CHARACTERIZATION OF NOVEL *LEGIONELLA PNEUMOPHILA* PHOSPHOINOSITIDE BINDING EFFECTOR PROTEINS

by

Adam Kerzner

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Bachelor of Science Honors Degree in Biological Sciences with Distinction

Spring 2024

CHARACTERIZATION OF NOVEL *LEGIONELLA PNEUMOPHILA* PHOSPHOINOSITIDE BINDING EFFECTOR PROTEINS

by

Adam Kerzner

Approved: Ramona Neunuebel

Ramona Neunuebel, PhD
Professor in charge of thesis on behalf of the Advisory Committee

Approved: John R. Jungck

John Jungck, PhD
Committee member from the Department of Biological Sciences

Approved: Carlton R Cooper

Carlton Cooper, PhD
Committee member from the Board of Senior Thesis Readers

Approved: _____

Michael Chajes, Ph.D.
Dean, University Honors Program

ACKNOWLEDGMENTS

This project represents an important milestone in my scientific journey, especially as I transition to pursuing my PhD. The first person I would like to thank is Dr. Ramona Neunuebel for giving me the opportunity to work in her lab for two and a half years. She not only made herself available at every available time she had when I had questions or to review a paper, she also gave me priceless advice navigating not only my time in her research lab, but also navigating my scientific career in graduate school and beyond. She has taught me irreplaceable knowledge that I can carry with me for the rest of my life. In addition, I would like to thank the current and former lab members of the Neunuebel lab: Abby Bolt, Jake Ellis, Marina Grossi, Summer Hackenburg, and Abby Richardson, for not only teaching me research techniques, but also answering every question I had with enthusiasm.

I would like to give a special acknowledgement to Abby Bolt for being my primary lab mentor throughout my time in the lab. Abby took me under her wing during my sophomore and junior years in the lab and granted me the opportunity to have a part in her project. She taught me countless laboratory techniques and always spent extra time explaining concepts to me to make sure I knew why I was doing each step. Many of the transferable skills I have developed is due to her commitment to being a good mentor.

I also would like to thank Dr. Jungck and Dr. Cooper as my second and third readers, respectively. Dr. Jungck introduced me to the field of bioinformatics, and he always took extra time to guide me through employing bioinformatics with my research. Because of his excitement for the field of bioinformatics, it is an area I would like to pursue in graduate school and possibly with my career. Dr. Cooper introduced me to the concepts of cancer biology and while it does not have any direct impact on this project, his encouragement of cancer biology has made me explore this area as well in graduate school.

I would also like to thank the University of Delaware and the Delaware Biosciences Association. These two organizations helped fund my work over the summers of 2023 and 2024 and much of the work in this project was completed during this time.

Finally, I would like to thank my family and friends for the countless amount of support they have given me during this project. Many of them not only gave me the encouragement to join Dr. Neunuebel's lab in the first place, but also took on the project of senior thesis. I appreciate all the support you have given me over the years.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Legionella pneumophila is a Gram-negative, intracellular pathogen that infects alveolar macrophages and causes Legionnaires' disease, a debilitating pneumonia. *L. pneumophila* manipulates host cell vesicular trafficking pathways to help construct a self-contained replicative niche known as the Legionella Containing Vacuole (LCV). Upon establishment of this LCV, *L. pneumophila* replicates freely inside of this vacuole without detection from the host cell until eventual host cell death allowing further infection of nearby macrophages. *L. pneumophila* employs its Icm/Dot type IV secretion system to translocate over 300 effector proteins into the host cell. Many of these effector proteins help construct the LCV and manipulate host vesicular pathways through the binding to phosphoinositides (PIs). PIs serve as crucial two-tailed

phospholipids that can be phosphorylated in key positions to control vesicular trafficking. Our goal is to identify and characterize PI-binding effector proteins as little is known about these proteins beyond their binding to PIs. Towards this goal, I have two objectives. The first objective was to create HaloTagged fusion constructs of three effector proteins hypothesized to bind to PIs. These fusion constructs will allow us to not only confirm PI-binding, but also discover when and where these proteins are secreted and if they bind to any other organelles that contribute to the establishment of the LCV. The second objective was to identify the specific region or residues of a PIbinding protein that allowed this binding to occur. This will help to elucidate the unique and specific properties of *L. pneumophila* PI-binding proteins as opposed to eukaryotic PI-binding proteins. Validating and characterizing the effector proteins that bind to PIs is crucial to not only understanding how *L. pneumophila* evades host cell detection, but also developing agents to specifically target and inhibit pathogenic proteins that bind to PIs to allow host cells to degrade *L. pneumophila*.

Chapter 1

INTRODUCTION

1.1 Background and Significance

1.1.1 Introduction to *Legionella Pneumophila*

Mammalian immune systems are exposed to countless microbial species including many pathogenic bacteria and parasites over their lifetime. Broadly, the immune system combats these pathogens expediently and effectively, returning the body to homeostasis. However, a few of these pathogens can be left undetected by the immune system, infecting numerous host cells, and causing diseases that are sometimes fatal. One such pathogen, *Legionella pneumophila*, can evade and take advantage of mammalian immune systems swiftly and easily causing a debilitating pneumonia known as Legionnaires' disease[1]. *L. pneumophila* is a Gram-negative rodshaped bacterium that can measure between 2-20 micrometers long[1]. *L. pneumophila* possess pili, which enable it to adhere to different environmental surfaces or to other cells, as well as a single, polar flagellum granting motility[1]. While *L. pneumophila* is commonly referred to as a fastidious bacterium due to its requirements for specific supplements such as L-cysteine and ferric iron, *L. pneumophila* still manages to replicate quite effectively in surface and drinking water or other water-based environments usually hostile to phagocytic cells[1]. *L. pneumophila* is usually present on the surface waters of rivers, lakes, or drinking water[1]. *L. pneumophila* is also particularly present in conjunction with free-living amoebae, which act as *L. pneumophila's* primary host in nature[1]. Additionally, *L. pneumophila* seems to possess a natural resistance to heat and chlorine as it can still be found in hot

water tanks and public fountains[1]. *L. pneumophila* carries particularly well in aerosols of water and thus cooling towers and evaporative condensers serve as outdoor sources of dissemination, and humidifiers and nebulizers serve as indoor sources of dissemination[1]. The first widely recognized outbreak of *L. pneumophila* occurred in 1976 at a Philadelphia hotel where the bacterium may have been disseminated through the hotel's heating, ventilation, and air conditioning system[2].

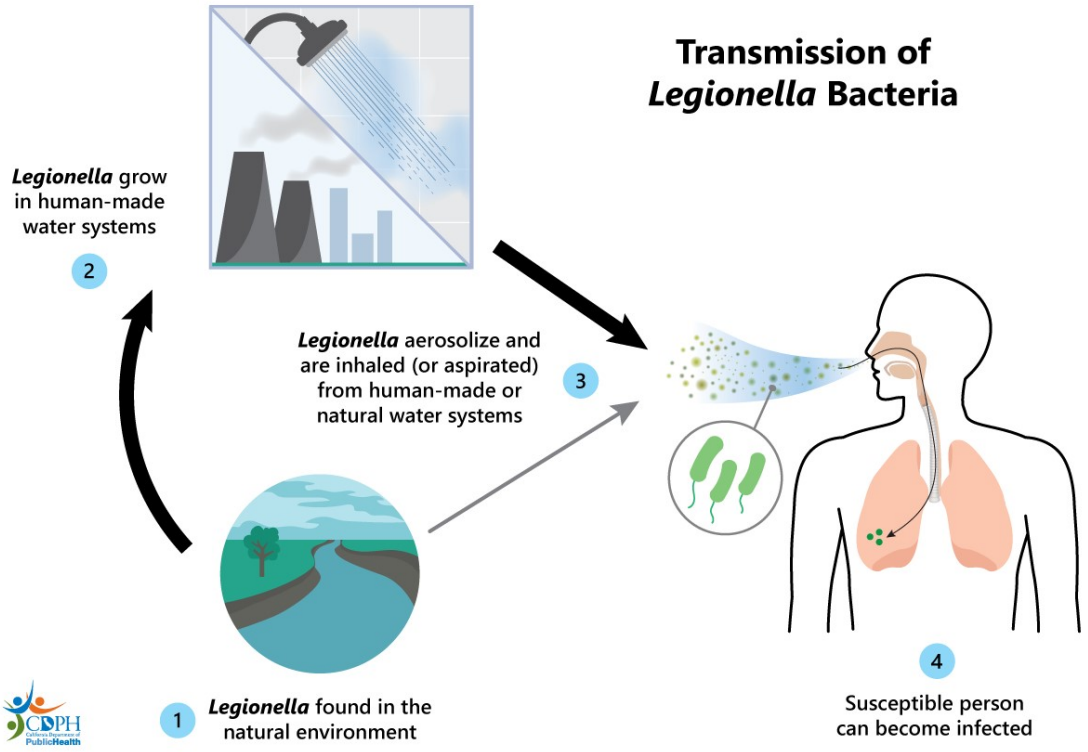


Figure 1 How *Legionella pneumophila* spreads. *L. pneumophila* naturally coexists with amoeba in freshwater environments where it can spread to manmade water systems. From here, aerosolized *L. pneumophila* bacteria can be inhaled and infect a person[3].

1.1.2 Clinical Relevance and Symptoms

Legionnaires' disease is a severe form of pneumonia or lung inflammation that is caused by infection from Legionella[4]. The bacterium can also cause an infected individual to develop Pontiac fever, which is a flu-like illness[4]. Upon exposure,

symptoms can develop in two to ten days and frequently begin with common symptoms such as headaches, muscle aches, and fever[4]. After two to three days of symptoms, other symptoms may occur such as a cough, shortness of breath, chest pain, gastrointestinal symptoms, and possibly confusion[4]. While Legionnaires' disease usually only impacts the lungs, it can potentially spread to other organs such as the heart[4]. While most individuals suffering from Legionnaires' disease can return to health upon swift treatment with antibiotics, some individuals display lasting effects, and untreated Legionella infection can be fatal[4]. Life-threatening complications caused by antibiotics, respiratory failure, septic shock, and acute kidney failure can occur without prompt treatment with antibiotics[4]. However, even with rapid treatment, longterm effects can still occur months and even years after initial diagnosis and treatment[5]. Years after recovery from Legionnaires' disease, individuals may be more likely to experience fatigue, neurological symptoms, or neuromuscular symptoms, and health-related quality of life could be diminished[5]. Individuals may also be more likely to develop lung abnormalities and reduced carbon monoxide transport even long after infection[6]. However, exposure to Legionella does not necessarily result in developing Legionnaires' disease. Certain risk factors including smoking, weakened immune system, chronic lung disease or ages over 50 can increase the risk of developing Legionnaires' disease and developing more severe symptoms and complications[4]. While antibiotics currently serve as an effective treatment if used quickly, Legionella strains are beginning to show high prevalence of resistance to commonly used antibiotics[7]. Because Legionella will continue to develop multi drug resistance, it is imperative that new therapeutics and agents are developed to prevent complications and even fatalities from Legionnaires' disease.

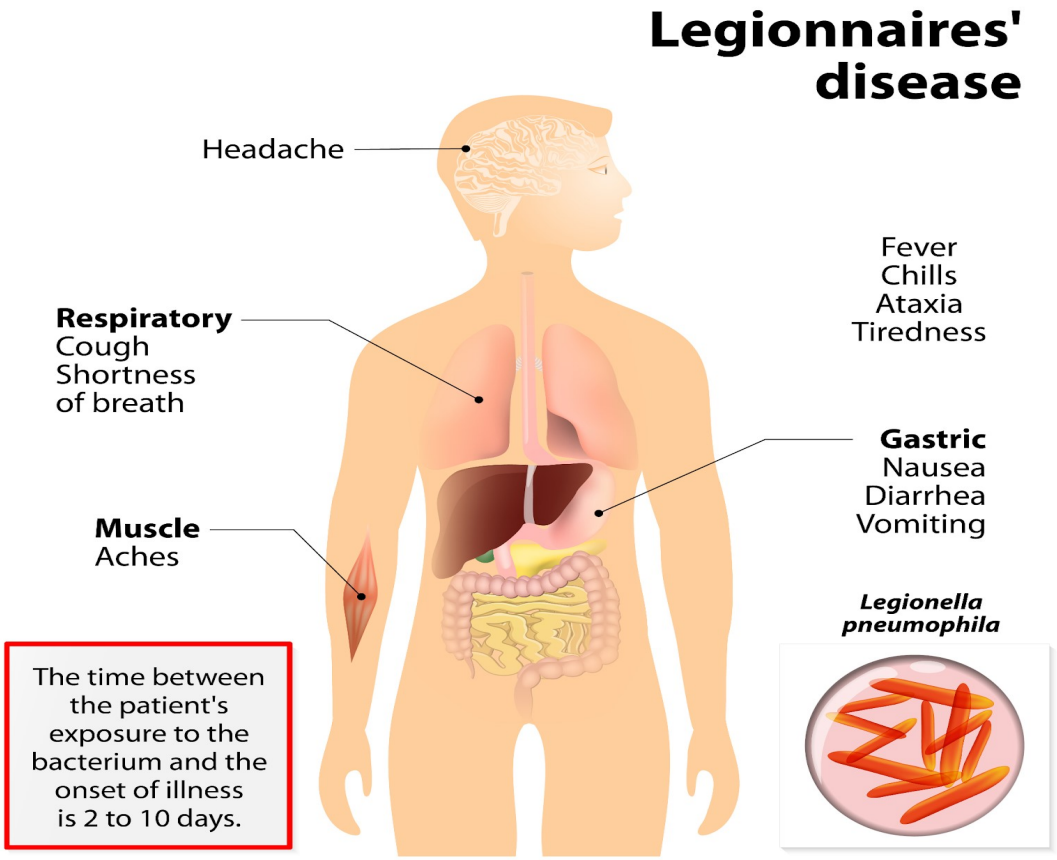


Figure 2 Legionnaires' disease symptoms. *L. pneumophila* infection can cause Legionnaires' disease, a sometimes-fatal pneumonia causing respiratory, muscle, and gastric symptoms[8].

