. endobioticum* does not grow under laboratory conditions and all infected material was provided from farm fields in PEI, it is possible that the tubers analyzed in this experiment could be contaminated with different pathotypes of potato wart fungus. In addition to that, it is hard to predict whether the infection progressed to the same stage within both the four tested varieties and the samples within each infected group. Furthermore, the resistance of the varieties being analyzed should be tested using one of the commonly used methods, such as the Glynne-Lemmerzahl method. This would allow assessment of the accuracy of resistance prediction acquired by proteomic analysis. Finally, the future work should concentrate on analysis of metabolites and lipids of the same sets of samples. While proteomic data collected in this study provided a lot of useful information that has not been documented before, conducting a metabolomic study would allow verification of the acquired data, while providing a further insight into specific pathways that are affected by the onset of infection.

The expression and structure analysis of AvrSen1 was partially successful. While the methodology for extraction and isolation of MBP-AvrSen1 fusion protein from the rest of bacterial proteins was developed successfully, the removal of MBP tag has shown to be challenging. Since AvrSen1 remains to be the only effector molecule identified in *S. endobioticum*, the future work should focus on trying different transformation systems. Alternatively, due to the numerous advantages of *E. coli* cells, it is also an option to try different tags and compare their efficiency to the polyhistidine and MBP tags. Following successful purification of AvrSen1 and analysis of its structure using X-ray crystallography approach, its function should be assessed using activity assays.

The final study of this master's project included the analysis of AvrSen1 interaction partners in Russet Kennebec variety. The obtained results suggest that AvrSen1 behaves differently in tubers and leaves as different proteins were identified for those samples. Additionally, it was determined that the two AvrSen1 proteins of different lengths also interact differently with potato proteins. While there were similarities between the number of protein interactors identified in leaf mixture for both half-sized protein and the full-length AvrSen1, the obtained results vary significantly for the proteins isolated from tuber tissue.

Due to the uncertainty regarding the resistance status of Russet Kennebec variety to potato wart, future work should also concentrate on repeating this experiment on a resistant type of potatoes. This will allow determination of the difference between the protein groups that AvrSen1 interacts with depending on which type of immune response gets activated in the host plant. It will also help to identify false positive interactions that might have been detected in this experiment. In addition, studying the difference between AvrSen1 interactors in resistant and susceptible tubers will allow determination of whether this effector molecule is more active in extracellular or intracellular matrix. Furthermore, the gene-for-gene concept described earlier in this work suggests that there is a protein that gets synthesized by the plant in response to AvrSen1. This protein is referred to as Sen1; however, its location in the chromosome has only been predicted so far and it has not been previously characterized. Performing co-immunoprecipitation experiment on the resistant variety might help to aid in its identification. This could be of high significance as the specificity of AvrSen1-Sen1 interaction might explain the limited host variability of *S. endobioticum*.

References

(1) Hyde, K. D. The numbers of fungi. *Fungal Diversity* **2022**, *114* (1), 1-1. DOI: 10.1007/s13225-022-00507-y.

(2) Phukhamsakda, C.; Nilsson, R. H.; Bhunjun, C. S.; de Farias, A. R. G.; Sun, Y.-R.; Wijesinghe, S. N.; Raza, M.; Bao, D.-F.; Lu, L.; Tibpromma, S.; et al. The numbers of fungi: contributions from traditional taxonomic studies and challenges of metabarcoding. *Fungal Diversity* **2022**, *114* (1), 327-386. DOI: 10.1007/s13225-022-00502-3.

(3) Selitrennikoff, C. P. Antifungal Proteins. *Applied and Environmental Microbiology* **2001**, *67* (7), 2883-2894. DOI: 10.1128/AEM.67.7.2883-2894.2001.

(4) Ferreira, R. B.; Monteiro, S.; Freitas, R.; Santos, C. N.; Chen, Z.; Batista, L. M.; Duarte, J.; Borges, A.; Teixeira, A. R. Fungal Pathogens: The Battle for Plant Infection. *Critical Reviews in Plant Sciences* **2006**, *25* (6), 505-524. DOI: 10.1080/07352680601054610.

(5) Montesinos, E.; Bonaterra, A.; Badosa, E.; Francés, J.; Alemany, J.; Llorente, I.; Moragrega, C. Plant-microbe interactions and the new biotechnological methods of plant disease control. *International Microbiology* **2002**, *5* (4), 169-175. DOI: 10.1007/s10123-002-0085-9.

(6) Koeck, M.; Hardham, A. R.; Dodds, P. N. The role of effectors of biotrophic and hemibiotrophic fungi in infection. *Cellular Microbiology* **2011**, *13* (12), 1849-1857. DOI: 10.1111/j.1462-5822.2011.01665.x.

(7) Ray, M.; Ray, A.; Dash, S.; Mishra, A.; Achary, K. G.; Nayak, S.; Singh, S. Fungal disease detection in plants: Traditional assays, novel diagnostic techniques and biosensors. *Biosensors and Bioelectronics* **2017**, *87*, 708-723. DOI: 10.1016/j.bios.2016.09.032.

(8) Mendgen, K.; Hahn, M. Plant infection and the establishment of fungal biotrophy. *Trends in Plant Science* **2002**, *7* (8), 352-356. DOI: 10.1016/S1360-1385(02)02297-5.

(9) Saur, I. M. L.; Hückelhoven, R. Recognition and defence of plant-infecting fungal pathogens. *Journal of Plant Physiology* **2021**, *256*, 153324. DOI: 10.1016/j.jplph.2020.153324.

(10) Mapuranga, J.; Zhang, N.; Zhang, L.; Chang, J.; Yang, W. Infection Strategies and Pathogenicity of Biotrophic Plant Fungal Pathogens. *Frontiers in Microbiology* **2022**, *13*, Review. DOI: 10.3389/fmicb.2022.799396.

(11) Masachis, S.; Segorbe, D.; Turrà, D.; Leon-Ruiz, M.; Fürst, U.; El Ghalid, M.; Leonard, G.; López-Berges, M. S.; Richards, T. A.; Felix, G.; Di Pietro, A. A fungal pathogen secretes plant alkalinizing peptides to increase infection. *Nature Microbiology* **2016**, *1* (6), 16043. DOI: 10.1038/nmicrobiol.2016.43.

(12) Perfect, S. E.; Green, J. R. Infection structures of biotrophic and hemibiotrophic fungal plant pathogens. *Molecular Plant Pathology* **2001**, *2* (2), 101-108. DOI: 10.1046/j.1364-3703.2001.00055.x.

(13) Panstruga, R.; Schulze-Lefert, P. Corruption of host seven-transmembrane proteins by pathogenic microbes: a common theme in animals and plants? *Microbes and Infection* **2003**, *5* (5), 429-437. DOI: 10.1016/S1286-4579(03)00053-4.

(14) Staples, R. C. Nutrients for a rust fungus: the role of haustoria. *Trends in Plant Science* **2001**, *6* (11), 496-498. DOI: 10.1016/S1360-1385(01)02126-4.

(15) Schulze-Lefert, P. Knocking on the heaven's wall: pathogenesis of and resistance to biotrophic fungi at the cell wall. *Current Opinion in Plant Biology* **2004**, *7* (4), 377-383. DOI: 10.1016/j.pbi.2004.05.004.

(16) Mendgen, K.; Schneider, A.; Sterk, M.; Fink, W. The Differentiation of Infection Structures as a Result of Recognition Events between some Biotrophic Parasites and their Hosts. *Journal of Phytopathology* **1988**, *123* (3), 259-272. DOI: 10.1111/j.1439-0434.1988.tb04476.x.

(17) Smith, F.; Smith, S. Membrane Transport at the Biotrophic Interface: an Overview. *Functional Plant Biology* **1989**, *16* (1), 33-43. DOI: 10.1071/PP9890033.

(18) SMITH, S. E.; SMITH, F. A. Structure and function of the interfaces in biotrophic symbioses as they relate to nutrient transport. *New Phytologist* **1990**, *114* (1), 1-38. DOI: 10.1111/j.1469-8137.1990.tb00370.x.

(19) Heath, M. C. Signal exchange between higher plants and rust fungi. *Canadian Journal of Botany* **1995**, *73* (S1), 616-623. DOI: 10.1139/b95-303.

(20) Spencer-Phillips, P. T. N. Function of Fungal Haustoria in Epiphytic and Endophytic Infections. In *Advances in Botanical Research*, Andrews, J. H., Tommerup, I. C., Callow, J. A. Eds.; Vol. 24; Academic Press, 1997; pp 309-333.

(21) Mengaud, J.; Ohayon, H.; Gounon, P.; Mege, R. M.; Cossart, P. E-cadherin is the receptor for internalin, a surface protein required for entry of L. monocytogenes into epithelial cells. *Cell* **1996**, *84* (6), 923-932. DOI: 10.1016/s0092-8674(00)81070-3 From NLM.

(22) Pascholati, S. F.; Yoshioka, H.; Kunoh, H.; Nicholson, R. L. Preparation of the infection court by Erysiphe graminis f. sp. hordei: cutinase is a component of the conidial exudate. *Physiological and Molecular Plant Pathology* **1992**, *41* (1), 53-59. DOI: 10.1016/0885-5765(92)90048-Z.

(23) Deising, H.; Nicholson, R. L.; Haug, M.; Howard, R. J.; Mendgen, K. Adhesion Pad Formation and the Involvement of Cutinase and Esterases in the Attachment of Uredospores to the Host Cuticle. *Plant Cell* **1992**, *4* (9), 1101-1111. DOI: 10.1105/tpc.4.9.1101 From NLM.

(24) Mendgen, K.; Hahn, M.; Deising, H. MORPHOGENESIS AND MECHANISMS OF PENETRATION BY PLANT PATHOGENIC FUNGI. *Annual Review of Phytopathology* **1996**, *34* (Volume 34, 1996), 367-386. DOI: 10.1146/annurev.phyto.34.1.367.

(25) Punja, Z. K. Genetic engineering of plants to enhance resistance to fungal pathogens—a review of progress and future prospects. *Canadian Journal of Plant Pathology* **2001**, *23* (3), 216-235. DOI: 10.1080/07060660109506935.

(26) Howard, R. J.; Ferrari, M. A.; Roach, D. H.; Money, N. P. Penetration of hard substrates by a fungus employing enormous turgor pressures. *Proceedings of the National Academy of Sciences* **1991**, *88* (24), 11281-11284. DOI: 10.1073/pnas.88.24.11281.