1.2 Pathogenesis of *L. pneumophila*

1.2.1 Mechanism of Infection

Legionella species have historically evolved to interact and replicate with different protozoa, especially amoeba, which coincidentally gives *L. pneumophila* the ability to replicate in human macrophages as an accidental pathogen[9]. Typically, microbes are phagocytosed by a macrophage and a phagosome will encapsulate the microbe until maturing into a digestive vacuole via the endocytic pathway[10]. Here, the vacuole will be acidified, and the microbe will be degraded without causing further

infection[10]. The endocytic pathway takes advantage of endosomal markers such as small GTPases to help form the phagosome[10]. *L. pneumophila* is not only able to delay this endosome formation, but also recruit vesicles and membrane lipids disabling the host cell's secretory pathway, which enables the construction of its own vacuole[11]. This vacuole is known as a Legionella Containing Vacuole (LCV) and allows *L. pneumophila* to replicate undisturbed by the traditional endocytic pathway[11]. However, while the LCV does not mature along the phagocytic or endocytic pathway, yet the LCV still interacts with host vesicles during its biogenesis and establishment as a replicative niche[11]. Following phagocytosis, the LCV recruits proteins derived from secretory vesicles that cycle between the Endoplasmic Reticulum (ER) and the Golgi Apparatus and ER-derived vesicles begin to fuse with the LCV[11]. Within 4 to 6 hours of ER-derived vesicle recruitment to the LCV, the LCV membrane is almost completely composed of the rough ER[11]. *L. pneumophila* can then freely replicate inside of this vacuole until *L. pneumophila* is ultimately released from the host cell to infect new cells[11].

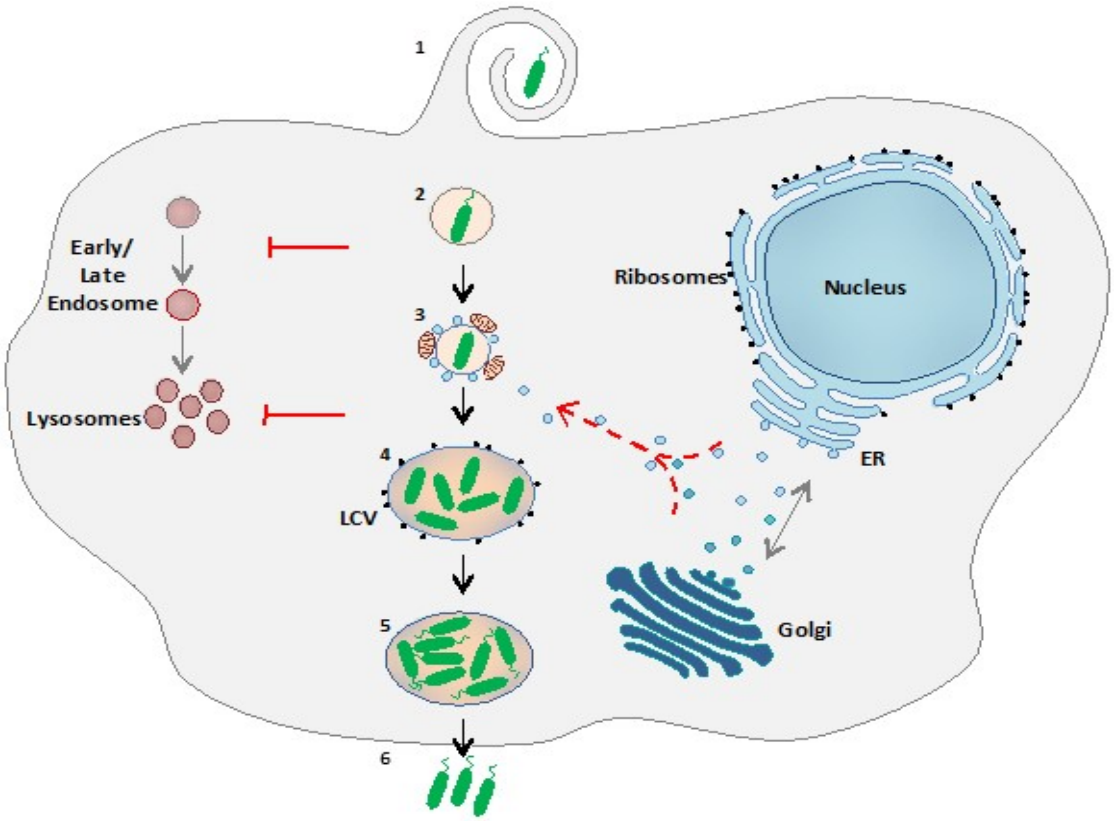


Figure 3 *L. pneumophila* avoids endosomal maturation and establishes a *Legionella*-containing vacuole. The host cell still phagocytizes *L. pneumophila*, however, the bacteria avoid maturation into an endosome and instead creates its own vacuole to replicate in[12].

1.2.2 *L. pneumophila* protein secretion

L. pneumophila's establishment of the LCV is enabled by its type IV secretion system (T4SS), called Dot/Icm, to translocate over 300 bacterial effector proteins into host cells[13]. Effector proteins are likely involved in all stages of *L. pneumophila's* life cycle and they serve as crucial weapons in *L. pneumophila's* arsenal to manipulate eukaryotic cellular processes[13]. The T4SS rapidly injects effector proteins into the host cell upon *L. pneumophila* infection[13]. These early secreted proteins allow prevention of the fusion between the LCV and endocytic vesicles or switches the LCV off the

classical endocytic pathway onto an alternative pathway[13]. These effector proteins also are crucial to help build the LCV through the recruitment of ER-derived vesicles that fuse to the LCV[13]. Some effector proteins specifically target host cell regulatory proteins that control vesicle traffic between the ER and the Golgi[13]. Upon subversion of host cell vesicular membrane transport, other effector proteins can control activation of host cell proteins to facilitate LCV fusion with the ER vesicles[13]. Many of these host cell proteins are small GTPases that cycle between an inactive GDPbound form and an active GTP-bound form[14]. Effector proteins can also function as GEFs (nuclear guanine exchange factor) or GAPs (GTPase activating protein) to control activation and deactivation of host cell small GTPases[14]. *L. pneumophila* effectors can also upregulate several host anti-apoptotic genes, which prevents host cell death[14]. The host cell's survival is critical for *L. pneumophila* survival and replication especially at early stages of infection[14]. *L. pneumophila's* employment of over 300 effector proteins enables not only the establishment of the LCV, but also enables avoidance of endosomal maturation and fusion of the LCV with ER-derived vesicles[14].

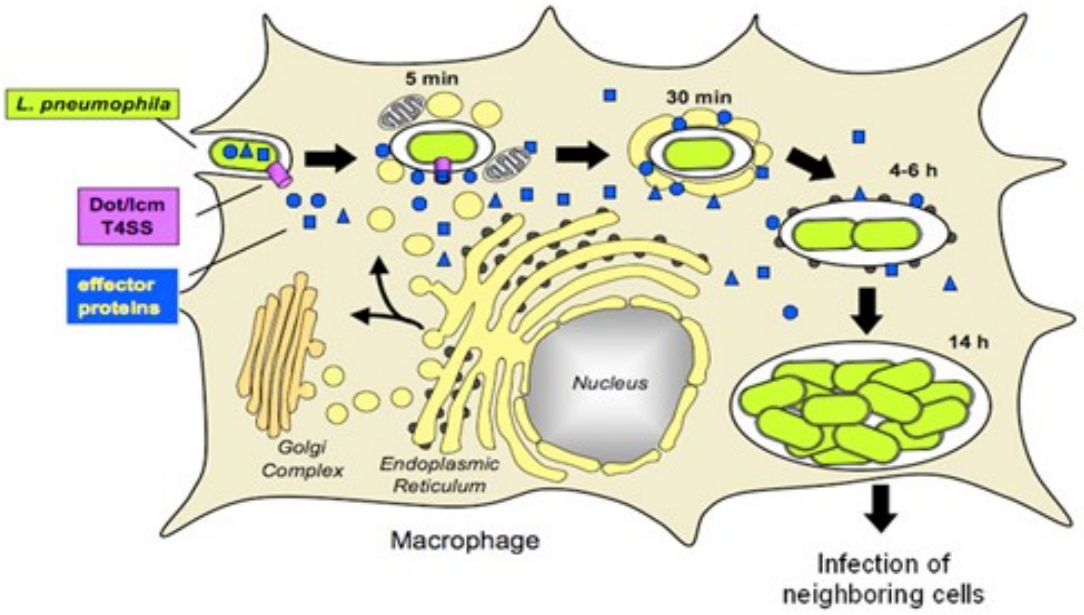


Figure 4 Illustration of *Legionella pneumophila* infection cycle in macrophages. Upon phagocytosis by macrophages, *L. pneumophila* secretes effector proteins that recruit and bind to different targets such as organelles (mitochondria, endoplasmic reticulum-Golgi vesicles), and lipids including phosphoinositides. These targets allow for the establishment of the *Legionella*-containing vacuole, where it replicates freely until egressing out of the host cell[15].

1.3 Phosphoinositides

1.3.1 Phosphoinositides as crucial regulators of eukaryotic cell membrane trafficking

L. pneumophila's ability to manipulate host cell trafficking is thanks in large parts due to its secretion of effector proteins that recruit and bind to phosphoinositides (PIs)[16]. PIs are phospholipids that comprise a tiny percentage of a eukaryotic cell's membrane yet have an immense role in vesicular trafficking and defining organelle identity[16]. PIs form a general class of phospholipids, but the key base moiety of PIs are phosphatidylinositols (PtdIns)[16]. PtdIns consist of diacylglycerol (DAG), which is a

glyceride consisting of two fatty acid chains covalently bonded to a glycerol molecule[16]. PtdIns also contain a D-myo-inositol 1-phosphate, which can be reversibly phosphorylated by organelle specific PI phosphatases and PI kinases at positions 3, 4 and/ or 5[16]. This enables seven classes of PtdIns to exist each with a unique mono- or poly- phosphorylated derivatives[16].

PIs recruit peripheral membrane proteins jointly with small GTPases in their active GDP-bound form[17]. These peripheral membrane proteins have specific PIbinding domains and thus this lipid-protein interaction along with specific adaptor proteins determines vesicular trafficking pathways and organelle identity[17]. Different subclasses of PIs localize to different subcellular organelles and pathways[17]. Some PtdIns are enriched at the plasma membrane, Golgi apparatus, or endoplasmic reticulum, while others accumulate in the endocytic pathway[17]. The enrichment of different PIs at different pathways and organelles defines these cellular components to enable correct vesicular trafficking[17].

The PI-metabolizing phosphatases and kinases travel to the cytoplasmic side of the plasma membrane by small GTPases[18]. These small GTPases activate the PImodulating enzymes and can determine where the PtdIns become phosphorylated and thus control which PtdIns are established[18]. The small GTPases themselves are activated by guanine nucleotide exchange factors (GEFs) and deactivated by GTPase activating proteins (GAPs)[18].

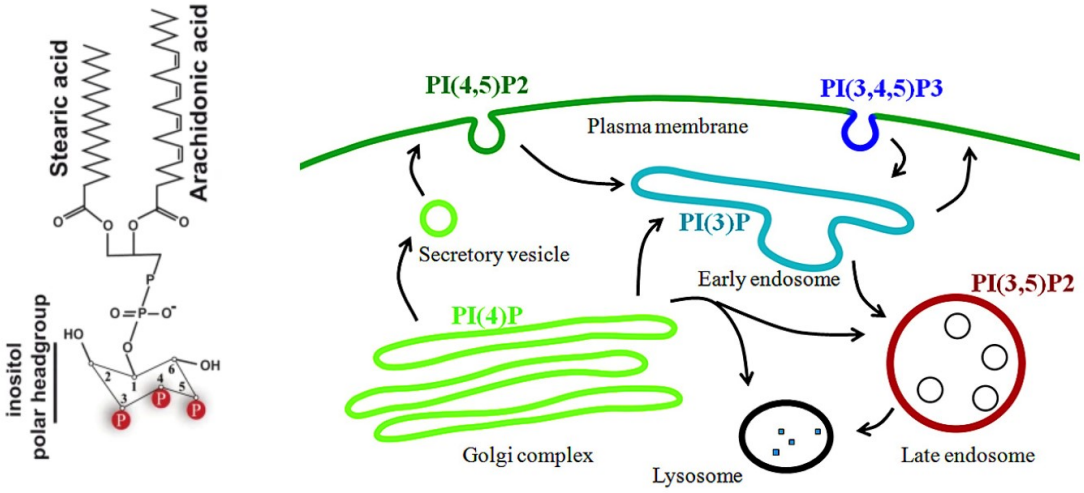


Figure 5 Phosphoinositides as crucial lipids for organelle identity. **(Left)** A phosphatidylinositol highlighting in red where PtdIns can be reversibly phosphorylated[19]. **(Right)** Model of cell showing which cellular compartments are enriched by which PI species[20].

1.3.2 Phosphoinositides in *L. pneumophila* infection

L. pneumophila does not possess PIPs in its biochemical composition, however, *L. pneumophila* effector proteins have the ability to modify PIPs at the LCV membrane either directly or by recruiting host PIP metabolizing enzymes[21]. PtdIns is crucial for not only *L. pneumophila's* establishment of the LCV, but also its manipulation of the host cell's classical endocytic maturation pathway[21]. Within minutes after *L. pneumophila's* uptake into a mammalian macrophage, PtdIns begin to accumulate on the LCV, specifically PtdIns(3)P[21]. These PIs usually are found on the ER membrane and on early and late endosomes helping to define endosome identity[21]. However, within around 2 hours of *L. pneumophila* uptake, the LCV loses this PI with PtdIns(4)P steadily accumulating on the LCV[21]. PtdIns(4)P are enriched on the LCV

for a prolonged period until *L. pneumophila* eventually exits the host cell[21]. PtdIns(4)P usually decorate the ER, Golgi, and secretory vesicles transporting proteins between the ER and Golgi[21]. PtdIns(4)P plays a crucial role in defining transporting vesicles as well as the ER and Golgi[21]. Taken together, the LCV undergoes a PI conversion taking the PtdIns(3)P that initially decorated the LCV and converting them to PtdIns(4)P[21]. This enables the LCV to change the macrophage defining the LCV as an endosome to defining the LCV as part of its natural host vesicular pathways[21].

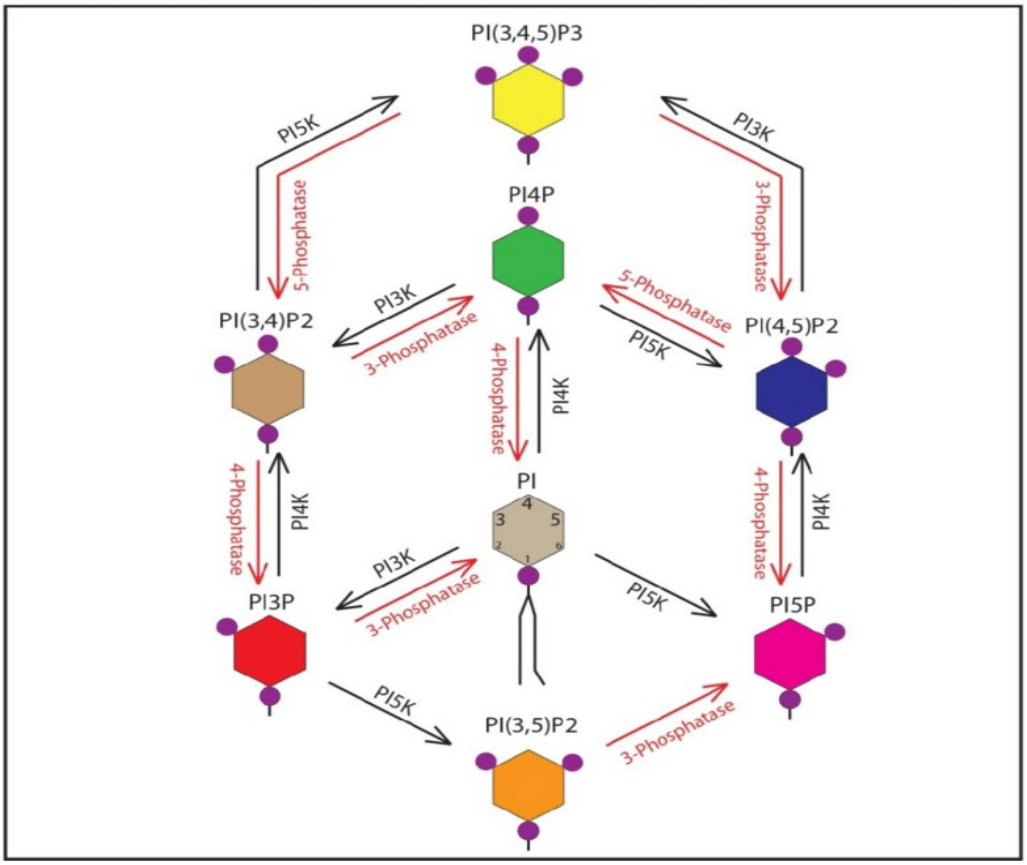


Figure 6 Phosphoinositide metabolizing schematic. Shows initial PI species and seven adapted PIP species along with which phosphatases and kinases modulate each PIP species[22]. PI3K, PI4K, and PI5K are kinases that phosphorylate the respective position of the inositol ring. 3-phosphate,

4phosphate, and 5-phosphate are phosphatases that dephosphorylate that respective position on the inositol ring[22].

1.3.3 Importance of identifying and characterizing *L. pneumophila* PI-binding proteins

L. pneumophila's secreted army of effector proteins allow the interaction with and subversion of host cell PIs[21]. Effector proteins can subvert PIs in a number of ways including direct binding to PIs, acting as bacterial PI-metabolizing enzymes (i.e. phosphatases, kinases, or phospholipases), or by recruiting host cell PI-metabolizing enzymes[21]. We envision that therapeutics can be developed that specifically target and inhibit *L. pneumophila* PI-binding effector proteins, but do not disrupt host cell PIprotein interactions[21]. Therefore, our goal is to identify which *L. pneumophila* effector proteins bind to PIs and characterize the function of these effector proteins during host infection[21].

1.4 Generation of three HaloTag fusion constructs

1.4.1 Introduction to the HaloTag system

Fluorescence microscopy is a prominent modality to visualize cell structure and dynamics[22]. Traditionally, naturally existing fluorescent proteins are utilized that emit specific emissions of light under certain conditions[22]. However, limitations exist with fluorescent proteins including low brightness and photostability, pH sensitivity, and an inability to optimize the biochemical properties of these proteins easily[22]. Therefore, a synthetic fluorophore offers numerous advantages and one such synthetic

fluorophore is HaloTag[22]. HaloTag, a 33 kDa bacterial enzyme, catalyzes the conversion of haloalkanes to hydroxy alkanes[22]. The HaloTag ligand, can be conjugated to a fluorophore and consists of a chlorohexane and two ethylene glycol motifs[22]. The active site of the HaloTag has been genetically modified and chloroalkane linker substrate (Halo ligand) can irreversibly (covalently) bind to the active site. Bound Halo ligands will fluoresce, while unbound HaloTag will not[22]. The HaloTag protein can be fused to the N- or C-terminus of a protein of interest and the tagged protein, using fluorescent Halo ligands, can thus be tracked in a cell with fluorescence microscopy[22]. HaloTag also offers the advantage of labeling translocated effector proteins, unlike GFP (Green fluorescent protein) or other fluorescent proteins, which are two large to properly pass-through Legionella type IV secretion system[22]. Labeling effector proteins that are translocated into the cell by *L. pneumophila*, presents the opportunity of live cell imaging to visualize highly dynamic processes in a cell[23].

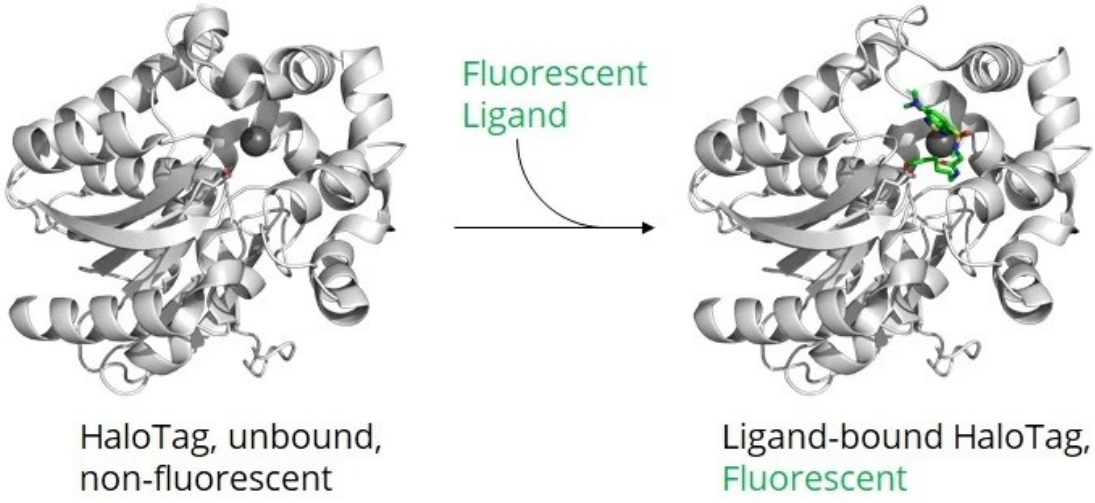


Figure 8 Bound HaloTag ligand fluoresces where unbound HaloTag ligand does not. When HaloTag is present in the cell, it needs to be bound to the ligand to fluoresce. HaloTag catalyzes the conversion of haloalkanes to hydroxy alkanes and the Chloroalkane linker substrates (ligand) are irreversibly bound to the active site. Once catalyzed the fluorophore can fluoresce under certain conditions, thus marking the protein of interest[24].

1.4.2 Preliminary results of Lpg1884, Lpg1958,and Lpg2511

The Neunuebel lab has discovered a number of PIP-binding effector protein candidates using a pull down assay followed by mass spectrometry . My objective was to follow up on three effector proteins of interest: Lpg1884, Lpg1958,and Lpg2511. After further testing using a protein-lipid overlay assay (also known as a PIP strip assay), these three effector proteins were validated to indeed bind phosphoinositides. However, while these proteins were confirmed to be bona fide PIP-binding proteins, their function remained to be characterized.

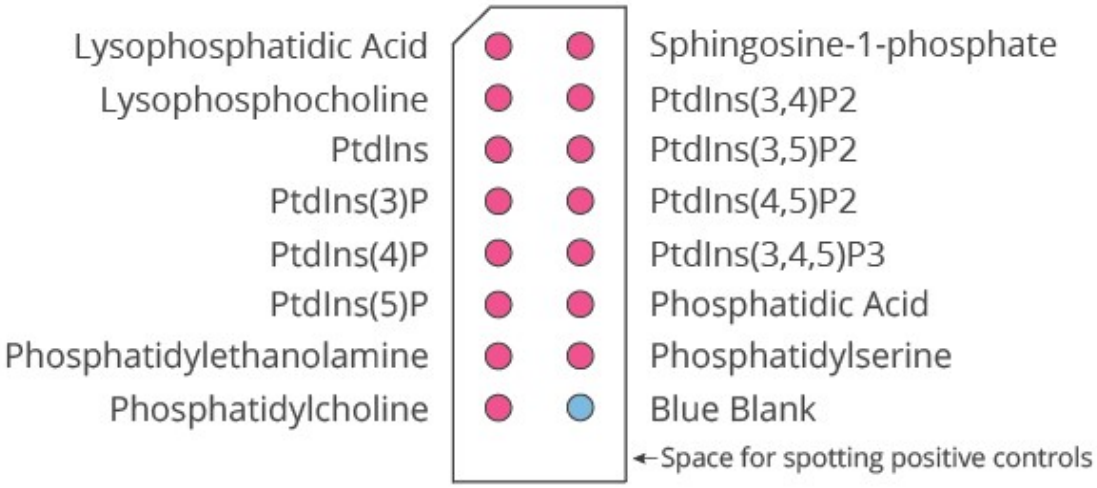


Figure 9 Protein lipid overlay assay. This assay allows for proteinphosphoinositide interaction and detects which phosphoinositide species the protein of interest can bind to[25].

1.4.3 Generating HaloTag fusions with *L. pneumophila* effectors Lpg1884, Lpg1958,and Lpg2511

We are focused on not only identifying novel phosphoinositide binding effector proteins but understanding the functional importance of these proteins. Relevant aspects we aim to determine regarding individual effectors are their timing of secretion during host infection, the cellular destination in the host cell, as well as discovering their specific biochemical targets. Because HaloTag opens the ability to image live cells, allowing visualization of dynamic interactions in the cell, our goal is to ultimately create a HaloTag library of *L. pneumophila* effector proteins. Towards this end, my objective was focused on the generation of three HaloTag fusion with *lpg1884*, *lpg1958*, and *lpg2511* that would contribute to this library. I attached the HaloTag protein to the N-terminal end of these three proteins using Gateway Cloning and confirmed cloning and expression of these tagged constructs with PCR, gel electrophoresis, and western blotting. These three constructs can thus be transfected or infected into a live cell to view the dynamic interactions that take place during *L. pneumophila* infection.

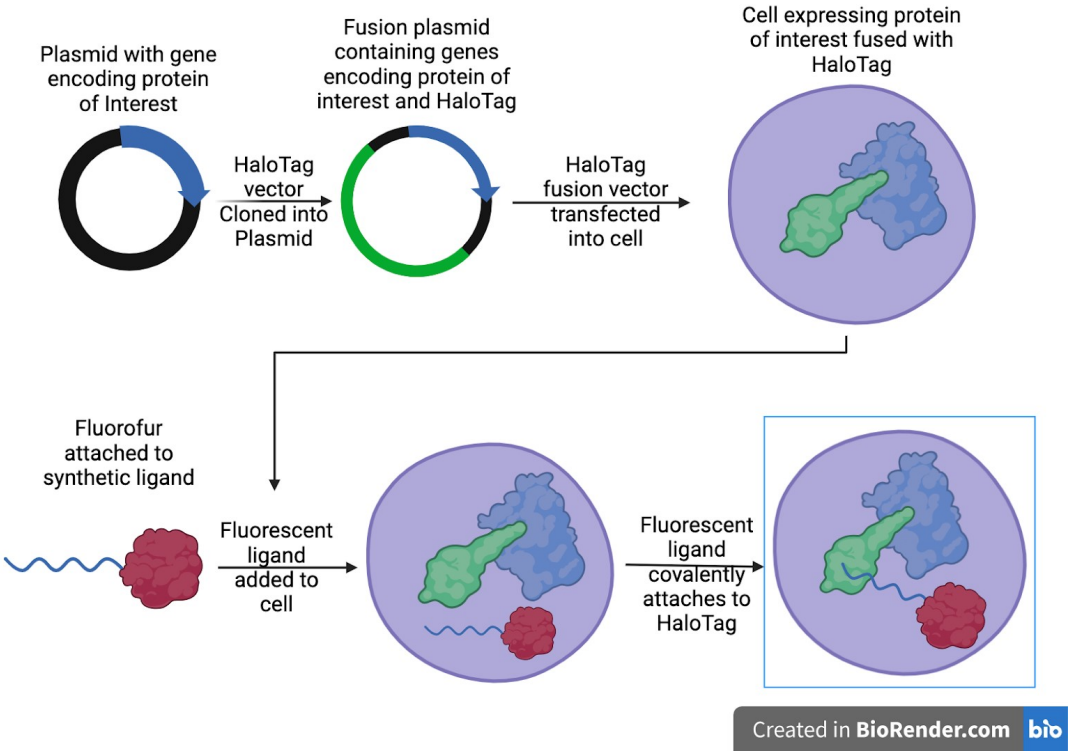


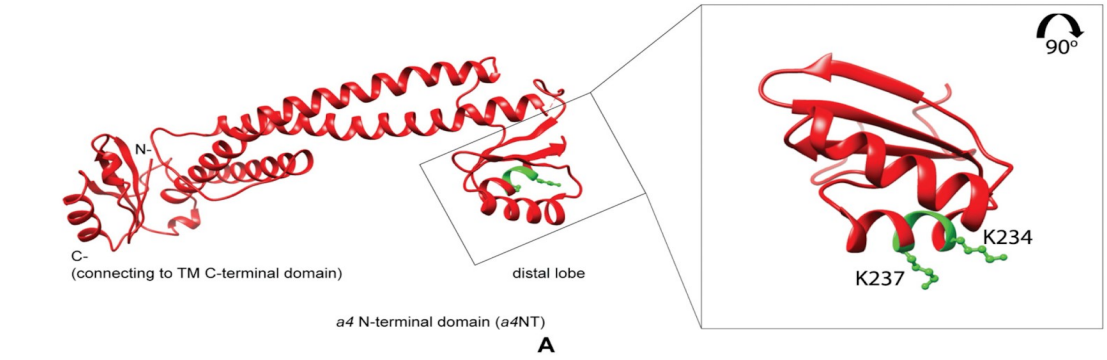
Figure 10 HaloTag fluorescence schematic. A protein of interest is first cloned into a gateway destination vector using gateway cloning. This fusion plasmid is then transfected and expressed in a mammalian cell generating the protein of interest fused with the HaloTag protein. A synthetic ligand, which has a fluorophore attached to it, is added to the mammalian cell where the synthetic ligand covalently binds to the HaloTag protein enabling marking of the protein of interest.