(27) FERREIRA, R. B.; MONTEIRO, S.; FREITAS, R.; SANTOS, C. N.; CHEN, Z.; BATISTA, L. M.; DUARTE, J.; BORGES, A.; TEIXEIRA, A. R. The role of plant defence proteins in fungal pathogenesis. *Molecular Plant Pathology* **2007**, *8* (5), 677-700. DOI: 10.1111/j.1364-3703.2007.00419.x.

(28) Vaghela, B.; Vashi, R.; Rajput, K.; Joshi, R. Plant chitinases and their role in plant defense: A comprehensive review. *Enzyme and Microbial Technology* **2022**, *159*, 110055. DOI: 10.1016/j.enzmictec.2022.110055.

(29) Chi, M.; Xiang, Y.; Liu, J.-J. Chapter 4 - Plant thaumatin-like protein family: Genome-wide diversification, evolution, and functional adaptation. In *Defense-Related Proteins in Plants*, Upadhyay, S. K. Ed.; Academic Press, 2024; pp 99-142.

(30) Pawar, M.; Dixit, S.; Vadassery, J. Chapter 7 - Plant protease inhibitors: Biological roles and applications in plant stress tolerance. In *Defense-Related Proteins in Plants*, Upadhyay, S. K. Ed.; Academic Press, 2024; pp 205-222.

(31) van de Vossenberg, B.; Prodhomme, C.; Vossen, J. H.; van der Lee, T. A. J. Synchytrium endobioticum, the potato wart disease pathogen. *Mol Plant Pathol* **2022**, *23* (4), 461-474. DOI: 10.1111/mpp.13183 From NLM.

(32) Schilberszky, K. Ein neuer Schorfparasit der Kartoffelknollen. *Ber Deut Botan Ges* **1896**, *14*, 36-37.

(33) Agency, C. F. I. *Potato wart in PEI*. 2021. https://www.canada.ca/en/food-inspection-agency/news/2021/11/potato-wart-in-pei.html.

(34) Russell, N. *A year later, P.E.I. potato growers still feeling impacts of U.S. border closure*. 2022. https://www.cbc.ca/news/canada/prince-edward-island/pei-potato-wart-border-closure-one-year-1.6656853.

(35) Sparrow, F. K. *Aguatic phycomycetes*; 1960.

(36) Barr, D. J. S. Chytridiomycota. In *Systematics and Evolution: Part A*, McLaughlin, D. J., McLaughlin, E. G., Lemke, P. A. Eds.; Springer Berlin Heidelberg, 2001; pp 93-112.

(37) Heckman, D. S.; Geiser, D. M.; Eidell, B. R.; Stauffer, R. L.; Kardos, N. L.; Hedges, S. B. Molecular Evidence for the Early Colonization of Land by Fungi and Plants. *Science* **2001**, *293* (5532), 1129-1133. DOI: 10.1126/science.1061457.

(38) Franc, G. D. Potato wart. **2007**. DOI: 10.1094/APSnetFeature-2007-0607.

(39) Lange, L.; Olson, L. W. Development of the resting sporangia ofSynchytrium endobioticum, the causal agent of potato wart disease. *Protoplasma* **1981**, *106* (1), 83-95. DOI: 10.1007/BF02115963.

(40) Lange, L.; Olson, L. W. Development of the zoosporangia ofSynchytrium endobioticum, the causal agent of potato wart disease. *Protoplasma* **1981**, *106* (1), 97-108. DOI: 10.1007/BF02115964.

(41) Lange, L.; Olson, L. W. Germination and parasitation of the resting sporangia ofSynchytrium endobioticum. *Protoplasma* **1981**, *106* (1), 69-82. DOI: 10.1007/BF02115962.

(42) Weiss, F. THE CONDITIONS OF INFECTION IN POTATO WART. *American Journal of Botany* **1925**, *12* (7), 413-443. DOI: 10.1002/j.1537-2197.1925.tb05846.x.

(43) Curtis, K. M.; Blackman, V. H. IX.— The life-history and cytology of synchytrium endobioticum (schilb.), perc., the cause of wart disease in potato. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character* **1921**, *210* (372-381), 409-478. DOI: 10.1098/rstb.1921.0009.

(44) Snyder, L. R.; Kirkland, J. J.; Dolan, J. W. *Introduction to modern liquid chromatography*; John Wiley & Sons, 2011.

(45) Scott, R. P. W. Modern liquid chromatography. *Chemical Society Reviews* **1992**, *21* (2), 137-145, 10.1039/CS9922100137. DOI: 10.1039/CS9922100137.

(46) Regnier, F. E. High-performance liquid chromatography of proteins. In *Methods in Enzymology*, Vol. 91; Academic Press, 1983; pp 137-190.

(47) Nawrocki, J. The silanol group and its role in liquid chromatography. *Journal of Chromatography A* **1997**, *779* (1), 29-71. DOI: 10.1016/S0021-9673(97)00479-2.

(48) Shi, Y.; Xiang, R.; Horváth, C.; Wilkins, J. A. The role of liquid chromatography in proteomics. *Journal of Chromatography A* **2004**, *1053* (1), 27-36. DOI: 10.1016/j.chroma.2004.07.044.

(49) De Hoffmann, E.; Stroobant, V. *Mass spectrometry: principles and applications*; John Wiley & Sons, 2007.

(50) Newell, D. B.; Tiesinga, E. The international system of units (SI). *NIST Special Publication* **2019**, *330*, 1-138. DOI: 10.6028/NIST.SP.330-2019

(51) Yamashita, M.; Fenn, J. B. Electrospray ion source. Another variation on the free-jet theme. *The Journal of Physical Chemistry* **1984**, *88* (20), 4451-4459. DOI: 10.1021/j150664a002.

(52) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, *246* (4926), 64-71. DOI: 10.1126/science.2675315.

(53) Covey, T. R.; Thomson, B. A.; Schneider, B. B. Atmospheric pressure ion sources. *Mass Spectrometry Reviews* **2009**, *28* (6), 870-897. DOI: 10.1002/mas.20246.

(54) Kebarle, P.; Verkerk, U. H. Electrospray: From ions in solution to ions in the gas phase, what we know now. *Mass Spectrometry Reviews* **2009**, *28* (6), 898-917. DOI: 10.1002/mas.20247.

(55) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Analytical Chemistry* **2013**, *85* (1), 2-9. DOI: 10.1021/ac302789c.

(56) Smith, J. N.; Flagan, R. C.; Beauchamp, J. L. Droplet Evaporation and Discharge Dynamics in Electrospray Ionization. *The Journal of Physical Chemistry A* **2002**, *106* (42), 9957-9967. DOI: 10.1021/jp025723e.

(57) Awad, H.; Khamis, M. M.; El-Aneed, A. Mass Spectrometry, Review of the Basics: Ionization. *Applied Spectroscopy Reviews* **2015**, *50* (2), 158-175. DOI: 10.1080/05704928.2014.954046.

(58) Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. Molecular beams of macroions. *The Journal of chemical physics* **1968**, *49* (5), 2240-2249. DOI: 10.1063/1.1670391.

(59) Iribarne, J.; Thomson, B. On the evaporation of small ions from charged droplets. *The Journal of chemical physics* **1976**, *64* (6), 2287-2294. DOI: 10.1063/1.432536.

(60) Karas, M.; Bahr, U.; Dülcks, T. Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. *Fresenius' Journal of Analytical Chemistry* **2000**, *366* (6), 669-676. DOI: 10.1007/s002160051561.

(61) Wilm, M.; Mann, M. Analytical Properties of the Nanoelectrospray Ion Source. *Analytical Chemistry* **1996**, *68* (1), 1-8. DOI: 10.1021/ac9509519.

(62) Miller, P. E.; Denton, M. B. The quadrupole mass filter: basic operating concepts. *Journal of chemical education* **1986**, *63* (7), 617. DOI: 10.1021/ed063p617.

(63) Campana, J. E. Elementary theory of the quadrupole mass filter. *International Journal of Mass Spectrometry and Ion Physics* **1980**, *33* (2), 101-117. DOI: 10.1016/0020-7381(80)80042-8.

(64) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry* **2005**, *40* (4), 430-443. DOI: 10.1002/jms.856.

(65) Scientific, T. F. *Q Exactive Orbitrap Mass Spectrometers (MS)*. 2016. https://www.thermofisher.com/ca/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-systems/orbitrap-lc-ms/q-exactive-orbitrap-mass-spectrometers.html.

(66) Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* **2005**, *4* (12), 2010-2021. DOI: 10.1074/mcp.T500030-MCP200 From NLM.

(67) Nolting, D.; Malek, R.; Makarov, A. Ion traps in modern mass spectrometry. *Mass Spectrometry Reviews* **2019**, *38* (2), 150-168. DOI: 10.1002/mas.21549.

(68) Eliuk, S.; Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annual Review of Analytical Chemistry* **2015**, *8* (Volume 8, 2015), 61-80. DOI: 10.1146/annurev-anchem-071114-040325.

(69) Scigelova, M.; Makarov, A. Orbitrap Mass Analyzer – Overview and Applications in Proteomics. *PROTEOMICS* **2006**, *6* (S2), 16-21. DOI: 10.1002/pmic.200600528.

(70) Diedrich, J. K.; Pinto, A. F. M.; Yates, J. R., III. Energy Dependence of HCD on Peptide Fragmentation: Stepped Collisional Energy Finds the Sweet Spot. *Journal of the American Society for Mass Spectrometry* **2013**, *24* (11), 1690-1699. DOI: 10.1007/s13361-013-0709-7.

(71) Dupree, E. J.; Jayathirtha, M.; Yorkey, H.; Mihasan, M.; Petre, B. A.; Darie, C. C. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field. *Proteomes* **2020**, *8* (3), 14. DOI: 10.3390/proteomes8030014.

(72) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R., III. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical Reviews* **2013**, *113* (4), 2343-2394. DOI: 10.1021/cr3003533.

(73) Miller, R. M.; Smith, L. M. Overview and considerations in bottom-up proteomics. *Analyst* **2023**, *148* (3), 475-486. DOI: 10.1039/d2an01246d From NLM.

(74) Smith, L. M.; Kelleher, N. L.; Linial, M.; Goodlett, D.; Langridge-Smith, P.; Ah Goo, Y.; Safford, G.; Bonilla*, L.; Kruppa, G.; Zubarev, R.; et al. Proteoform: a single term describing protein complexity. *Nature Methods* **2013**, *10* (3), 186-187. DOI: 10.1038/nmeth.2369.

(75) Cupp-Sutton, K. A.; Wu, S. High-throughput quantitative top-down proteomics. *Molecular Omics* **2020**, *16* (2), 91-99, 10.1039/C9MO00154A. DOI: 10.1039/C9MO00154A.