1.5 Identification of PIP-binding regions

1.5.1 Why we want to identify PIP-binding regions

One long-term goal of our lab is to identify how and why *L. pneumophila* effector proteins bind to phosphoinositides and how these proteins could be different from eukaryotic phosphoinositide binding proteins. If a therapeutic was developed targeting all phosphoinositide binding proteins the host cell's vesicular trafficking

pathways would also be altered and could possibly cause more damage than even *L. pneumophila* infection. Identifying what makes *L. pneumophila* PIP-binding effector proteins different might allow a more specific therapeutic to be produced and help to prevent the establishment of the LCV during *L. pneumophila* without inhibiting the host cell's normal trafficking pathways. However, most proteins have long stretches of amino acids and therefore these proteins have multiple domains with multiple functions[26]. It proves difficult to compare PIP-binding proteins without first identifying which region of a protein is responsible for the protein's binding to a phosphoinositide.



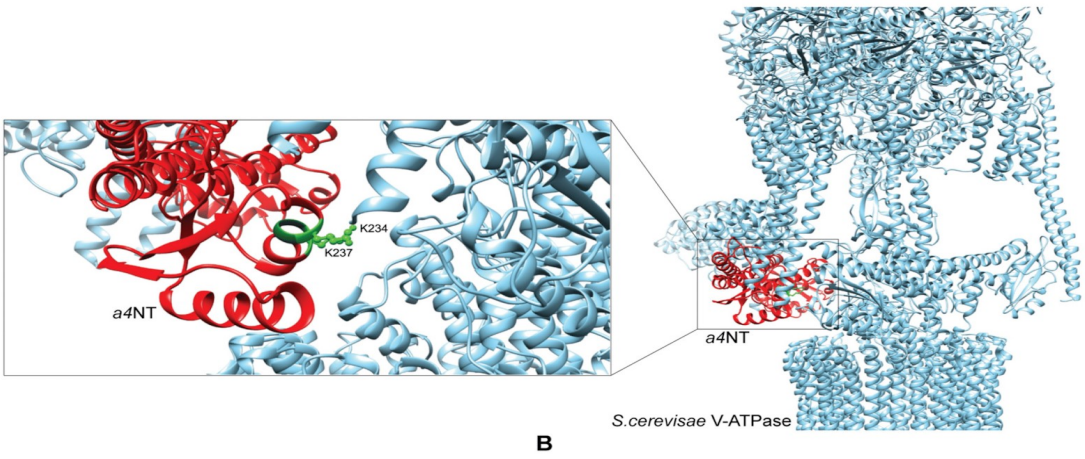


Figure 11 Characterization of a PIP Binding Site in the N-Terminal Domain of VATPase a4. The PIP-binding site is highlighted in green. The characterization and identification of this PIP-binding site proved critical to understanding V-ATPase a4 and its role in plasma membrane association[27].

1.5.2 How PIP-binding regions have previously been identified

Bioinformatics has previously been deployed to help elucidate not only possible phosphoinositide binding proteins, but also characterize effector protein functions overall. Other studies used *Legionella* genomes from different legionella species to identify common effectors present in all species[28]. This type of analysis allowed for identification of new conserved effector domains and described previously unknown effector functions[28]. While this study provided a good roadmap on how to analyze the genomes, this did not elucidate specific PIP-binding regions and these regions may not be well conserved across species. Another study used *Legionella* effector domains, which are conserved across *Legionella* species, but not in other known proteins, to see if these LEDs function as PI-binding domains[29]. Three LEDs were identified, which are present in a dozen *L. pneumophila* effectors and in more than 200 predicted effectors in the 41 Legionella species examined suggesting the LEDs are responsible for PIP-binding[29]. However, these studies examine whole Legionella species genomes to characterize function and identify PIP-binding domains, where we started with experimental data.

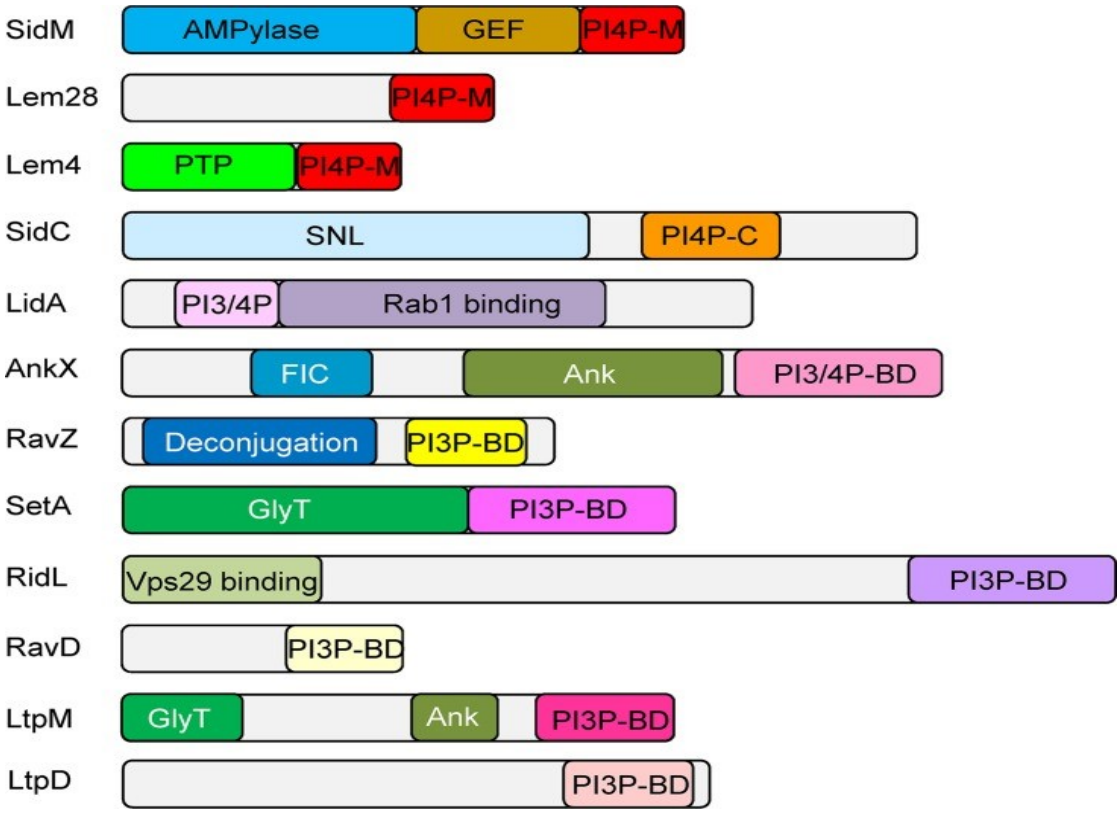


Figure 12 Domain architecture of *L. pneumophila* effectors that bind PIs. Each effector protein has a schematic attached to it showing the validated domain that binds a PI-species. A domain that displays X-BD is a binding domain, where other colored regions represent other functions. Demonstrates how effector proteins are split into domains[26].

1.5.3 How we are identifying PIP-binding regions

As previously mentioned, The Neunuebel lab has discovered several PIPbinding effector protein candidates using a pull-down assay followed by mass spectrometry. Many of these effector proteins had already been identified to be PIPbinders, but 28 of these effector proteins had not been identified as having PIP-binding domains. We aimed to determine what common features of protein structure or protein sequence made these proteins similar and whether we could predict the PIP-

binding region of each of these proteins. While other bioinformatic studies focused on deriving PIP-binding proteins from genome wide data, we focused on experimentally found PIP-binding proteins. We aimed to discover what part of these proteins made them responsible for PI-binding and elucidate why other bioinformatic techniques might have missed these effector proteins containing PIP-binding domains.

Chapter 2

METHODS

2.1 Strains and Media

The two model organisms used in this project were *Escherichia coli* and *Legionella pneumophila*. Construction of the expression plasmids was performed in *E.coli* in One Shot™ ccdB Survival™ 2 T1R Competent Cells (from Thermo fisher), where validation of protein expression was performed in *L. pneumophila* strains Lp02 and Lp03 initially derived from the National Institute of Health. *E.coli* containing plasmid of interest was plated on LB plates with 0.1% ampicillin grown in liquid cultures with LB broth containing ampicillin. LB broth is a mixture of yeast extract, tryptone, and NaCl in a 1:2:2 ratio respectively in 1 liter of water, where LB plates is a mixture of yeast extract, tryptone, NaCl, and agar in 1:2:2:3 ratio respectively in 1 liter of water. pDONOR plasmids with *Legionella* effector encoding gene or destination vector encoding HaloTag encoding gene were previously constructed by other members of the Neunuebel lab. From glycerol stocks, plasmids were plated on LB + ampicillin plates and stored at 37 C overnight. Following a colony would be grown in 3 mL of LB broth and 3 microliters of ampicillin and stored overnight in a shaking incubator at 37 C. Subsequently, all plasmid DNA was isolated with the NucleoSpin® Plasmid / Plasmid (NoLid): Isolation of high-copy plasmid DNA from *E. coli* protocol. Two strains of *L. pneumophila* were used with the Lp02 strain acting as our experimental strain and the Lp03 strain acting as our control as its translocation deficient cannot

secrete effector proteins. Lp02 and Lp03 are both thymidine auxotroph's meaning they need the additional supplement of thymidine to have the proper nutrients to replicate. However, the HaloTag gene codes for the expression of thymidine so once the HaloTag plasmid is transformed into either *L. pneumophila* strain, both strains no longer need the addition of thymidine to replicate. Lp02 and Lp03 without the HaloTag plasmid are plated on CYET plates, which are made by combining yeast extract, ACES (N-(2-acetamido)-2-aminoethanesulfonic acid), charcoal, and agar, in a 5:5:1:9 ratio respectively in 1 liter of water. This solution then has KOH added until the pH is 6.9. After heating and cooling, the supplements ferric nitrate, cysteine, and thymidine are added in a 1:1:1 ratio. These plates are stored in an incubator at 37 C for 3-4 days. After colonies appear, *L. pneumophila* patches were then streaked onto CYET plates, and stored in an incubator at 37 C for two days. *L. pneumophila* strains were subsequently grown in liquid cultures in AYE broth with appropriate supplements. AYE broth is composed of yeast extract and ACES in 1:1 ratio in 1 liter of water and pH is brought to 6.9 with KOH. *L. pneumophila* is gown in a liquid culture through a serial dilution where a *L. pneumophila* patch is resuspended across 5 3mL solutions of AYE media combined a 1:1:1 ratio of ferric nitrate, cysteine, and thymidine. This serial dilution is stored in the shaking incubator for 18-20 hours at 37 C. An *E.coli* plasmid with the HaloTag vector can be transformed into the resulting solution from the serial dilution with an OD600 between 0.8 and 1.2. Growing *L. pneumophila* with the HaloTag vector already transformed into requires the same protocol except not adding thymidine as a supplement to the CYE plates or the AYE

media.

2.2 Construction of Expression Plasmids

All the plasmids used in this project were generated using the Gateway Cloning Technology. Previous members of the Neunuebel lab generated all entry clones used in this project. To generate an entry clone of a gene of interest, which encoded a legionella effector, a BP reaction would occur. Here *attB* sites flanking each end of the gene of interest were cloned into *ccdB* 2 T1R *Escherichia Coli* Competent Cells. This *E.coli* strain is resistant to the product of the *ccdB* gene, which is toxic for other *E.coli* strains. This BP reaction would result in the *ccdB* gene being replaced by an *L. pneumophila* effector-encoding gene. The plasmids carrying the gene of interest were gateway cloned into pJB908-HaloTag destination plasmids using an LR reaction. Here an entry vector (carrying the gene of interest) could be combined with a destination vector (carrying the *ccdB* gene) to generate an expression vector carrying the gene of interest with the desired destination plasmid. 150 nanograms of a donor vector was resuspended with 150 nanograms of a destination plasmid with 2 microliters of proprietary LR clonase and enough LR reaction buffer to create a 10-microliter solution. This solution was incubated for 1 hour at 25 C. Subsequently 1 microliter of proteinase K was resuspended in the solution and then incubated for 10 minutes at 37 C. The expression clone was subsequently transformed into competent *E.coli* 2T 1 cell. 5 microliters of expression plasmid DNA were added to the Eppendorf tube of competent cells. Competent cells were kept on ice for 30 minutes, heat shocked at 42

C for 30 seconds, and put back on ice for 2 minutes. Following 250 microliters of SOC broth was added to the competent cells. Competent cells were recovered at 37 C in the Thermoshaker for 90 minutes. Competent cells were plated on LB + ampicillin and stored at 37 C overnight. Following verification of expression vectors with colony PCR, the expression vector was transformed into a *L. pneumophila* strain. 1 mL of liquid culture containing *L. pneumophila* strain at an OD600 between 0.8 and 1.2 was spun down in a centrifuge for 1 minute at 11,000 g at 4 C. The resulting pellet was washed with 500 microliters of 10% glycerol a total of three times. The final pellet was resuspended in 50 microliters of 10% glycerol. 2,000 nanograms of the expression plasmid DNA was then resuspended into the resulting solution. The resulting solution was electroporated and then dispensed into 1 mL of AYE broth. This suspension was transferred into an Eppendorf tube placed in the Thermoshaker at 37 C for 2 hours and 30 minutes. This suspension was spun down in a centrifuge at 11,000 g for 30 seconds and 800 microliters of supernatant was removed from the resulting suspension. The resulting supernatant was resuspended with the pellet and plated on a CYE plate with the appropriate supplements. The plate was stored in an incubator at 37 C for approximately four days. The resulting colonies were *L. pneumophila* cells expressing the desired expression plasmid. These plasmids had to then be verified with Western blotting to confirm correct protein expression.