(76) Wysocki, V. H.; Resing, K. A.; Zhang, Q.; Cheng, G. Mass spectrometry of peptides and proteins. *Methods* **2005**, *35* (3), 211-222. DOI: 10.1016/j.ymeth.2004.08.013.

(77) Roepstorff, P.; Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical Mass Spectrometry* **1984**, *11* (11), 601-601. DOI: 10.1002/bms.1200111109.

(78) Marcotte, E. M. How do shotgun proteomics algorithms identify proteins? *Nature Biotechnology* **2007**, *25* (7), 755-757. DOI: 10.1038/nbt0707-755.

(79) Inc., T. F. S. *Protein Sample Preparation for Mass Spectrometry*. n.d. https://www.thermofisher.com/ca/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/sample-preparation-mass-spectrometry.html.

(80) Tsiatsiani, L.; Heck, A. J. R. Proteomics beyond trypsin. *The FEBS Journal* **2015**, *282* (14), 2612-2626. DOI: 10.1111/febs.13287.

(81) Swaney, D. L.; Wenger, C. D.; Coon, J. J. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res* **2010**, *9* (3), 1323-1329. DOI: 10.1021/pr900863u From NLM.

(82) Vandermarliere, E.; Mueller, M.; Martens, L. Getting intimate with trypsin, the leading protease in proteomics. *Mass Spectrometry Reviews* **2013**, *32* (6), 453-465. DOI: 10.1002/mas.21376.

(83) Krasny, L.; Huang, P. H. Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology. *Molecular omics* **2021**, *17* (1), 29-42. DOI: 10.1039/d0mo00072h.

(84) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & Cellular Proteomics* **2012**, *11* (6), O111.016717. DOI: 10.1074/mcp.O111.016717 From NLM.

(85) Barkovits, K.; Pacharra, S.; Pfeiffer, K.; Steinbach, S.; Eisenacher, M.; Marcus, K.; Uszkoreit, J. Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Library-based Data-independent Acquisition. *Molecular & Cellular Proteomics* **2020**, *19* (1), 181-197. DOI: 10.1074/mcp.RA119.001714 From NLM.

(86) Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B. C.; Aebersold, R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Molecular Systems Biology* **2018**, *14* (8), e8126. DOI: 10.15252/msb.20178126.

(87) Rosano, G. L.; Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology* **2014**, *5*, Review. DOI: 10.3389/fmicb.2014.00172.

(88) Johnston, C.; Martin, B.; Fichant, G.; Polard, P.; Claverys, J.-P. Bacterial transformation: distribution, shared mechanisms and divergent control. *Nature Reviews Microbiology* **2014**, *12* (3), 181-196. DOI: 10.1038/nrmicro3199.

(89) Mercenier, A.; Chassy, B. M. Strategies for the development of bacterial transformation systems. *Biochimie* **1988**, *70* (4), 503-517. DOI: 10.1016/0300-9084(88)90086-7.

(90) Francis, D. M.; Page, R. Strategies to Optimize Protein Expression in E. coli. *Current Protocols in Protein Science* **2010**, *61* (1), 5.24.21-25.24.29. DOI: 10.1002/0471140864.ps0524s61.

(91) Kapust, R. B.; Tözsér, J.; Copeland, T. D.; Waugh, D. S. The P1′ specificity of tobacco etch virus protease. *Biochemical and Biophysical Research Communications* **2002**, *294* (5), 949-955. DOI: 10.1016/S0006-291X(02)00574-0.

(92) di Guana, C.; Lib, P.; Riggsa, P. D.; Inouyeb, H. Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. *Gene* **1988**, *67* (1), 21-30. DOI: 10.1016/0378-1119(88)90004-2.

(93) Young, C. L.; Britton, Z. T.; Robinson, A. S. Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. *Biotechnology Journal* **2012**, *7* (5), 620-634. DOI: 10.1002/biot.201100155.

(94) Kapust, R. B.; Waugh, D. S. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci* **1999**, *8* (8), 1668-1674. DOI: 10.1110/ps.8.8.1668 From NLM.

(95) Crowe, J.; Dobeli, H.; Gentz, R.; Hochuli, E.; Stiiber, D.; Henco, K. 6xffis-Ni-NTA Chromatography as a Superior Technique in Recombinant Protein Expression/Purification. In *Protocols for Gene Analysis*, Harwood, A. J. Ed.; Humana Press, 1994; pp 371-387.

(96) (NEB), N. E. B. *TEV Protease*. n.d. https://www.neb.com/en/products/p8112-tev-protease.

(97) Bolanos-Garcia, V. M.; Davies, O. R. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2006**, *1760* (9), 1304-1313. DOI: 10.1016/j.bbagen.2006.03.027.

(98) Laboratories, B.-R. *Nickel Columns for Chromatography*. n.d. https://www.bio-rad.com/en-ca/feature/nickel-columns-nickel-resin.html.

(99) Cytiva. *Affinity Chromatography*. 2021. https://cdn.cytivalifesciences.com/api/public/content/digi-11495-pdf.

(100) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583-589. DOI: 10.1038/s41586-021-03819-2.

(101) Varadi, M.; Bertoni, D.; Magana, P.; Paramval, U.; Pidruchna, I.; Radhakrishnan, M.; Tsenkov, M.; Nair, S.; Mirdita, M.; Yeo, J.; et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research* **2023**, *52* (D1), D368-D375. DOI: 10.1093/nar/gkad1011.

(102) Fekete, S.; Beck, A.; Veuthey, J.-L.; Guillarme, D. Theory and practice of size exclusion chromatography for the analysis of protein aggregates. *Journal of Pharmaceutical and Biomedical Analysis* **2014**, *101*, 161-173. DOI: 10.1016/j.jpba.2014.04.011.

(103) Brusotti, G.; Calleri, E.; Colombo, R.; Massolini, G.; Rinaldi, F.; Temporini, C. Advances on Size Exclusion Chromatography and Applications on the Analysis of Protein Biopharmaceuticals and Protein Aggregates: A Mini Review. *Chromatographia* **2018**, *81* (1), 3-23. DOI: 10.1007/s10337-017-3380-5.

(104) Uliyanchenko, E. Size-exclusion chromatography—from high-performance to ultra-performance. *Analytical and Bioanalytical Chemistry* **2014**, *406* (25), 6087-6094. DOI: 10.1007/s00216-014-8041-z.

(105) Hong, P.; Koza, S.; Bouvier, E. S. Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *J Liq Chromatogr Relat Technol* **2012**, *35* (20), 2923-2950. DOI: 10.1080/10826076.2012.743724 From NLM.

(106) Laboratories, B.-R. *NGC Chromatography Systems* n.d. https://www.bio-rad.com/sites/default/files/webroot/web/pdf/lsr/literature/Bulletin_6326.pdf/

(107) Balotf, S.; Wilson, R.; Tegg, R. S.; Nichols, D. S.; Wilson, C. R. Shotgun Proteomics as a Powerful Tool for the Study of the Proteomes of Plants, Their Pathogens, and Plant-Pathogen Interactions. *Proteomes* **2022**, *10* (1). DOI: 10.3390/proteomes10010005 From NLM.

(108) Manikandan, R.; Harish, S.; Karthikeyan, G.; Raguchander, T. Comparative Proteomic Analysis of Different Isolates of Fusarium oxysporum f.sp. lycopersici to Exploit the Differentially Expressed Proteins Responsible for Virulence on Tomato Plants. *Frontiers in Microbiology* **2018**, *9*, Original Research. DOI: 10.3389/fmicb.2018.00420.

(109) Murawska, Z.; Dębski, J.; Szajko, K.; Lebecka, R. Isolation of proteins from potato tubers. *Plant Breeding and Seed Science* **2017**, *75*, 23-27. DOI: 10.1515/plass-2017-0005.

(110) Pradhan, A.; Ghosh, S.; Sahoo, D.; Jha, G. Fungal effectors, the double edge sword of phytopathogens. *Current Genetics* **2021**, *67* (1), 27-40. DOI: 10.1007/s00294-020-01118-3.

(111) Rouxel, T.; Balesdent, M.-H. Avirulence Genes. In *Encyclopedia of Life Sciences*, 2010.

(112) Figueroa, M.; Ortiz, D.; Henningsen, E. C. Tactics of host manipulation by intracellular effectors from plant pathogenic fungi. *Current Opinion in Plant Biology* **2021**, *62*, 102054. DOI: 10.1016/j.pbi.2021.102054.

(113) Rocafort, M.; Fudal, I.; Mesarich, C. H. Apoplastic effector proteins of plant-associated fungi and oomycetes. *Current Opinion in Plant Biology* **2020**, *56*, 9-19. DOI: 10.1016/j.pbi.2020.02.004.

(114) Kausar, H.; Sharma, G.; Bhatt, B. Role of fungal effector proteins for disease expression in plants. **2023**.

(115) Tariqjaveed, M.; Mateen, A.; Wang, S.; Qiu, S.; Zheng, X.; Zhang, J.; Bhadauria, V.; Sun, W. Versatile effectors of phytopathogenic fungi target host immunity. *Journal of Integrative Plant Biology* **2021**, *63* (11), 1856-1873. DOI: 10.1111/jipb.13162.

(116) Luderer, R.; Joosten, M. H. A. J. Avirulence proteins of plant pathogens: determinants of victory and defeat. *Molecular Plant Pathology* **2001**, *2* (6), 355-364. DOI: 10.1046/j.1464-6722.2001.00086.x.

(117) Laugé, R.; De Wit, P. J. Fungal avirulence genes: structure and possible functions. *Fungal Genet Biol* **1998**, *24* (3), 285-297. DOI: 10.1006/fgbi.1998.1076 From NLM.

(118) van de Vossenberg, B.; Prodhomme, C.; van Arkel, G.; van Gent-Pelzer, M. P. E.; Bergervoet, M.; Brankovics, B.; Przetakiewicz, J.; Visser, R. G. F.; van der Lee, T. A. J.; Vossen, J. H. The Synchytrium endobioticum AvrSen1 Triggers a Hypersensitive Response in Sen1

Potatoes While Natural Variants Evade Detection. *Mol Plant Microbe Interact* **2019**, *32* (11), 1536-1546. DOI: 10.1094/mpmi-05-19-0138-r From NLM.

(119) van de Vossenberg, B. T. L. H.; Prodhomme, C.; Vossen, J. H.; van der Lee, T. A. J. Synchytrium endobioticum, the potato wart disease pathogen. *Molecular Plant Pathology* **2022**, *23* (4), 461-474. DOI: 10.1111/mpp.13183.

(120) Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **2015**, *10* (6), 845-858. DOI: 10.1038/nprot.2015.053.

(121) Singh, R.; Liyanage, R.; Gupta, C.; Lay, J. O.; Pereira, A.; Rojas, C. M. The Arabidopsis Proteins AtNHR2A and AtNHR2B Are Multi-Functional Proteins Integrating Plant Immunity With Other Biological Processes. *Frontiers in Plant Science* **2020**, *11*, Original Research. DOI: 10.3389/fpls.2020.00232.

(122) Dunham, W. H.; Mullin, M.; Gingras, A.-C. Affinity-purification coupled to mass spectrometry: Basic principles and strategies. *PROTEOMICS* **2012**, *12* (10), 1576-1590. DOI: 10.1002/pmic.201100523.

(123) Lampugnani, E. R.; Wink, R. H.; Persson, S.; Somssich, M. The Toolbox to Study Protein–Protein Interactions in Plants. *Critical Reviews in Plant Sciences* **2018**, *37* (4), 308-334. DOI: 10.1080/07352689.2018.1500136.

(124) Biolabs, N. E. *Isolation of MBP-fusion protein using Anti-MBP Magnetic Beads*. 2024. https://www.neb.com/en/protocols/0001/01/01/isloation-of-mbp-fusion-protein-using-anti-mbp-magnetic-beads#.

(125) Bergen, P. U. a. U. o. On beads digestion (immunoprecipitate). n.d.

(126) Hernandez, P.; Binz, P.-A.; Wilkins, M. R. Protein Identification in Proteomics. In *Proteome Research: Concepts, Technology and Application*, Wilkins, M. R., Appel, R. D., Williams, K. L., Hochstrasser, D. F. Eds.; Springer Berlin Heidelberg, 2007; pp 41-67.

(127) Lao, Y. W.; Gungormusler-Yilmaz, M.; Shuvo, S.; Verbeke, T.; Spicer, V.; Krokhin, O. V. Chromatographic behavior of peptides containing oxidized methionine residues in proteomic LC–MS experiments: Complex tale of a simple modification. *Journal of Proteomics* **2015**, *125*, 131-139. DOI: 10.1016/j.jprot.2015.05.018.

Appendices

Table S1. The list of all the interaction partners identified from tuber tissue material for both CPG103 and CPG106.

The list includes the name of the proteins, Uniprot IDs, information regarding presence or absence of the protein in the sample, as well as description of the protein function. The following list excludes the proteins that were detected as interaction partners in MBP sample.

Protein list	CPG103+tubers	CPG106+tubers	Description
Granule-bound starch synthase 1, chloroplastic/amyloplastic (SSG1_SOLTU)	Present	Not detected	amylose synthesis
ATP synthase subunit gamma, mitochondrial (M1CAI4_SOLTU)	Present	Not detected	ATP production
ATP synthase subunit beta (M1D096_SOLTU)	Present	Not detected	ATP production
Probable UDP-arabinopyranose mutase 2 (RGP2_SOLTU)	Present	Not detected	biosynthesis of cell wall components
S-formylglutathione hydrolase (M1AYK4_SOLTU)	Present	Not detected	cell detoxification
NADPH quinone oxidoreductase (M1C1K7_SOLTU)	Present	Not detected	cell detoxification
V-type proton ATPase catalytic subunit A (M1CII0_SOLTU)	Present	Not detected	cell pH homeostasis and energy transport
Small molecular heat shock protein (M0ZT86_SOLTU)	Present	Not detected	chaperone
Low molecular weight heat-shock protein (M1DJ69_SOLTU)	Present	Not detected	chaperone
Heat shock cognate protein 80 (M1C075_SOLTU)	Present	Not detected	chaperone
Heat shock cognate 70 kDa protein (M0ZI18_SOLTU)	Present	Not detected	chaperone
DnaJ (Q38HT9_SOLTU)	Present	Not detected	chaperone
17.6 kDa class I heat shock protein (M1CZ82_SOLTU)	Present	Not detected	chaperone
101 kDa heat shock protein (M1CAJ2_SOLTU)	Present	Not detected	chaperone
Beta-mannosidase (M1A467_SOLTU)	Present	Not detected	degradation of cell wall
Glyceraldehyde-3-phosphate dehydrogenase (M1ASG7_SOLTU)	Present	Not detected	glycolysis
Fibrillarin homolog (M1AMW1_SOLTU)	Present	Not detected	interacts with plant viruses
C2 domain-containing protein (M1AMV5_SOLTU)	Present	Not detected	membrane targeting
Major intrinsic protein 2 (M1AE81_SOLTU)	Present	Not detected	membrane transport
Guanosine nucleotide diphosphate dissociation inhibitor (M1B2L7_SOLTU)	Present	Not detected	membrane transport

Protein list	CPG103+tubers	CPG106+tubers	Description
ADP, ATP carrier protein, mitochondrial (ADT1_SOLTU)	Present	Not detected	membrane transport
UTP--glucose-1-phosphate uridylyltransferase (M1B150_SOLTU)	Present	Not detected	metabolism
Sucrose-phosphate synthase (M1CPB7_SOLTU)	Present	Not detected	metabolism
Sucrose synthase (SUS1_SOLTU)	Present	Not detected	metabolism
Sucrose synthase (M1A8J5_SOLTU)	Present	Not detected	metabolism
Succinic semialdehyde reductase isofom1 (M1BT61_SOLTU)	Present	Not detected	metabolism
Succinate--CoA ligase [ADP-forming] subunit beta, mitochondrial (M1BSG4_SOLTU)	Present	Not detected	metabolism
Short chain alcohol dehydrogenase (M1AFF5_SOLTU)	Present	Not detected	metabolism
S-methyl-5-thioribose kinase (M1CFQ4_SOLTU)	Present	Not detected	metabolism
S-(hydroxymethyl)glutathione dehydrogenase (M1ACT7_SOLTU)	Present	Not detected	metabolism
Pyruvate dehydrogenase E1 component subunit alpha, mitochondrial (ODPA_SOLTU)	Present	Not detected	metabolism
Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit beta (PFPB_SOLTU)	Present	Not detected	metabolism
Pyridoxine biosynthesis protein isoform A (M1CZB0_SOLTU)	Present	Not detected	metabolism
Phosphoenolpyruvate carboxylase (CAPP_SOLTU)	Present	Not detected	metabolism
Phospho-2-dehydro-3-deoxyheptonate aldolase 1, chloroplastic (AROF_SOLTU)	Present	Not detected	metabolism
Phospho-2-dehydro-3-deoxyheptonate aldolase (M1AXT4_SOLTU)	Present	Not detected	metabolism
isopentenyl-diphosphate Delta-isomerase (M1C547_SOLTU)	Present	Not detected	metabolism
homogentisate 1,2-dioxygenase (M1B8K1_SOLTU)	Present	Not detected	metabolism
Glyoxysomal fatty acid beta-oxidation multifunctional protein (M1A065_SOLTU)	Present	Not detected	metabolism
Glutathione transferase (M1ANN8_SOLTU)	Present	Not detected	metabolism
Glucose-6-phosphate isomerase (M1AZE4_SOLTU)	Present	Not detected	metabolism
Fructose-bisphosphate aldolase (M1CQP6_SOLTU)	Present	Not detected	metabolism
formate--tetrahydrofolate ligase (M1D576_SOLTU)	Present	Not detected	metabolism

Protein list	CPG103+tubers	CPG106+tubers	Description
Citrate synthase (M1CTV7_SOLTU)	Present	Not detected	metabolism
Chorismate synthase (M1BB91_SOLTU)	Present	Not detected	metabolism
beta-fructofuranosidase (E1AXT4_SOLTU)	Present	Not detected	metabolism
Aldose 1-epimerase (M0ZLG1_SOLTU)	Present	Not detected	metabolism
Aldo/keto reductase (M1A7Q3_SOLTU)	Present	Not detected	metabolism
Aldehyde dehydrogenase (M1C9T0_SOLTU)	Present	Not detected	metabolism
Aldehyde dehydrogenase (M1BFP7_SOLTU)	Present	Not detected	metabolism
Alcohol dehydrogenase 1 (ADH1_SOLTU)	Present	Not detected	metabolism
ADP/ATP translocase (K7WP02_SOLTU)	Present	Not detected	metabolism
Adenosylhomocysteinase (M0ZZV7_SOLTU)	Present	Not detected	metabolism
Adenosine kinase (M1BY04_SOLTU)	Present	Not detected	metabolism
Aconitate hydratase (M1AGW8_SOLTU)	Present	Not detected	metabolism
6-phosphogluconate dehydrogenase, decarboxylating (M1DSQ9_SOLTU)	Present	Not detected	metabolism
5-methyltetrahydropteroyltriglutamate- -homocysteine S-methyltransferase (M1AGR8_SOLTU)	Present	Not detected	metabolism
3-ketoacyl CoA thiolase 2 (M1CWA6_SOLTU)	Present	Not detected	metabolism
(S)-2-hydroxy-acid oxidase (M1BWS8_SOLTU)	Present	Not detected	metabolism
transketolase (M1A9Z4_SOLTU)	Present	Not detected	metabolism and photosynthesis
transaldolase (M1CPM6_SOLTU)	Present	Not detected	metabolism and photosynthesis
Histidine triad (Hit) protein (M1CQE7_SOLTU)	Present	Not detected	metabolism and signalling
Mitochondrial processing peptidase (M0ZLN7_SOLTU)	Present	Not detected	mitochondrial protein import
Polyadenylate-binding protein (M1B8V3_SOLTU)	Present	Not detected	mRNA metabolism
Tubulin beta chain (M1BNE9_SOLTU)	Present	Not detected	multiple functions, including adaptation to salt stress
Tubulin beta chain (M1ARQ6_SOLTU)	Present	Not detected	multiple functions, including adaptation to salt stress
Tubulin beta chain (B5M4B1_SOLTU)	Present	Not detected	multiple functions, including adaptation to salt stress