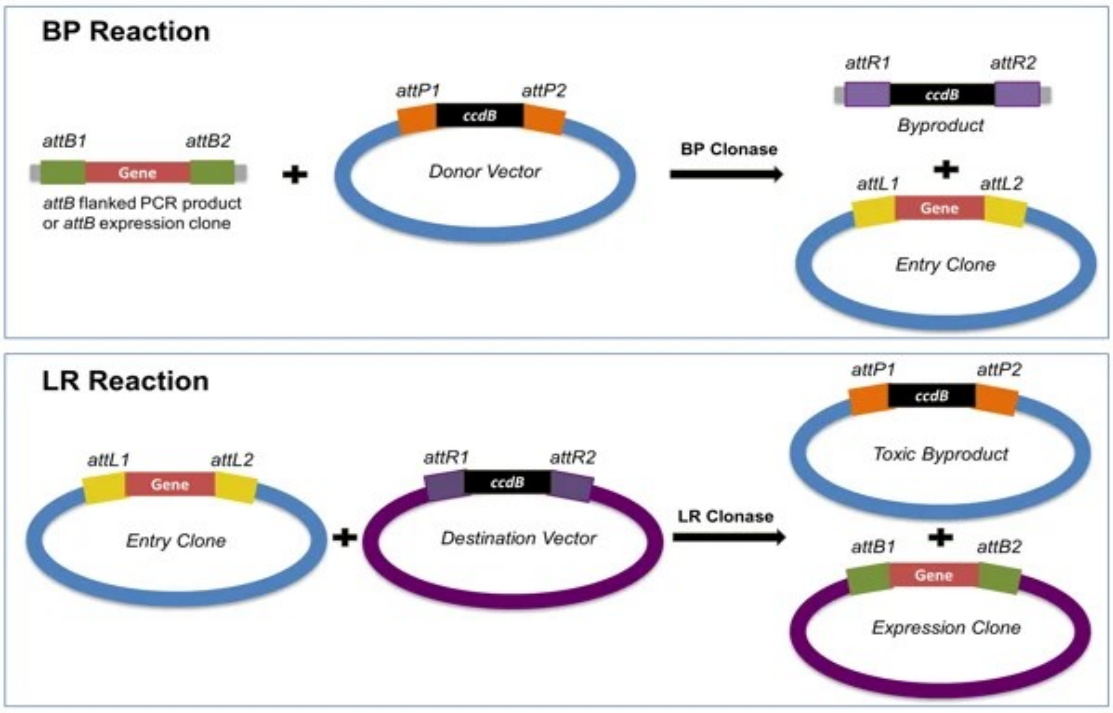


Figure 13 The Gateway system adopts phage integration into the BP and LR reactions. The BP reaction creates an attL-flanked entry clone. The LR reaction creates an expression clone with all the components necessary for gene expression[30].

2.3 Colony PCR

After a plate was incubated overnight with colonies hosting plasmids with a gene of interest, one colony was obtained with a sterile loop and resuspended in 30 microliters of sterile water. Subsequently, the same loop was used to inoculate another plate. The suspension was then put in heat shock for 10 minutes at 99 C. Afterwards, the suspension was centrifuged at 14.8 g at 4 C for 10 minutes. 0.8 microliters of the resulting suspension were used as the template DNA for the PCR. This template DNA was resuspended with 0.2 microliters of vector specific forward primer, 0.2 microliters

of vector specific reverse primer, 3.8 microliters of water, and 5 microliters of Thermo Scientific™ PCR Master Mix, which includes Taq polymerase, reaction buffer, magnesium chloride, and dNTPs. This suspension was subsequently loaded into the Thermocycler with the following parameters: (95°C for 3 min, 95°C for 30 s, 53°C for 30 s (annealing phase), 72°C for 1-2 min, 72°C for 5 min). PCR products were separated by agarose gel electrophoresis on 0.7% agarose gel.

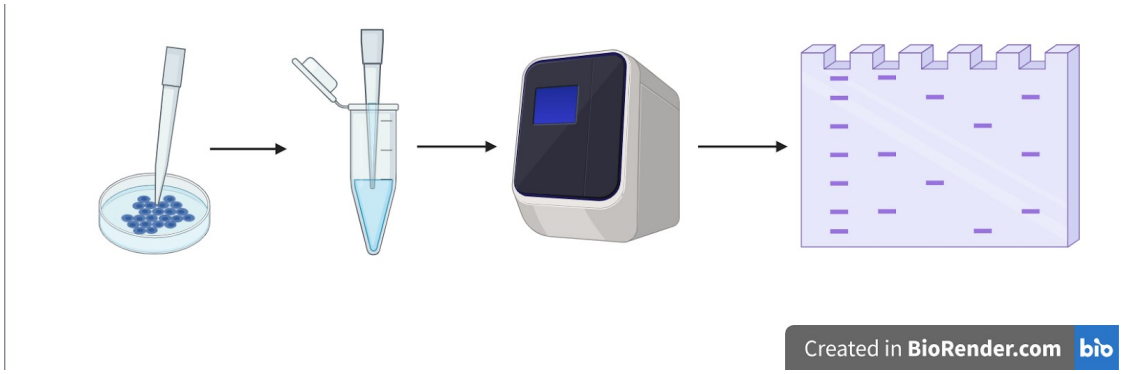


Figure 14 General workflow of colony PCR. After incubation of colonies hosting plasmid containing gene of interest, a colony is resuspended in 0.2 microliters of vector specific forward primer, 0.2 microliters of vector specific reverse primer, 3.8 microliters of water, and 5 microliters of Thermo Scientific™ PCR Master Mix. The resuspension is added to the Thermocycler with the appropriate parameters and products were separated with agarose gel electrophoresis.

2.4 Western Blot

After the legionella-effector encoding gene has been correctly cloned into the HaloTag destination plasmid, and this plasmid has been verified to be the correct size after being transformed into *L. pneumophila*, the protein that should be overexpressed by *L. pneumophila* needs to be verified through western blotting. To run this *L. pneumophila* expression test, a serial dilution was created with an *L. pneumophila* strain containing a colony PCR verified legionella-effector fused HaloTag construct. The dilution with OD600 of 0.8-1.2 inoculated a 5mL solution of AYE with appropriate supplements to an OD600 of 0.3. The inoculated culture was stored in the shaking incubator at 37 C for about 4 hours until the OD600 was between 0.6-1.0. 500 microliters of this uninduced sample was saved through spinning down the culture at 11,000 g for 1 minute, removing the supernatant, and saving the cell pellet in the freezer at -20 C. The remaining solution was induced with IPTG to a final concentration of 0.5mM and this culture was grown overnight at 37 C in the shaking incubator. The next morning the OD600 of the sample solutions were recorded and induced sample cells were saved equal to the uninduced sample. The induced cultures were spun down in the centrifuge at 11,000 g for 1 min, where the supernatant was removed, and the pellet was saved. The uninduced and induced pellets were resuspended in 50 microliters of a solution of PBS (Phosphate buffered saline), 4x

Laemmli sample buffer (from BIO-RAD), and BME (2-Mercaptoethanol)(from BIORAD) in a 27:9:1 ratio. After resuspension, samples were boiled at 99 C for 10 minutes and spun down in a centrifuge at 17,000 g (confirm this is max speed). The samples were loaded onto an SDS page-gel (insert link for how to make SDS page gel). The gel was run at 150 volts for 20 minutes and then 200 volts for 20 minutes. The resulting protein on the gel was then transferred to a membrane (say membrane name) using the Trans-Blot Turbo Transfer System. The membrane was blocked in 5% milk for 1 hour on a rocking platform. After removing the milk, the membrane was incubated with the primary antibody (∝-HaloTag) in 1:1000 dilution with 5%milk for 1 hour on a rocking platform. The milk solution was removed from the membrane and then washed with PBST (Phosphate-Buffered saline with 0.1% Tween (Polysorbate 20)) for five minutes a total of three times. Secondary antibody (∝mouse) was added in a 1:20,000 dilution with 5% milk for 1 hour on a rocking platform. The milk solution was removed from the membrane and then washed with PBST for five minutes a total of three times. All liquid was removed from the membrane and PICO agents in a 1:1 ratio was added and to the membrane for 3 minutes. The reagents were then removed, and the western blot was imaged.

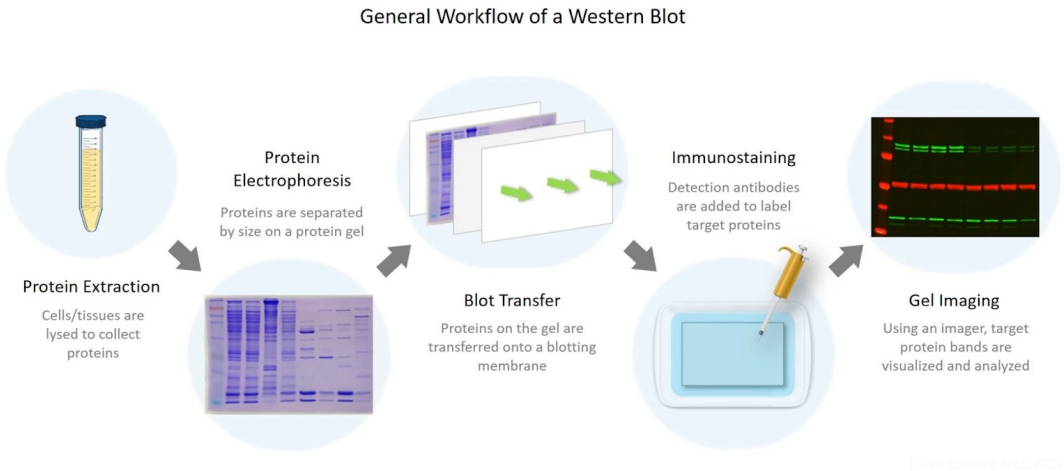


Figure 15 General workflow of a western blot. All proteins are first extracted from *L. pneumophila* overexpressing the protein of interest. Proteins are then separated by gel electrophoresis and transferred to a blotting membrane. Primary and secondary antibodies are added to label and detect the overexpressed protein and gel imaging is performed to visualize the target protein bands[31].

2.5 Bioinformatics

We first used the full length FASTA file (protein sequence code) of all 28 PIPbinding effector proteins identified by the Neunuebel lab to generate a maximum likelihood phylogenetic tree using the MEGA[32] software. We used this same technique to analyze the PIP-binding region of all the published PIP-binding proteins. To predict which domain or fragment of each PIP-binding protein was responsible for binding to phosphoinositides, we first ran an HHpred[33] prediction on all 28 effector proteins, which predicted domains for most proteins. We used HHpred in conjunction with the 3D predicted structure of each protein provided by AlphaFold[34] to help elucidate domains that could not be predicted with HHpred. Both AlphaFold and HHpred use gene homology to predict domains and 3D structure. Subsequently, we predicted the function of each of the fragments generated by utilizing the predicted functions of the domains provided by HHpred along with predictions provided by ProteInfer[35], which uses machine learning, neural networks, and gene ontology to generate protein predictions. We generated another maximum likelihood tree with every predicted domain using the MEGA software. We also ran a CONSURF[36] analysis on all 28 effector proteins, which projects conservation scores onto 3D predicted structures using homology analysis to other proteins.

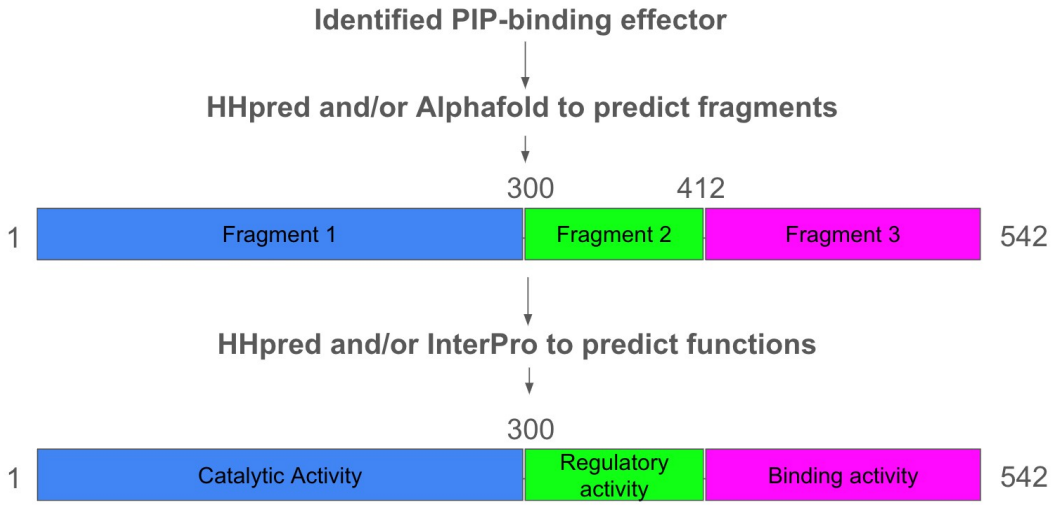


Figure 16 Schematic of how full-length PIP-binding effector proteins were separated into fragments with predicted function. HHpred and Alpha Fold predictions were run on the 28 effector proteins identified to be PIPbinding effector by the Neunuebel lab to separate the full-length protein into fragments with predicted domains. HHpred and Interpro were used to predict the functions of each of these fragments.