Protein list	CPG103+tubers	CPG106+tubers	Description
Tubulin alpha chain (M0ZYR0_SOLTU)	Present	Not detected	multiple functions, including adaptation to salt stress
H1 histone (M0ZW28_SOLTU)	Present	Not detected	multiple functions, including modulation of plant immune responses
Ubiquitin extension protein (O04829_SOLTU)	Present	Not detected	multiple functions, including plant defense
Stem 28 kDa glycoprotein (M1BD67_SOLTU)	Present	Not detected	multiple functions, including plant defense
Probable linoleate 9S-lipoxygenase 3 (LOX13_SOLTU)	Present	Not detected	multiple functions, including plant defense
Leucine-rich repeat family protein (M1A6U3_SOLTU)	Present	Not detected	multiple functions, including plant defense
Dehydroascorbate reductase (Q3HVN5_SOLTU)	Present	Not detected	multiple functions, including plant defense
NAD-dependent malic enzyme 59 kDa isoform, mitochondrial (MAON_SOLTU)	Present	Not detected	multiple functions, including stress response
Monodehydroascorbate reductase (M1AWV2_SOLTU)	Present	Not detected	multiple functions, including stress response
Monodehydroascorbate reductase (M0ZSA3_SOLTU)	Present	Not detected	multiple functions, including stress response
Importin subunit alpha (M1B7C9_SOLTU)	Present	Not detected	multiple functions, including stress response
Hydroxyproline-rich glycoprotein family protein (M0ZJT7_SOLTU)	Present	Not detected	multiple functions, including stress response
Elongation factor 1B alpha-subunit (M1CM67_SOLTU)	Present	Not detected	multiple functions, including stress response
Elongation factor 1-gamma (M1BA84_SOLTU)	Present	Not detected	multiple functions, including stress response
Cysteine proteinase inhibitor (M1C2L8_SOLTU)	Present	Not detected	multiple functions, including stress response
Aspartic protease (M1B7H6_SOLTU)	Present	Not detected	multiple functions, including stress response
Aminoaldehyde dehydrogenase 2 (M1CA90_SOLTU)	Present	Not detected	multiple functions, including stress response
Glutamine synthetase (M1B0M9_SOLTU)	Present	Not detected	nitrogen metabolism
Oxygen-evolving enhancer protein 2, chloroplastic (PSBP_SOLTU)	Not detected	Present	photosynthesis
Ran binding protein-1 (M1AEP8_SOLTU)	Present	Not detected	plant defense
Proteinase inhibitor type-2 K (IP2K_SOLTU)	Present	Not detected	plant defense
Proteinase inhibitor PTI (IP21_SOLTU)	Present	Present	plant defense
Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial (M1AY79_SOLTU)	Present	Not detected	plant development and stress response

Protein list	CPG103+tubers	CPG106+tubers	Description
Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial (M1B744_SOLTU)	Present	Not detected	plant development and stress response
Cell division cycle protein 48 (M1BJD2_SOLTU)	Present	Not detected	plant growth
Xyloglucan endotransglucosylase/hydrolase (M1BZ99_SOLTU)	Present	Not detected	plant growth and development
Cysteine protease (M1ADW0_SOLTU)	Present	Not detected	plant growth and development
Cysteine protease (M0ZWN3_SOLTU)	Present	Not detected	plant growth and development
Actin-58 (M1C5F5_SOLTU)	Present	Not detected	plant growth and development
Actin-101 (M1CWV5_SOLTU)	Present	Not detected	plant growth and development
Uridine 5'-monophosphate synthase (M1BQD4_SOLTU)	Present	Not detected	plant growth and development
Threonine synthase, chloroplastic (THRC_SOLTU)	Present	Not detected	plant growth and development
Rop guanine nucleotide exchange factor (M0ZG60_SOLTU)	Present	Not detected	plant growth and development
Ripening regulated protein DDTFR10 (Q308A7_SOLTU)	Present	Not detected	plant growth and development
Pyruvate decarboxylase (M1D0M1_SOLTU)	Present	Not detected	plant growth and development
Leucine-rich repeat/extensin (M1A616_SOLTU)	Present	Not detected	plant growth and development
GDP-mannose 3',5'-epimerase (A0A0B4J3K8_SOLTU)	Present	Not detected	plant growth and development
Adenylyl cyclase-associated protein (M1D7P9_SOLTU)	Present	Not detected	plant growth and development
3-isopropylmalate dehydratase (M1B1N3_SOLTU)	Present	Not detected	plant growth and development
Pectinesterase (M1AIV9_SOLTU)	Present	Not detected	plant growth, development and metabolism
2-isopropylmalate synthase (M1BCJ8_SOLTU)	Present	Not detected	plant growth, development and metabolism
Zinc finger protein (M1A753_SOLTU)	Present	Not detected	plant growth, development and stress response
PR10 (M0ZMA9_SOLTU)	Present	Not detected	plant growth, development and stress response
Phospholipase D (M1CS26_SOLTU)	Present	Not detected	plant growth, development and stress response
Phospholipase D (M1AKN6_SOLTU)	Present	Not detected	plant growth, development and stress response
Nucleoside diphosphate kinase (M1D2W7_SOLTU)	Present	Not detected	plant growth, development and stress response
Ketol-acid reductoisomerase (M1ADA6_SOLTU)	Present	Not detected	plant growth, development and stress response

Protein list	CPG103+tubers	CPG106+tubers	Description
Glycine-rich protein 2 (M0ZMR8_SOLTU)	Present	Not detected	plant growth, development and stress response
Glutaredoxin (B3F8F4_SOLTU)	Present	Not detected	plant growth, development and stress response
GAST (M1CV60_SOLTU)	Present	Not detected	plant growth, development and stress response
ATP-binding cassette transporter (M0ZQF4_SOLTU)	Present	Not detected	plant homeostasis and stress response
Protein tyrosine phosphatase (M1CLX9_SOLTU)	Present	Not detected	plant signaling
Cysteine protease inhibitor 2 (Fragment) (CPI2_SOLTU)	Not detected	Present	plants defense
Glucose-6-phosphate 1-dehydrogenase, cytoplasmic isoform (G6PD_SOLTU)	Present	Not detected	prevention of oxidative damage
Ferritin (M1CL13_SOLTU)	Present	Not detected	prevention of oxidative damage
Ferritin (M1AH14_SOLTU)	Present	Not detected	prevention of oxidative damage
Serine protease inhibitor 5 (SPI5_SOLTU)	Present	Present	probable plant defense
Probable serine protease inhibitor 6 (SPI6_SOLTU)	Present	Not detected	probable plant defense
Lipoxygenase (M1BVW6_SOLTU)	Present	Not detected	probable plant defense
Linoleate 9S-lipoxygenase 2 (LOX12_SOLTU)	Present	Not detected	probable plant defense
S-adenosylmethionine synthase 2 (METK2_SOLTU)	Present	Not detected	protein biosynthesis
S-adenosylmethionine synthase (M1CQT1_SOLTU)	Present	Not detected	protein biosynthesis
Aspartic proteinase nepenthesin-1 (M1A3D8_SOLTU)	Present	Not detected	protein degradation
26S proteasome non-ATPase regulatory subunit 2 homolog (M1AJV5_SOLTU)	Present	Not detected	protein degradation
26S proteasome non-ATPase regulatory subunit (M1AE01_SOLTU)	Present	Not detected	protein degradation
26S protease regulatory subunit 7 homolog A (M1A1A0_SOLTU)	Present	Not detected	protein degradation
Peptidyl-prolyl cis-trans isomerase (M1A2T5_SOLTU)	Present	Not detected	protein folding
Disulfide-isomerase (M0ZSL7_SOLTU)	Present	Not detected	protein folding
26S protease regulatory subunit 6A homolog (M1CAU1_SOLTU)	Present	Not detected	protein folding
Nascent polypeptide associated complex alpha (M1CPW0_SOLTU)	Present	Not detected	protein folding and targeting
Cucumisin (M1AJY3_SOLTU)	Present	Not detected	protein inhibition
Tom (M1B069_SOLTU)	Present	Not detected	protein recognition

Protein list	CPG103+tubers	CPG106+tubers	Description
phenylalanine--tRNA ligase (M0ZXT4_SOLTU)	Present	Not detected	protein synthesis
Coatomer subunit alpha (M0ZWP5_SOLTU)	Present	Not detected	protein transport
Serine hydroxymethyltransferase (M1BAA6_SOLTU)	Present	Not detected	resistance to biotic and abiotic stress
Serine hydroxymethyltransferase (M1A0B7_SOLTU)	Present	Not detected	resistance to biotic and abiotic stress
Ascorbate peroxidase (M1CY58_SOLTU)	Present	Not detected	response to abiotic stresses
catechol oxidase (M1BMR6_SOLTU)	Present	Not detected	responsible to browning in plants
Thaliana 60S ribosomal protein L7 (M1BV98_SOLTU)	Present	Not detected	ribosomal
Ribosomal protein S9 (M1CY42_SOLTU)	Present	Not detected	ribosomal
Ribosomal protein S9 (M0ZY89_SOLTU)	Present	Not detected	ribosomal
Ribosomal protein S14 (M1CQ57_SOLTU)	Present	Not detected	ribosomal
Ribosomal protein L5 (M1BUM3_SOLTU)	Present	Not detected	ribosomal
Ribosomal protein L3 (Q2VCJ2_SOLTU)	Present	Not detected	ribosomal
Ribosomal protein L28 (M0ZRT6_SOLTU)	Present	Not detected	ribosomal
Ribosomal protein L27a (M1BJP4_SOLTU)	Present	Not detected	ribosomal
Ribosomal protein L15 (M1BHF8_SOLTU)	Present	Not detected	ribosomal
Ribosomal protein (M1CEF5_SOLTU)	Present	Not detected	ribosomal
60S ribosomal protein L9 (M1CD37_SOLTU)	Present	Not detected	ribosomal
60S ribosomal protein L8 (M1BZ12_SOLTU)	Present	Not detected	ribosomal
60S ribosomal protein L6 (M1AFN7_SOLTU)	Present	Not detected	ribosomal
60S ribosomal protein L35a (M1AVU1_SOLTU)	Present	Not detected	ribosomal
60S ribosomal protein L35 (M1C5W2_SOLTU)	Present	Not detected	ribosomal
60S ribosomal protein L22 (M1CAU6_SOLTU)	Present	Not detected	ribosomal
60S ribosomal protein L18a (Q308A3_SOLTU)	Present	Not detected	ribosomal
60S ribosomal protein L18a (K7WTZ1_SOLTU)	Present	Not detected	ribosomal
60S ribosomal protein L17 (Q3HRY2_SOLTU)	Present	Not detected	ribosomal

Protein list	CPG103+tubers	CPG106+tubers	Description
60S ribosomal protein L13a (Q3HRW1_SOLTU)	Present	Not detected	ribosomal
60S ribosomal protein L12 (K7WTX4_SOLTU)	Present	Not detected	ribosomal
60S ribosomal protein L1 (M1CP75_SOLTU)	Present	Not detected	ribosomal
60S acidic ribosomal protein P0 (M1C5E6_SOLTU)	Present	Not detected	ribosomal
50S ribosomal protein L22, chloroplastic (Q307X7_SOLTU)	Present	Not detected	ribosomal
50S ribosomal protein L15 (M1B5R1_SOLTU)	Present	Not detected	ribosomal
40S ribosomal protein S9 (K7WJW2_SOLTU)	Present	Not detected	ribosomal
40S ribosomal protein S8 (M1B539_SOLTU)	Present	Not detected	ribosomal
40S ribosomal protein S5 (M1AF95_SOLTU)	Present	Not detected	ribosomal
40S ribosomal protein S26 (M1CDE9_SOLTU)	Present	Not detected	ribosomal
40S ribosomal protein S2 (M1CBF6_SOLTU)	Present	Not detected	ribosomal
40S ribosomal protein S19 (M1CA24_SOLTU)	Present	Not detected	ribosomal
40S ribosomal protein S17 (M1ACD6_SOLTU)	Present	Not detected	ribosomal
40S ribosomal protein S11 (M1A2N0_SOLTU)	Present	Not detected	ribosomal
30S ribosomal protein S15, chloroplastic (Q2V995_SOLTU)	Present	Not detected	ribosomal
RNA helicase (M1AG92_SOLTU)	Present	Not detected	RNA metabolism
Ribonuclease (M1B4H0_SOLTU)	Present	Not detected	RNA metabolism
Pre-mRNA-splicing factor SF2 (M1AT81_SOLTU)	Present	Not detected	RNA metabolism
Nuclear RNA binding protein (M1BTZ0_SOLTU)	Present	Not detected	RNA metabolism
ATP-dependent RNA helicase (M1CUZ8_SOLTU)	Present	Not detected	RNA metabolism
FCP1 homology domain-containing protein (M0ZS68_SOLTU)	Present	Not detected	RNA processing
1,4-alpha-glucan-branching enzyme (GLGB_SOLTU)	Present	Not detected	starch biosynthesis
Starch-granule-bound R1 protein (M1ACN9_SOLTU)	Present	Not detected	starch metabolism
4-alpha-glucanotransferase, chloroplastic/amyloplastic (DPEP_SOLTU)	Present	Not detected	starch metabolism
Probable inactive patatin-3-Kuras 1 (PT3K1_SOLTU)	Present	Not detected	storage protein + plant defense

Protein list	CPG103+tubers	CPG106+tubers	Description
Patatin-2-Kuras 3 (PT2K3_SOLTU)	Present	Not detected	storage protein + plant defense
Patatin-1-Kuras 2 (PT1K2_SOLTU)	Present	Not detected	storage protein + plant defense
Wound/stress protein (M1B3I7_SOLTU)	Present	Not detected	stress response
Universal stress protein family protein (M1AUH6_SOLTU)	Present	Not detected	stress response
Universal stress protein (M1BZ56_SOLTU)	Present	Not detected	stress response
Thioredoxin h (M1BYR1_SOLTU)	Present	Not detected	stress response
Osmotic stress-activated protein kinase (J9ULI1_SOLTU)	Present	Not detected	stress response
14-3-3-like protein 16R (14335_SOLTU)	Present	Not detected	stress response
Aldo/keto reductase 2 (M0ZQP2_SOLTU)	Present	Not detected	stress response and metabolism
Hsp70-interacting protein 1 (M1A0V9_SOLTU)	Present	Not detected	stress response and protein folding
Pom14 protein (M1AKX9_SOLTU)	Present	Not detected	unknown function
Ly200 (M0ZQH0_SOLTU)	Present	Not detected	unknown function
F17H15.1/F17H15.1 (M1CV01_SOLTU)	Present	Not detected	unknown function
DAG protein (M1BWW8_SOLTU)	Present	Not detected	unknown function

Table S2. The list of all then proteins that were isolated from leaf material and were identified as interaction partners for both CPG103 and CPG106.

The list includes the name of the proteins, Uniprot IDs, information regarding presence or absence of the protein in the sample, as well as description of the protein function. The following list excludes the proteins that were detected as interaction partners in MBP sample.

Protein list	CPG103+leaves	CPG106+leaves	Description
Arginine/serine-rich splicing factor (M1CHP4_SOLTU)	Present	Present	alternative splicing
ATP synthase subunit b, chloroplastic (ATPF_SOLTU)	Present	Not detected	ATP production
ATP synthase subunit b, chloroplastic (ATPF_SOLTU)	Not detected	Present	ATP production
ATP synthase subunit beta (M1A5X1_SOLTU)	Present	Present	ATP production
Thiamine thiazole synthase, chloroplastic (M1BNZ3_SOLTU)	Not detected	Present	biosynthesis of thiamine
Chloroplast heat shock protein 70-2 (M1CBM0_SOLTU)	Present	Not detected	chaperone
Heat shock cognate protein 80 (M1C075_SOLTU)	Present	Not detected	chaperone
Chaperonin 21 (M0ZT88_SOLTU)	Not detected	Present	chaperone
Heat shock protein 70-3 (M0ZHT9_SOLTU)	Not detected	Present	chaperone
Peptidyl-prolyl cis-trans isomerase (M1A2T5_SOLTU)	Not detected	Present	chaperone
S-adenosyl-L-methionine Mg-protoporphyrin IX methyltranserase (M1B4J2_SOLTU)	Present	Not detected	chlorophyll biosynthesis
Chitinase (M0ZH08_SOLTU)	Not detected	Present	defense mechanism
Chitinase (M1BRE5_SOLTU)	Not detected	Present	defense mechanism
Linoleate 13S-lipoxygenase 2-1, chloroplastic (LOX21_SOLTU)	Present	Not detected	defense response
Lipoxygenase (M1D597_SOLTU)	Not detected	Present	defense response
Heparanase (M1CQY8_SOLTU)	Present	Not detected	degradation of ECM
EMB2394 (EMBRYO DEFECTIVE 2394) (M1C7B2_SOLTU)	Not detected	Present	embryo development
fructose-bisphosphate aldolase (M1CJ86_SOLTU)	Present	Not detected	glycolysis
Glyceraldehyde-3-phosphate dehydrogenase (Q8LK04_SOLTU)	Present	Not detected	glycolysis
Glyceraldehyde-3-phosphate dehydrogenase (M1CVH1_SOLTU)	Not detected	Present	glycolysis
Fructose-bisphosphate aldolase (M1C0V6_SOLTU)	Present	Present	glycolysis
WD-repeat protein (M1CAV4_SOLTU)	Not detected	Present	growth and development
glutamate-1-semialdehyde 2,1-aminomutase (M1D5Y7_SOLTU)	Present	Not detected	maintenance of essential biological processes

Protein list	CPG103+leaves	CPG106+leaves	Description
Chloroplast sedoheptulose-1,7-bisphosphatase (M1CL27_SOLTU)	Not detected	Present	metabolism
Cysteine synthase, chloroplastic/chromoplastic (CYSKP_SOLTU)	Not detected	Present	metabolism
Ferredoxin-dependent glutamate synthase 1 (M1AL00_SOLTU)	Not detected	Present	metabolism
UTP--glucose-1-phosphate uridylyltransferase (UGPA_SOLTU)	Not detected	Present	metabolism
Malate dehydrogenase (M1ANW0_SOLTU)	Not detected	Present	metabolism
Pyridoxine biosynthesis protein isoform A (M1CZB0_SOLTU)	Present	Present	metabolism
Succinic semialdehyde reductase isofom1 (M1BT61_SOLTU)	Present	Present	metabolism
NAD dependent epimerase/dehydratase (M1CK21_SOLTU)	Present	Present	metabolism, and potentially stress response
Tubulin alpha chain (M0ZLG7_SOLTU)	Present	Not detected	multiple functions, including adaptation to salt stress
Tubulin beta chain (B5M4B1_SOLTU)	Not detected	Present	multiple functions, including adaptation to salt stress
Xylem serine proteinase 1 (M1AAX9_SOLTU)	Not detected	Present	multiple functions, including immune response
Histone H1 (M1BRI8_SOLTU)	Not detected	Present	multiple functions, including modulation of plant immune responses
Leucine-rich repeat protein (M0ZL78_SOLTU)	Not detected	Present	multiple functions, including plant defense
Germin-like protein (M1B3M1_SOLTU)	Not detected	Present	multiple functions, including plant defense
Lipase (M1C4J2_SOLTU)	Present	Present	multiple functions, including plant defense
Cysteine proteinase inhibitor (M1C2L8_SOLTU)	Not detected	Present	multiple functions, including stress response
Biotin carboxylase (M1BCQ9_SOLTU)	Not detected	Present	multiple functions, including stress response
Elongation factor 1-gamma (M1BA84_SOLTU)	Not detected	Present	multiple functions, including stress response
mRNA binding protein (M1AZB4_SOLTU)	Not detected	Present	multiple functions, including stress response
Glutamine synthetase (M0ZZ31_SOLTU)	Present	Present	nitrogen metabolism
Ribulose bisphosphate carboxylase small subunit, chloroplastic 1 (RBS1_SOLTU)	Present	Not detected	photosynthesis
CP12 (M1AB57_SOLTU)	Not detected	Present	photosynthesis
Chlorophyll a-b binding protein, chloroplastic (M1ABI3_SOLTU)	Not detected	Present	photosynthesis
Chlorophyll a-b binding protein, chloroplastic (M1AF38_SOLTU)	Not detected	Present	photosynthesis