Chapter 3

RESULTS

3.1 Construction and Confirmation of Gateway Destination Plasmids

3.1.1 Construction of Gateway Destination Plasmids

To fuse the HaloTag gene with the three *L. pneumophila* effector encoding genes of interest, we successfully cloned 3 *L. pneumophila* effector encoding genes into the pJB908-HaloTag Destination plasmid using primers. PCR products were separated by gel electrophoresis on a 0.7% agarose gel. (Figure 17-18)

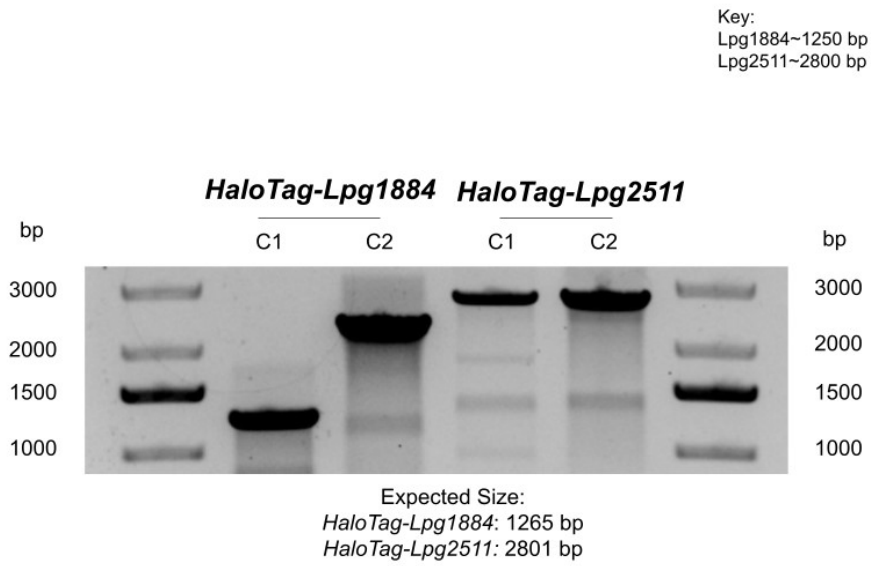


Figure 17 Colony PCR performed on 2T1 *E.coli* cells containing pJB908-HaloTag*lpg1884*, and pJB908-HaloTag-*lpg2511*. Lanes 2-3 represent two different colonies from the cultured cells containing pJB908-HaloTag*lpg1884*. Lanes 4-5 represent two different colonies from the cultured cells containing pJB908-HaloTag-*lpg2511*. The colony in lane 2 for pJB908-HaloTag-*lpg1884* has the appropriate band size considering the additional 50 bp added when the primers bind the regions upstream

and downstream of the gene (1265 bp). The colony in lane 3 for pJB908HaloTag-*lpg2511* has the incorrect band size. The colonies in lanes 4-5 for pJB908-HaloTag-*lpg2511* have the appropriate band size considering the additional 50 bp added when the primers bind the regions upstream and downstream of the gene (2801 bp). The faint bands most likely represent non-specific binding of the primers due to a low annealing temperature.

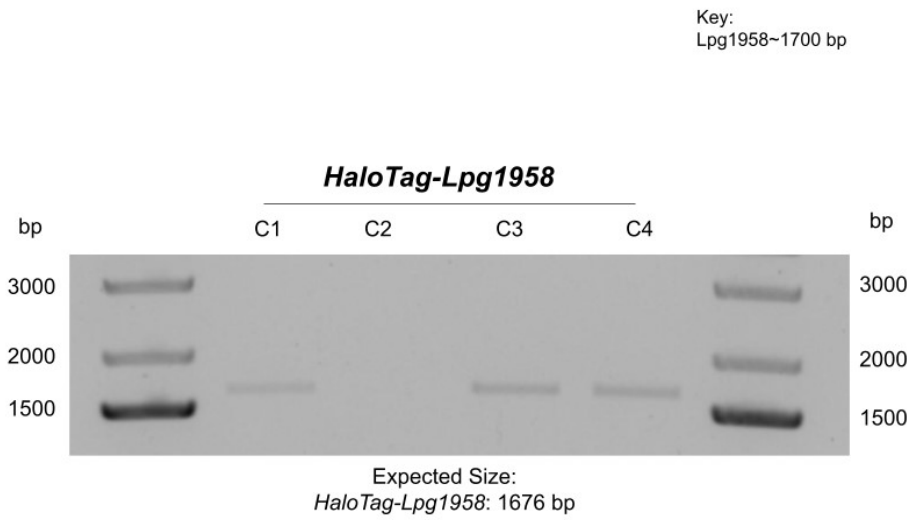


Figure 18 Colony PCR of 2T1 *E.coli* cells containing pJB908-HaloTag-*lpg1958*. Lanes 2-5 represent four different colonies from the cultured cells containing pJB908-HaloTag-*lpg1958*. The colonies in lanes 2-5 for pJB908-HaloTag-*lpg1958* have the appropriate band size considering the additional 50 bp added when the primers bind the regions upstream and downstream of the gene (1676 bp). The faint bands most likely represent non-specific binding of the primers due to a low annealing temperature.

3.1.2 Transformation of Constructs into *L. pneumophila*

To verify that the HaloTag fused *L. pneumophila* effector constructs were transformed correctly into both Lp02 and Lp03, colony PCR was run on *L.*

pneumophila colonies expressing this construct. 3 effector constructs were amplified in two different *L. pneumophila* strains using primers. PCR products were separated by gel electrophoresis on a 0.7% agarose gel. (Figures 19-24)

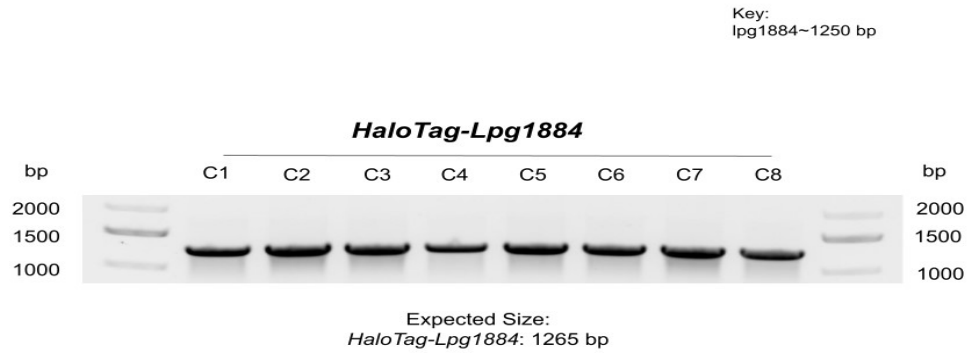


Figure 19 Colony PCR was performed on Lp02 overexpressing pJB908-HaloTag*lpg1884*. Lanes 2-9 represent 8 different colonies from the cultured cells overexpressing pJB908-HaloTag-*lpg1884*. The colonies in lanes 2-9 for pJB908-HaloTag-*lpg1884* have the appropriate band size considering the additional 50 bp added when the primers bind the regions upstream and downstream of the gene (1265 bp).

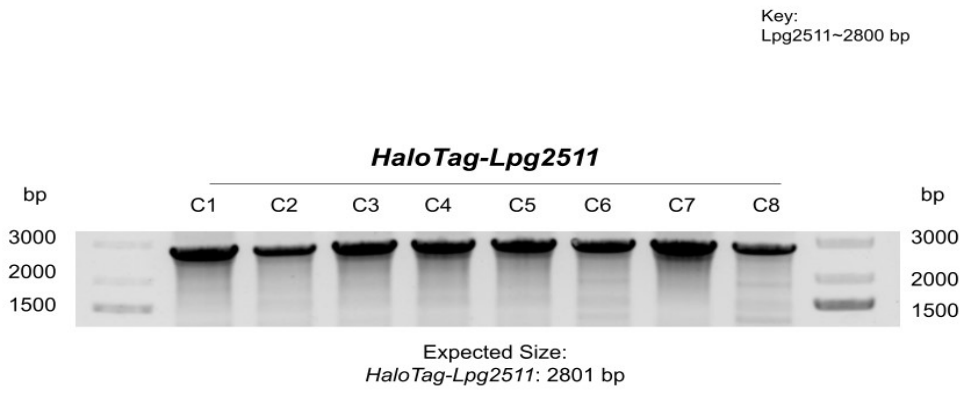


Figure 20 Colony PCR performed on Lp02 overexpressing pJB908-HaloTag*lpg2511*. Lanes 2-9 represent 8 different colonies from the cultured cells overexpressing pJB908-HaloTag-*lpg2511*. The colonies in lanes 2-9 for pJB908-HaloTag-*lpg2511* have the appropriate band size considering the additional 50 bp added when the primers bind the regions upstream and downstream of the gene (2801 bp). The faint bands most likely represent non-specific binding of the primers due to a low annealing temperature.

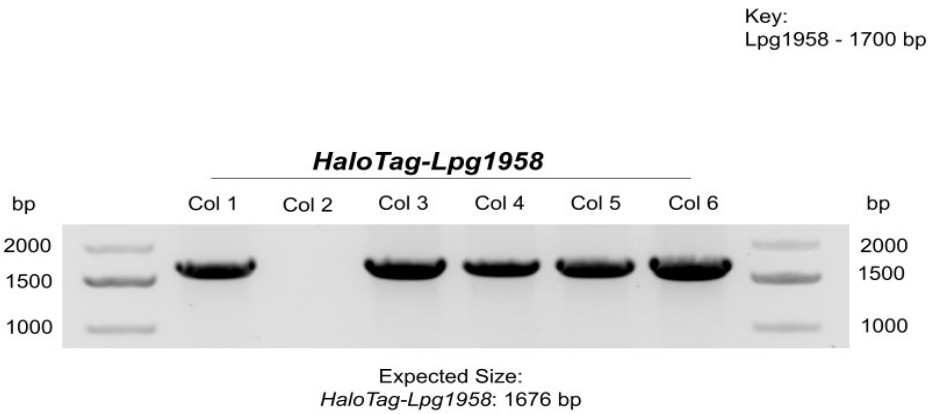


Figure 21 Colony PCR was performed on Lp02 overexpressing pJB908-HaloTag*lpg1958*. Lanes 2-7 represent 6 different colonies from the cultured cells overexpressing pJB908-HaloTag-*lpg1958*. The colonies in lanes 2 and 47 for pJB908-HaloTag-*lpg1958* have the appropriate band size considering the additional 50 bp added when the primers bind the regions upstream and downstream of the gene (2801 bp). The colonies in lane 3 for pJB908-HaloTag-*lpg1958* had no band indicating the primers did not attach to the gene correctly for this colony.

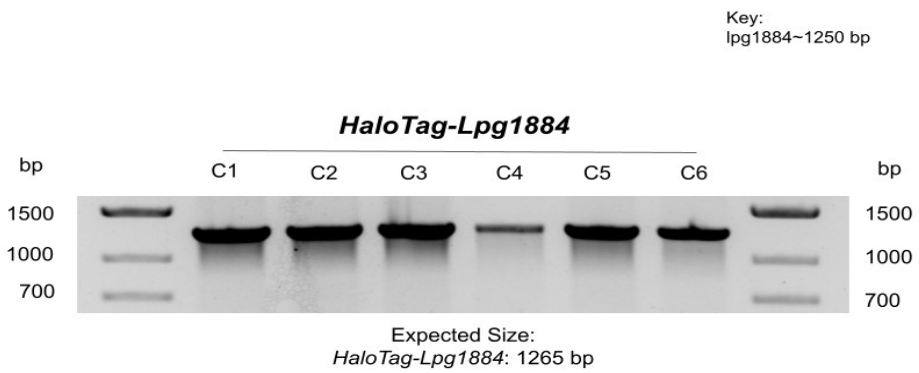


Figure 22 Colony PCR was performed on Lp03 overexpressing pJB908-HaloTag*lpg1884*. Lanes 2-7 represent 6 different colonies from the cultured cells overexpressing pJB908-HaloTag-*lpg1884*. The colonies in lanes 2-7 for pJB908-HaloTag-*lpg1884* have the appropriate band size considering the additional 50 bp added when the primers bind the regions upstream and downstream of the gene (1265 bp).

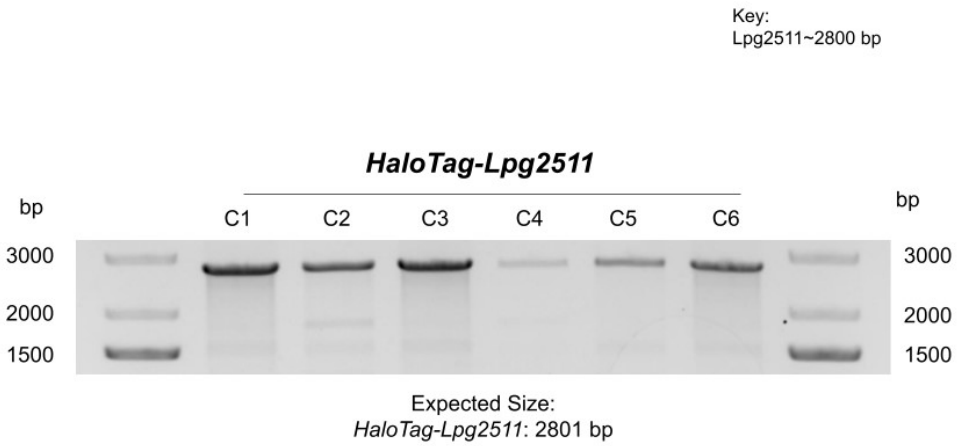


Figure 23 Colony PCR was performed on Lp03 overexpressing pJB908-HaloTag*lpg2511*. Lanes 2-7 represent 6 different colonies from the cultured cells overexpressing pJB908-HaloTag-*lpg2511*. The colonies in lanes 2-7 for pJB908-HaloTag-*lpg2511* have the appropriate band size considering the additional 50 bp added when the primers bind the regions upstream and downstream of the gene (2801 bp). The faint bands most likely represent non-specific binding of the primers due to a low annealing temperature.