Protein list	CPG103+leaves	CPG106+leaves	Description
Chlorophyll a-b binding protein, chloroplastic (M1AY18_SOLTU)	Not detected	Present	photosynthesis
Chlorophyll a-b binding protein, chloroplastic (M1AY18_SOLTU)	Not detected	Present	photosynthesis
Oxygen-evolving enhancer protein 1, chloroplastic (PSBO_SOLTU)	Not detected	Present	photosynthesis
Photosystem I P700 chlorophyll, apoprotein A1 (PSAA_SOLTU)	Not detected	Present	photosynthesis
Photosystem I reaction center subunit (M1A511_SOLTU)	Not detected	Present	photosynthesis
Photosystem I reaction center subunit IV B isoform 2 (M1BZP9_SOLTU)	Not detected	Present	photosynthesis
Photosystem II 10 kDa polypeptide, chloroplastic (A0A0C4B0N5_SOLTU)	Not detected	Present	photosynthesis
Photosystem II 10 kDa polypeptide, chloroplastic (PSBR_SOLTU)	Not detected	Present	photosynthesis
Photosystem II 22 kDa protein, chloroplastic (M1BHC8_SOLTU)	Not detected	Present	photosynthesis
Photosystem II D2 protein (PSBD_SOLTU)	Not detected	Present	photosynthesis
Photosystem Q(B) protein (M1DUW7_SOLTU)	Not detected	Present	photosynthesis
Electron transporter (M0ZIN1_SOLTU)	Present	Present	photosynthesis
Chlorophyll a-b binding protein, chloroplastic (M1BWY5_SOLTU)	Present	Present	photosynthesis
Chlorophyll a-b binding protein, chloroplastic (M1CIH8_SOLTU)	Present	Present	photosynthesis
Photosystem I reaction center subunit II, chloroplastic (Q70PN9_SOLTU)	Present	Present	photosynthesis
Photosystem II CP47 chlorophyll apoprotein (M1DZ44_SOLTU)	Present	Present	photosynthesis
Oligopeptidase (M1BGY6_SOLTU)	Not detected	Present	plant defense
Proline-rich protein (M0ZT17_SOLTU)	Not detected	Present	plant defense
Proteinase inhibitor type-2 T (IP2T_SOLTU)	Not detected	Present	plant defense
Fruit protein PKIWI502 (M1D095_SOLTU)	Not detected	Present	plant growth and development
Actin (M1CZ42_SOLTU)	Not detected	Present	plant growth and development
Actin-71 (ACT6_SOLTU)	Not detected	Present	plant growth and development
Expansin (M0ZM03_SOLTU)	Present	Present	plant growth and development
3-isopropylmalate dehydratase (M1B1N3_SOLTU)	Present	Present	plant growth and development
Phosphoglucomutase, cytoplasmic (PGMC_SOLTU)	Not detected	Present	plant growth, development and metabolism
Hydrolase, hydrolyzing O-glycosyl compounds (M1CLM2_SOLTU)	Present	Not detected	plant growth, development and stress response
Elongation factor G, chloroplastic (M1A0N5_SOLTU)	Present	Not detected	protein biosynthesis

Protein list	CPG103+leaves	CPG106+leaves	Description
Elongation factor Tu (M1DPA6_SOLTU)	Present	Not detected	protein biosynthesis
S-adenosylmethionine synthase 1 (METK1_SOLTU)	Not detected	Present	protein biosynthesis
Elongation factor Tu (M1DPA6_SOLTU)	Not detected	Present	protein biosynthesis
Aspartic protease inhibitor 10 (API10_SOLTU)	Not detected	Present	protein degradation
Aspartic proteinase nepenthesin-1 (M1A3D8_SOLTU)	Not detected	Present	protein degradation
Nascent polypeptide associated complex alpha (M1CPW0_SOLTU)	Present	Present	protein folding and targeting
Cucumisin (M1A6Q8_SOLTU)	Not detected	Present	protein inhibition
Thioredoxin m(Mitochondrial)-type (M1B8P1_SOLTU)	Present	Present	redox signaling
ATP-dependent Clp protease ATP-binding subunit clpA homolog CD4B, chloroplastic (M1B5I1_SOLTU)	Not detected	Present	removal of damaged and misfolded proteins
ATP-dependent Clp protease proteolytic subunit (M1CDL1_SOLTU)	Not detected	Present	removal of damaged and misfolded proteins
ATP-dependent clp protease ATP-binding subunit clpx (M1AIM3_SOLTU)	Present	Present	removal of damaged and misfolded proteins
Replicase (M1C030_SOLTU)	Not detected	Present	replication of viral proteins
Catalase (M1CVH4_SOLTU)	Present	Present	response to stimulus
30S ribosomal protein S10, chloroplastic (M1B535_SOLTU)	Present	Not detected	ribosomal
60S ribosomal protein L35a (M1AVU1_SOLTU)	Present	Not detected	ribosomal
60S ribosomal protein L6 (M1AFN7_SOLTU)	Present	Not detected	ribosomal
Ribosomal protein (M1CEF5_SOLTU)	Not detected	Present	ribosomal
30S ribosomal protein S1, chloroplastic (M0ZL57_SOLTU)	Not detected	Present	ribosomal
40S ribosomal protein S15a-1 (Q3HRZ3_SOLTU)	Not detected	Present	ribosomal
40S ribosomal protein S19 (M1CA24_SOLTU)	Not detected	Present	ribosomal
40S ribosomal protein S30 (M0ZQX8_SOLTU)	Not detected	Present	ribosomal
40S ribosomal protein S8 (M1BS92_SOLTU)	Not detected	Present	ribosomal
50S ribosomal protein L3, chloroplastic (M1B9U2_SOLTU)	Not detected	Present	ribosomal
50S ribosomal protein L4, chloroplastic (M1AA49_SOLTU)	Not detected	Present	ribosomal
60S ribosomal protein L4/L1 (M1ARJ5_SOLTU)	Not detected	Present	ribosomal
Ribosomal protein L5 (M1BUM3_SOLTU)	Not detected	Present	ribosomal
Phosphoribulokinase (M1AJF6_SOLTU)	Present	Present	ribosomal

| Protein list	CPG103+leaves	CPG106+leaves	Description
Ribosomal protein S7 (M1AUW2_SOLTU)	Present	Present	ribosomal
30S ribosomal protein S15, chloroplastic (Q2V995_SOLTU)	Present	Present	ribosomal
40S ribosomal protein S4 (RS4_SOLTU)	Present	Present	ribosomal
40S ribosomal protein S8 (M1BUI8_SOLTU)	Present	Present	ribosomal
50S ribosomal protein L17, chloroplastic (M1CNA2_SOLTU)	Present	Present	ribosomal
50S ribosomal protein L22, chloroplastic (RK22_SOLTU)	Present	Present	ribosomal
60S ribosomal protein L18a (M1BT85_SOLTU)	Present	Present	ribosomal
60S ribosomal protein L22 (M1CAU6_SOLTU)	Present	Present	ribosomal
Ribosomal protein L27a (M1BJP4_SOLTU)	Present	Present	ribosomal
Ribosomal protein S9 (M1BEQ3_SOLTU)	Present	Present	ribosomal
ATP-dependent RNA helicase (M1AEQ1_SOLTU)	Not detected	Present	ribosome biogenesis, can also play role in stress response
Extracellular Ca2+ sensing receptor (M1C9W6_SOLTU)	Present	Not detected	signaling, stress response, and photosynthesis
ATP binding protein (M1CT07_SOLTU)	Not detected	Present	takes part in numerous cellular activities
Translation initiation factor IF-3 (M1ARZ2_SOLTU)	Present	Not detected	translation regulation
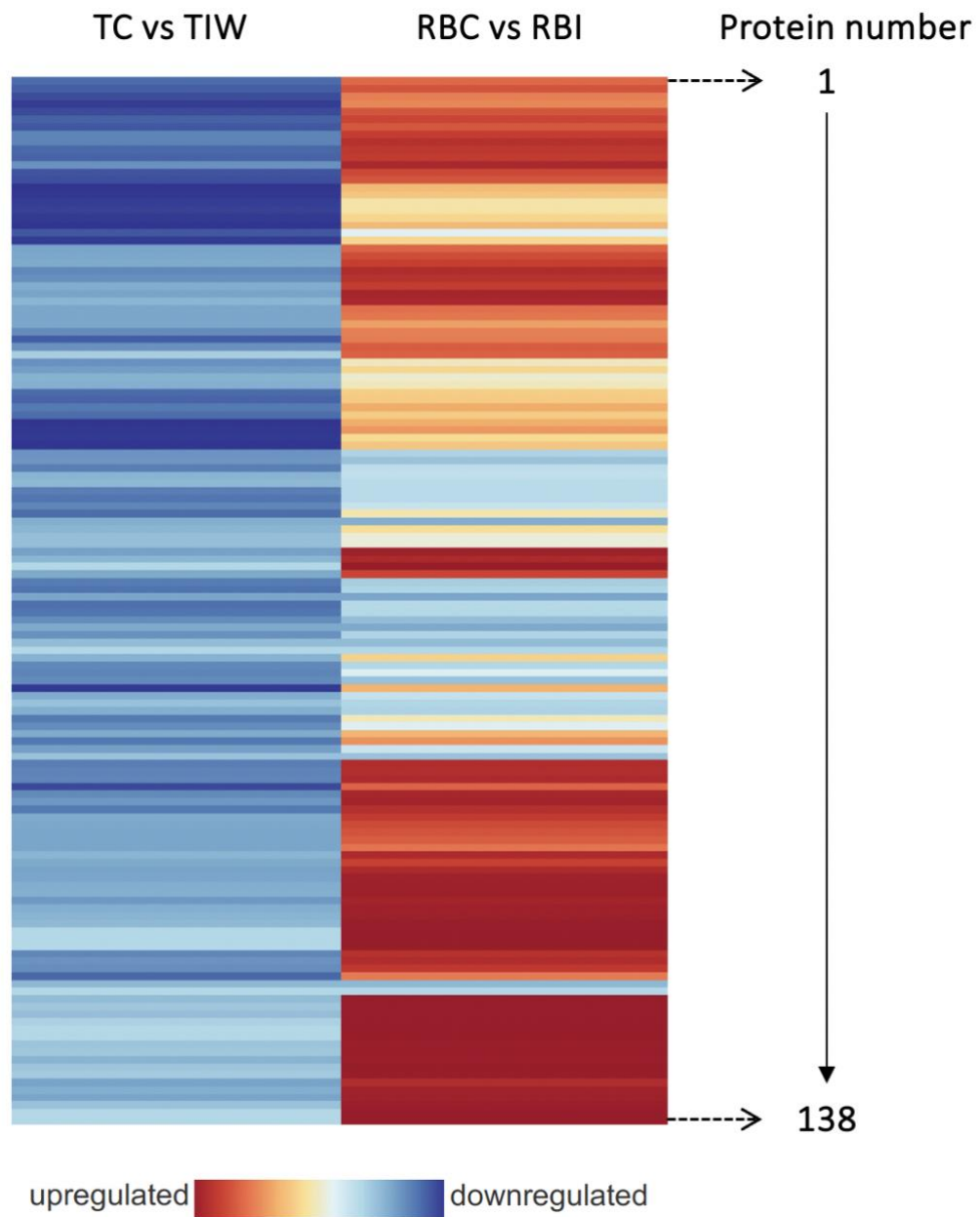
Figure S1. The numbering of proteins in the heat map that illustrates the expression levels of proteins present in both Russet Burbank and Targhee Russet varieties.

Table S3. The list of protein names and their functions that are present in both Russet Burbank and Targhee Russet varieties.

The list below includes the information about names and functions of the proteins that are present in both varieties of potatoes and were used to construct a heat map shown in Figures 8 and S1.