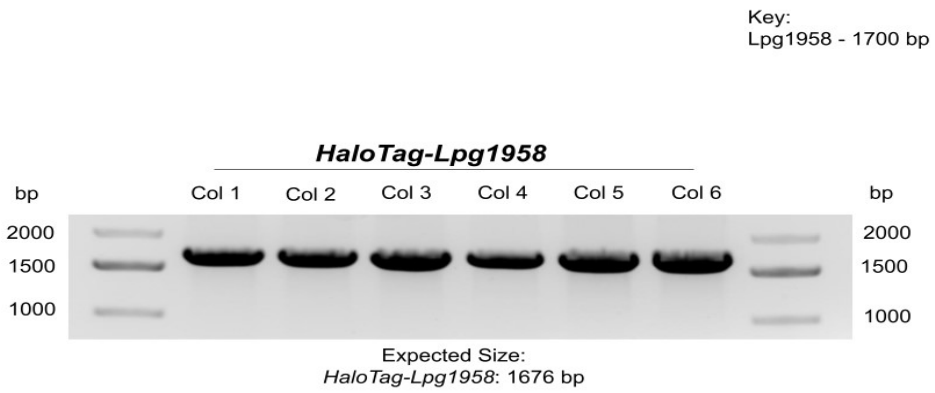


Figure 24 Colony PCR performed on Lp03 overexpressing pJB908-HaloTag*lpg1958*. Lanes 2-7 represent 6 different colonies from the cultured cells overexpressing pJB908-HaloTag-*lpg1958*. The colonies in lanes 2-7 for pJB908-HaloTag-*lpg1958* have the appropriate band size considering the additional 50 bp added when the primers bind the regions upstream and downstream of the gene (1676 bp).

3.1.3 Verification of correct protein expression from constructs

To verify that the HaloTag fused *L. pneumophila* effector constructs were being correctly expressed by both Lp02 and Lp03, western blotting was performed on Lp02 and Lp03 overexpressing the HaloTag fused *L. pneumophila* effector constructs. Protein expression for 3 effector constructs in two different *L. pneumophila* strains was verified. (Figures 12-17)

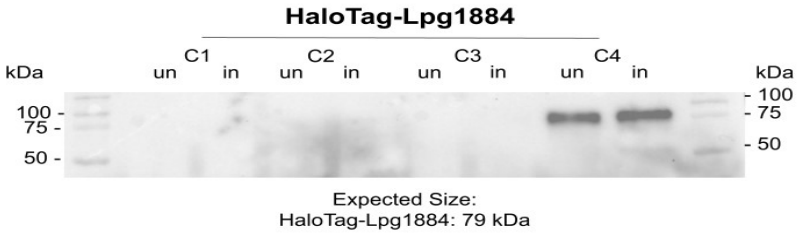


Figure 25 Western blot was performed on Lp02 overexpressing pJB908-HaloTagLpg1884. Lanes 2-3, 4-5, 6-7, and 8-9 represent 4 different colonies from the cultured cells overexpressing pJB908-HaloTag-lpg1884. Lanes 2, 4, 6, and 8 represent the uninduced colony, and lanes 3, 5, 7, and 9 represent the colony induced with IPTG. The concentration of the primary antibody used was a 1:1000 dilution of a mouse anti-HaloTag antibody. The concentration of the secondary antibody was a 1:20,000 dilution of an anti-Mouse IgG antibody. Lanes 2-7 show no bands indicating that for colonies 1-3 either the proteins were not transferred to the blot correctly or the primary or secondary antibody washed off the blot completely. The band size in lanes 8-9 for colony 4 show pJB908HaloTag-lpg1884 being correctly expressed in kda (79kDa).

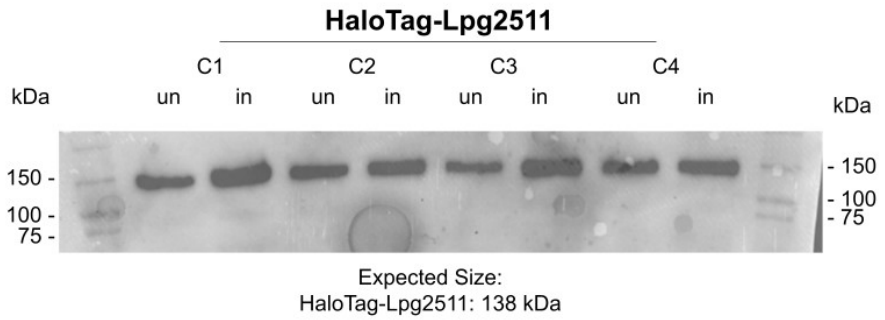


Figure 26 Western blot was performed on Lp02 overexpressing pJB908-HaloTaglpg2511. Lanes 2-3, 4-5, 6-7, and 8-9 represent 4 different colonies from the cultured cells overexpressing pJB908-HaloTag-lpg2511. Lanes 2, 4, 6, and 8 represent the uninduced colony, and lanes 3, 5, 7, and 9 represent the colony induced with IPTG. The concentration of the primary antibody used was a 1:1000 dilution of a mouse anti-HaloTag antibody. The concentration of the secondary antibody was a 1:20,000 dilution of an anti-Mouse IgG antibody. The band size in lanes 2-9 for colonies 1-4 show pJB908-HaloTag-lpg2511 being correctly expressed in kda (138kDa).

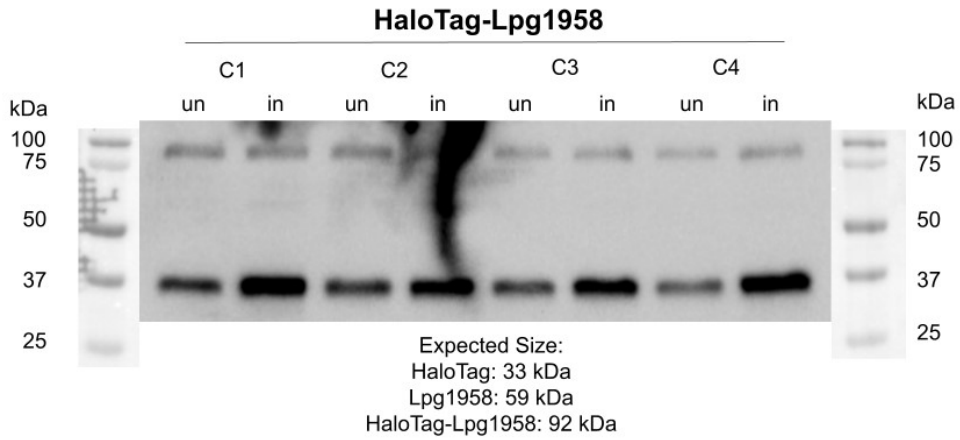


Figure 27 Western blot was performed on Lp02 overexpressing pJB908-HaloTaglpg1958. Lanes 2-3, 4-5, 6-7, and 8-9 represent 4 different colonies from the cultured cells overexpressing pJB908-HaloTag-lpg2511. Lanes 2, 4, 6, and 8 represent the uninduced colony, and lanes 3, 5, 7, and 9 represent the colony induced with IPTG. The concentration of the primary antibody used was a 1:1000 dilution of a mouse anti-HaloTag antibody. The concentration of the secondary antibody was a 1:20,000 dilution of an anti-Mouse IgG antibody. The fainter band size above the much darker bands in lanes 2-9 for colonies 1-4 show pJB908-HaloTaglpg2511 being correctly expressed in kda (92kDa). The darker bands represent the size in kDa of the HaloTag protein indicating that a large portion of the HaloTag protein became detached or fell off from lpg2511. The dark streak in the middle of the blot is excess reagent and water and does not represent any proteins.

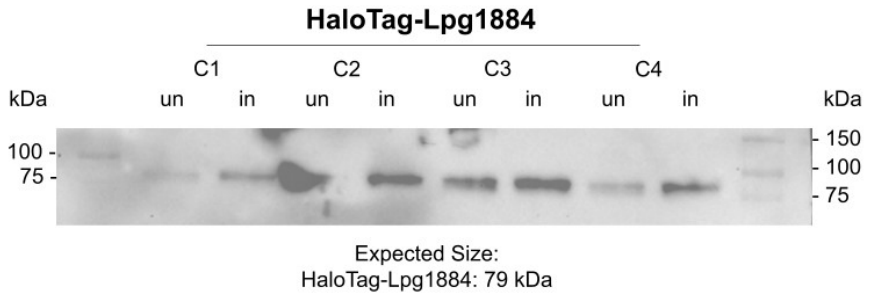


Figure 28 Western blot was performed on Lp03 overexpressing pJB908-HaloTaglpg1884. Lanes 2-3, 4-5, 6-7, and 8-9 represent 4 different colonies from the cultured cells overexpressing pJB908-HaloTag-lpg1884. Lanes 2, 4, 6, and 8 represent the uninduced colony, and lanes 3, 5, 7, and 9 represent the colony induced with IPTG. The concentration of the primary antibody used was a 1:1000 dilution of a mouse anti-HaloTag antibody. The concentration of the secondary antibody was a 1:20,000 dilution of an anti-Mouse IgG antibody. The band size in lanes 2-9 for colonies 1-4 show pJB908-HaloTag-lpg1884 being correctly expressed in kda (79kDa). Bands have different intensities due to more or less antiHaloTag antibody binding to the HaloTag protein.

Figure 29 Western blot was performed on Lp03 overexpressing pJB908-HaloTagLpg2511. Lanes 2-3, 4-5, 6-7, and 8-9 represent 4 different colonies from the cultured cells overexpressing pJB908-HaloTag-lpg2511. Lanes 2, 4, 6, and 8 represent the uninduced colony, and lanes 3, 5, 7, and 9 represent the colony induced with IPTG. The concentration of the primary antibody used was a 1:1000 dilution of a mouse anti-HaloTag antibody. The concentration of the secondary antibody was a 1:20,000 dilution of an anti-Mouse IgG antibody. The band size in lanes 2-9 for colonies 1-4 show pJB908-HaloTag-lpg2511 being correctly expressed in kda (138kDa). Bands have different intensities due to more or less antiHaloTag antibody binding to the HaloTag protein.

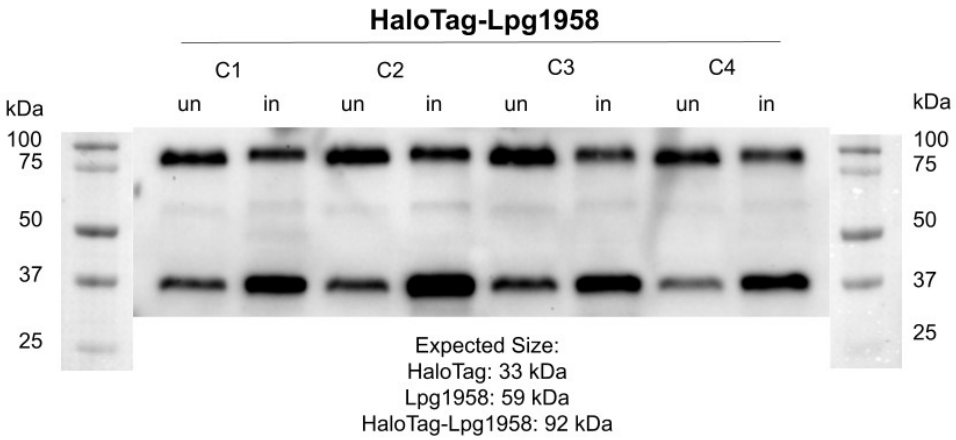


Figure 30 Western blot was performed on Lp03 overexpressing pJB908-HaloTaglpg1958. Lanes 2-3, 4-5, 6-7, and 8-9 represent 4 different colonies from the cultured cells overexpressing pJB908-HaloTag-lpg1958. Lanes 2, 4, 6, and 8 represent the uninduced colony, and lanes 3, 5, 7, and 9 represent the colony induced with IPTG. The concentration of the primary antibody used was a 1:1000 dilution of a mouse anti-HaloTag antibody. The concentration of the secondary antibody was a 1:20,000 dilution of an anti-Mouse IgG antibody. The bands size on the upper end of the blot above the bands on the lower end of the blot in lanes 2-9 for colonies 1-4 show pJB908-HaloTag-lpg1958 being correctly expressed in kda (92kDa). The lower bands represent the size in kDa of the HaloTag protein indicating that a large portion of the HaloTag protein became detached or fell off from lpg1958.

3.2 Analysis of PIP-binding domains

3.2.1 Phylogenetic trees of full-length PIP-binding proteins

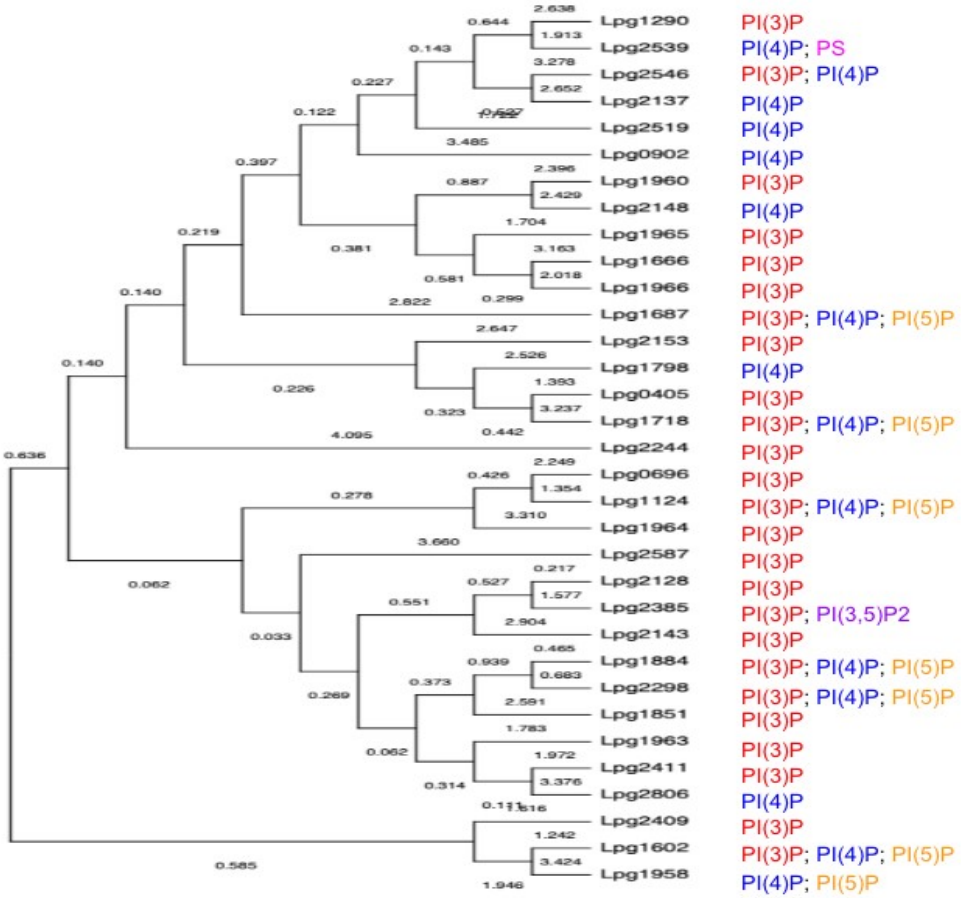


Figure 31 Maximum likelihood tree generated in MEGA of the full length protein sequence of 28 effector proteins secreted by *L. pneumophila* identified to bind to phosphoinositides. The numbers on each branch represent the relative number of changes that each sequence has from branching off from its respective branch point. Red represents an effector protein that binds PI(3)P, blue represents an effector that binds PI(4)P, yellow represents an effector that binds PI(5)P, and pink represents an effector that binds another PIP species.