Protein number	Protein name	Protein function
1	Phosphoglycerate kinase	Metabolism
2	PR-10	Multiple functions, including plant defense
3	Malate dehydrogenase	Metabolism
4	Lactoylglutathione lyase	Plant defense
5	Neutral leucine aminopeptidase protein	Protein degradation and recycling
6	60S acidic ribosomal protein	Ribosomal
7	Protein disulfide-isomerase	Protein folding
8	ATP synthase subunit beta	Synthesis of ATP
9	Actin	Plant growth and development
10	Elongation factor 1-alpha	Protein biosynthesis
11	Glyceraldehyde-3-phosphate dehydrogenase	Metabolism
12	Catechol oxidase	Plays role in enzymatic browning
13	60S acidic ribosomal protein P0	Ribosomal
14	6-phosphogluconate dehydrogenase, decarboxylating	Metabolism and stress response
15	4-alpha-glucanotransferase, chloroplastic/amyloplastic	Starch metabolism
16	Aspartic protease	Plant defense
17	Serine protease inhibitor 5	Plant defense
18	Dehydroascorbate reductase	Stress response
19	Alpha-1,4-glucan phosphorylase	Starch degradation
20	UTP-glucose-1-phosphate uridylyltransferase	Biosynthesis
21	Cysteine protease inhibitor	Plant defense
22	Cysteine protease inhibitor	Plant defense
23	ADP/ATP translocase, mitochondrial	ATP/ADP exchange across membrane
24	Apyrase	Plant growth and development
25	Pyruvate decarboxylase	Metabolism
26	Annexin	Stress response
27	Malic enzyme	Multiple functions, including stress response
28	Ci21A protein	Stress response

Protein number	Protein name	Protein function
29	Glyceraldehyde-3-phosphate dehydrogenase	Metabolism
30	Formate dehydrogenase, mitochondrial	Multiple functions, including stress response
31	Heat shock protein	Chaperone
32	30S ribosomal protein S15, chloroplastic	Ribosomal
33	RuBisCO large subunit-binding protein subunit alpha, chloroplastic	Photosynthesis
34	Probable linoleate 9S-lipoxygenase 3	Multiple functions, including plant defense
35	Cysteine protease inhibitor	Plant defense
36	Multicystatin	Plant defense
37	Catalase isozyme 2	Stress response
38	Fasciclin-like arabinogalactan protein 13	Plant development and stress response
39	Patatin	Storage and plant defense
40	Miraculin	Taste modification
41	Linoleate 9S-lipoxygenase 2	Plant defense
42	25 kDa protein dehydrin	Stress response
43	Probable inactive patatin-3-Kuras 1	Plant defense and storage
44	Transaldolase	Metabolism
45	Glutaredoxin	Multiple functions, including stress response
46	Annexin	Stress response
47	Phosphopyruvate hydratase	Metabolism
48	Low molecular weight heat-shock protein	Chaperone
49	Linoleate 9S-lipoxygenase 1	Plant defense
50	Serine protease inhibitor 2	Plant defense
51	Patatin-07	Storage and plant defense
52	Temperature-induced lipocalin	Stress response
53	Probable linoleate 9S-lipoxygenase 8	Multiple functions, including plant defense
54	Miraculin	Taste modification
55	Kunitz-type enzyme inhibitor S9C11	Plant defense
56	Induced stolon tip protein	Plant growth and development (tuber formation)
57	14-3-3 protein	Plant defense
58	Chilling-responsive protein	Stress response
59	Aldehyde dehydrogenase	Stress response

Protein number	Protein name	Protein function
60	60S ribosomal protein L1	Ribosomal
61	Histone H2B	Plant growth and development
62	Protein disulfide-isomerase	Protein folding
63	Frustose-bisphosphate aldolase	Metabolism
64	Patatin T-5	Storage and plant defense
65	Acidic ribosomal protein P1a	Ribosomal
66	Abscisic acid and environmental stress-inducible protein TAS14	Stress response
67	Glutaredoxin-dependent peroxiredoxin	Multiple functions, including stress response
68	Aspartic protease inhibitor 1	Plant defense
69	Heat shock protein	Chaperone
70	Patatin-B2	Storage and plant defense
71	Patatin-2-Kuras 1	Storage and plant defense
72	Patatin	Storage and plant defense
73	Acyl-CoA-binding protein	Plant development and stress response
74	Aspartic protease inhibitor	Plant defense
75	Protease C56	Plant development and stress response
76	Protease inhibitor PTI	Plant defense
77	Cysteine protease inhibitor 10 or 8	Plant defense
78	Proteinase inhibitor 2	Plant defense
79	Proteinase inhibitor 1	Plant defense
80	Cysteine protease inhibitor	Plant defense
81	Proteinase inhibitor type 2	Plant defense
82	Wound-induced proteinase inhibitor 1	Plant defense
83	Patatin	Storage and plant defense
84	Patatin	Storage and plant defense
85	Glycine-rich RNA binding protein	Plant growth and development
86	Patatin-2-Kuras 3	Storage and plant defense
87	Aspartic protease inhibitor 8 or 11	Plant defense
88	Phoshoglucomutase, cytoplasmic	Metabolism
89	Phoshoglucomutase	Metabolism
90	Ran binding protein-1	Protein transport
91	Triosephosphate isomerase	Metabolism
92	Malate dehydrogenase	Metabolism
93	NEDD9 RUB2	Regulation of protein activity
94	14-3-3	Plant defense
95	Fructose-bisphosphate aldolase	Metabolism
96	Tubulin alpha chain	Plant growth and development
97	Glucose-1-phosphate adenylyltransferase	Plant growth and development

Protein number	Protein name	Protein function
98	Gamma interferon inducible lysosomal thiol reductase family protein	Stress response
99	Stem 28 kDa glycoprotein	Plant growth and development
100	Isopentenyl-disphosphate delta-isomerase	Biosynthesis
101	14-3-3	Plant defense
102	Chloroplast small heat shock protein class I	Chaperone
103	Nascent polypeptide associated complex alpha	Transcription
104	Thaliana 60S ribosomal protein L7	Ribosomal
105	Histone H4	Plant growth and development
106	Transketolase	Metabolism
107	Molecular chaperone Hsp90-1	Chaperone
108	1,4-alpha-glucan branching enzyme	Starch synthesis
109	Superoxide dismutase [Cu-Zn]	Stress response
110	Nucleoside disphosphate kinase	Multiple functions, including stress response
111	Aminoaldehyde dehydrogenase 2	Metabolism
112	Pyrophosphate-fructose 6-phosphate 1-phosphotransferase subunit beta	Metabolism
113	Ribosomal protein S14	Ribosomal
114	D-3-phosphoglycerate dehydrogenase	Metabolism
115	Aquaporin	Transport of water and neutral molecules
116	Adenosylhomocysteinase	Biosynthesis
117	Heat shock protein 70-3	Chaperone
118		
119	Peptidyl-propyl cis-trans isomerase	Protein folding
120	Thioredoxin H-type 2	Homeostasis
121	Annexin	Stress response
122	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	Multiple functions, including stress response
123	40S ribosomal protein S3a	Ribosomal
124	GTP-binding protein	RNA export and protein transport
125	Glutathione transferase	Multiple functions, including stress response
126	Thaumatin	Plant defense
127	NAD-dependent malic enzyme 62 kDa isoform, mitochondrial	Multiple functions, including stress response
128	Translationally-controlled tumor protein homolog	Homeostasis

Protein number	Protein name	Protein function
129	Elongation factor 1-gamma	Protein biosynthesis
130	Triosephosphate isomerase	Metabolism
131	5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase	Metabolism
132	Protease subunit alpha type	Multiple functions, including stress response
133	Monodehydroascorbate reductase (NADH)	Metabolism
134	Uncharacterized protein	Unknown
135	Heat shock protein hsp70	Chaperone
136	CBS domain-containing protein	Multiple functions, including stress response
137	Endochitinase 3 or 4	Plant defense
138	Chitinase	Plant defense

Curriculum Vitae

Daria Baskova

M.Sc. Candidate in Analytical/Physical Chemistry

Agriculture and Agri-Food Canada (AAFC) and Western University

Post-Secondary Degrees

MSc, Analytical Chemistry September 2022-Present
Western University – London, ON

Bachelor of Applied Technology – Biotechnology September 2018-April 2022
Fanshawe College – London, ON

Honours and Awards

Dean's Honour Roll September 2019 – December 2019

Dean's Honour Roll January 2019 – April 2019

International Entrance Degree Program Scholarship January 2019 – April 2019

Work Experience

Teaching Assistant (Chemistry 2213A and Chemistry 3372G)   September 2022 – April 2024
Western University-London, ON

- Supervised organic chemistry and analytical chemistry laboratory sessions for groups of up to 24 second- and third-year students

- Gave weekly pre-lab instructional talks

- Created marking schemes to mark reports and marked lab reports in a timely manner

- Provided instructions to the students on prelab/report submissions

- Respond to students' emails on lab-related questions

FSWEP Student October 2021 – August 2022
Agriculture and Agri-Food Canada-London, ON

- Prepared grain samples for the extraction, quantitative and qualitative analysis of mycotoxins and other contaminants

- Assisted in the development and validation of analytical methods and their transfer to a new automated robotics platform

- Analyzed samples by high-resolution mass spectrometry and gained experience with method development of full and tandem MS methods

- Performed literature review, initial data analysis of results and generation of results in a written report

- Aided in general laboratory cleaning and organization, as well as maintenance and quality assurance of analytical instruments

- Learned to aseptically propagate plant tissue from parent organism using micropropagation technique

Research Assistant   March 2021 – August 2021
Fanshawe College – London, ON

- Performed Oxitest analysis to determine oxidation times of lipids present in cold brew caffeinated samples

- Assisted in streaking Petri dishes, inoculation of petrifilm plates, and running PCR reactions followed by DNA sequencing to characterize the fungal and bacterial species identified in the samples

Conferences

Canadian Chemistry Conference and Exhibition   June 2024
RBC Convention Centre Winnipeg – Winnipeg, MB

- Presented a poster that summarizes the main finding on the original master's project

- Attended talks presented in the Analytical Chemistry General Symposiun, Nanomaterials, Natural Product Biosynthesis, and Biomedical Mass Spectrometry sections and participated in discussions during questioning period

Trent Conference on Mass Spectrometry   August 2023
Trent University – Peterborough, ON

- Presented a 15-minute presentation on original research

- Attended talks on new findings and novel methods used in my field of study

Post-ASMS Symposium   November 2022
Four Points by Sheraton Toronto Airport – Toronto, ON

- Attended seminars on novel mass spectrometry methods and results

Summary of Coursework

Chemistry 9713Q – Advanced Medicinal Chemistry – 90

Biochemistry 9522R – Applied Proteomics – 88

Chemistry 9544B – Mass Spectrometry: Fundamentals and Biochemical Applications – 95