We first wanted to see if the full-length protein sequences of 28 effector proteins identified by the Neunuebel lab as PIP-binders clustered with any specific pattern or if PIP species clustered together. We ran a multiple sequence alignment with ClustalW[37] of the 28 full length effector protein sequence and then generated a maximum likelihood tree from the multiple sequence alignment with the MEGA software (Figure 31). While some PI(4)P binding proteins clustered, most proteins did not cluster according to the PIP species they bound. We then wanted to see if the published PIP-binding protein domain clustered together according to the PIP-species they bound and the LED that was found to be conserved in its PIP-domain.

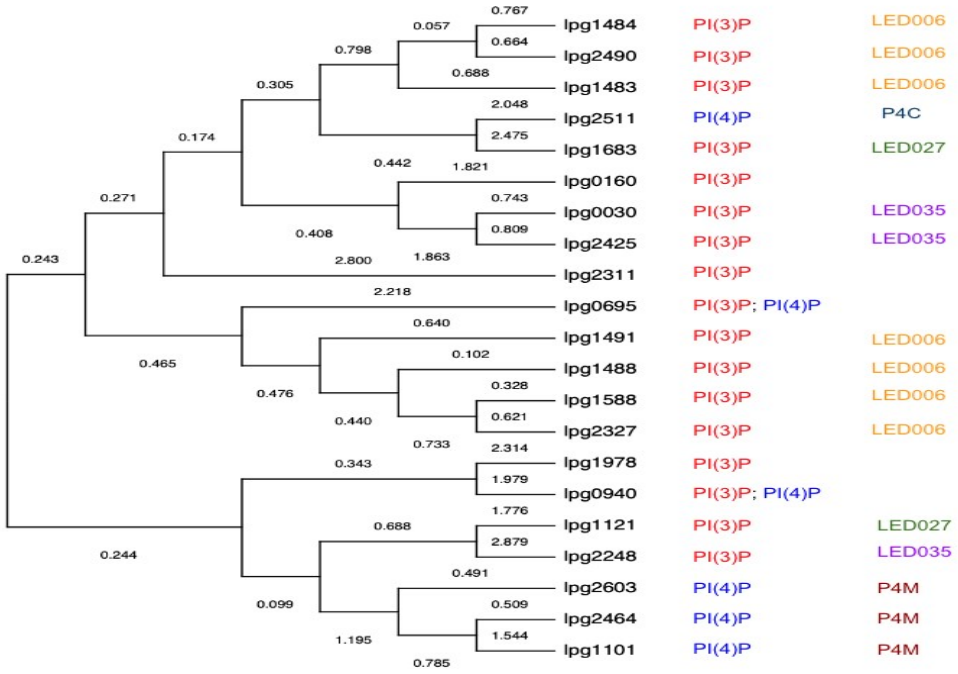


Figure 32 Maximum likelihood tree generated in MEGA of the PIP-binding domains from protein sequence of published PIP-binding proteins secreted by *L. pneumophila*. The numbers on each branch represent the

relative number of changes that each sequence has from branching off from its respective branch point. Red represents an effector protein that binds PI(3)P, and blue represents an effector that binds PI(4)P. Yellow represents PIP-binding domains that share conservation termed LED006, green represents PIP-binding domains that share conservation termed LED027, and purple represents PIP-binding domains that share conservation termed LED035. Brown and navy represent a PI(4)P binding domain.

The validated PIP-binding region has been identified and published for 21 different effector proteins secreted by *L. pneumophila*[26]. We took the protein sequences of each of these PIP-binding domains and ran a multiple sequence alignment using ClustalW and then generated a maximum likelihood tree from the multiple sequence alignment (Figure 32). This phylogenetic tree showed clear clustering of not only proteins binding to PI(3)P versus PI(4)P, but also showed clustering of the conserved LEDs. This result indicates that proper identification of PIP-binding domains and properly split full-length protein sequences is crucial to generate a maximum likelihood tree that shows clustering of proteins binding to the same PIP-species.

3.2.2 Domain Predictions of PIP-binding proteins

We then aimed to predict which region of the 28 PIP-binding proteins identified by the Neunuebel were responsible for allowing these proteins to bind to PIPs. We used a combination of bioinformatic software including HHpred, Alpha Fold, CONSURF and ProteInfer to not only split each protein into predicted domains,

but also predict the function of each of these domains. We then took the protein sequence of each of these domains, ran a multiple sequence alignment with ClustalW, and generated a maximum likelihood tree using the MEGA software (Figure 33).

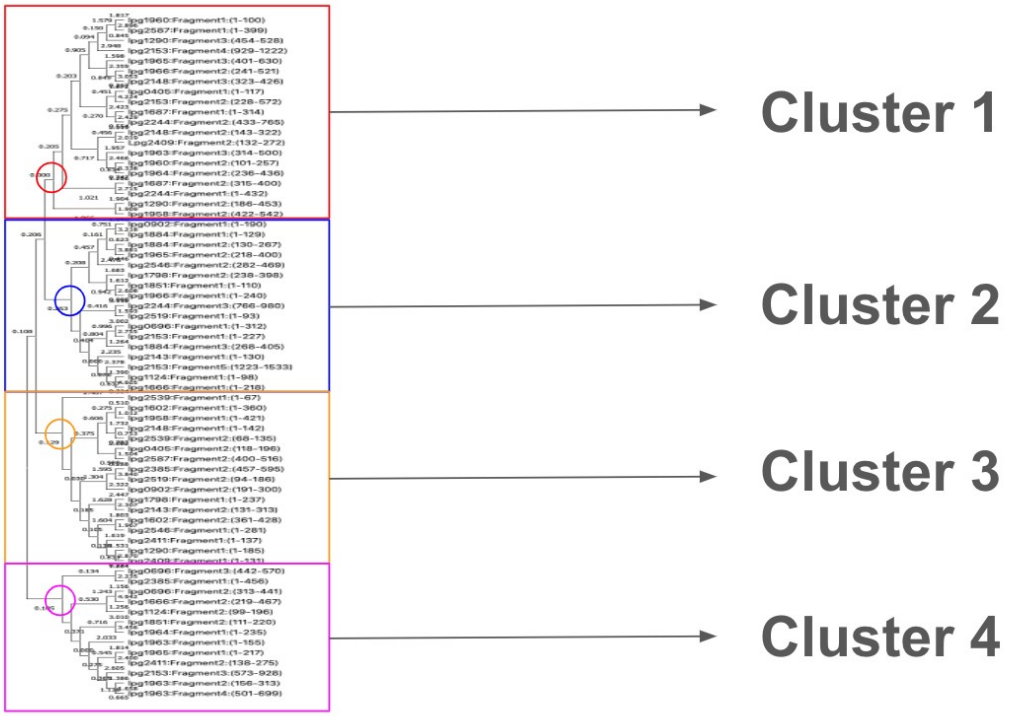


Figure 33 Maximum likelihood tree generated in MEGA of the split protein sequences of 28 effector proteins secreted by *L. pneumophila* identified to bind to phosphoinositides. The phylogenetic tree is split into four clusters. The top red cluster is cluster 1 and the red circle represents where cluster 1 branches off from the rest of the tree. The blue cluster is cluster 2 and the blue circle represents where cluster 2 branches off from the rest of the tree. The yellow cluster is cluster 3 and the yellow circle represents where cluster 3 branches off from the rest of the tree. The pink cluster is cluster 4 and the pink circle represents where cluster 4 branches off from the rest of the tree.

The phylogenetic tree generated from the split fragments from the 28 effector proteins identified by the Neunuebel lab to be PIP-binders revealed four clusters that branch off from the rest of the tree. Each circle in this phylogenetic represents each of the 4 cluster's origins. We split this phylogenetic tree into four clusters and added what the predicted function of each fragment was and which PIP species the protein the fragment is from binds to.

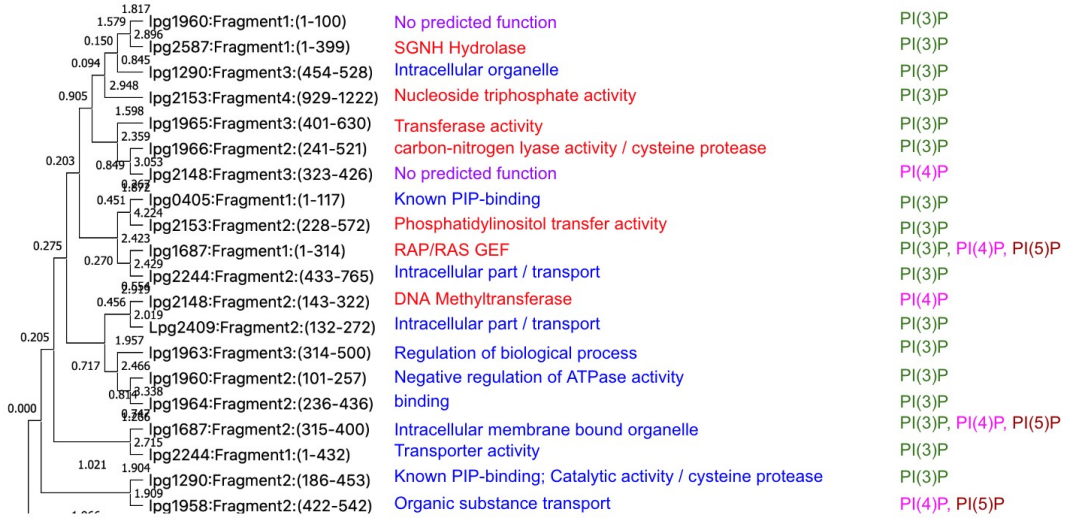


Figure 34 Cluster 1 of the Maximum likelihood tree generated in MEGA of the split protein sequences of 28 effector proteins secreted by *L. pneumophila* identified to bind to phosphoinositides. Directly to the right of the tree are the predicted function of each split fragment and to the right of the predicted function is the PIP species the protein was found to bind. Blue represents functions predicted to be involved with transport activity or an intracellular organelle part. Red represents functions predicted to have catalytic activity, and purple represents protein fragments that do not have a predicted function. Green represents an effector protein that binds PI(3)P, pink represents an effector that binds PI(4)P, and maroon represents an effector that binds any other PIP species.

Cluster 1 represents the top cluster from the phylogenetic tree in figure 34. There appears to be clustering of the different fragments predicted function. The bottom portion of the cluster consists of protein fragments predicted to be involved with transport activity or an intracellular organelle part. Conversely, the top portion of the cluster consists of protein fragments predicted to have catalytic activity. However, there appears to be no clustering of the protein fragments according to the PIP species the protein binds too.

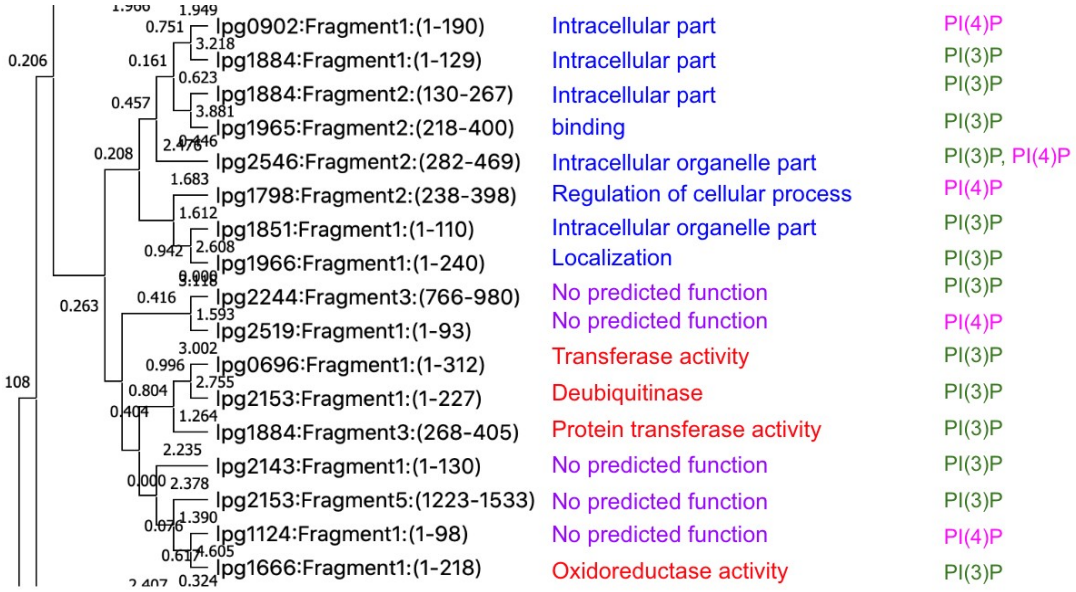


Figure 35 Cluster 2 of the Maximum likelihood tree generated in MEGA of the split protein sequences of 28 effector proteins secreted by *L. pneumophila* identified to bind to phosphoinositides. Directly to the right of the tree are the predicted function of each split fragment and to the right of the predicted function is the PIP species the protein was found to bind. Blue represents functions predicted to be involved with transport activity or an intracellular organelle part. Red represents functions predicted to have catalytic activity, and purple represents protein fragments that do not have a predicted function. Green represents an effector protein that binds PI(3)P, and pink represents an effector that binds PI(4)P.

Cluster 2 represents the second cluster down from the phylogenetic tree in figure 34. Again, there appears to be clustering of the different fragments predicted function. The top portion of the cluster consists of protein fragments predicted to be involved with transport activity or an intracellular organelle part. Conversely, the bottom portion of the cluster consists of protein fragments predicted to have catalytic activity or have no predicted function. However, there appears to be no clustering of the protein fragments according to the PIP species the protein binds too.

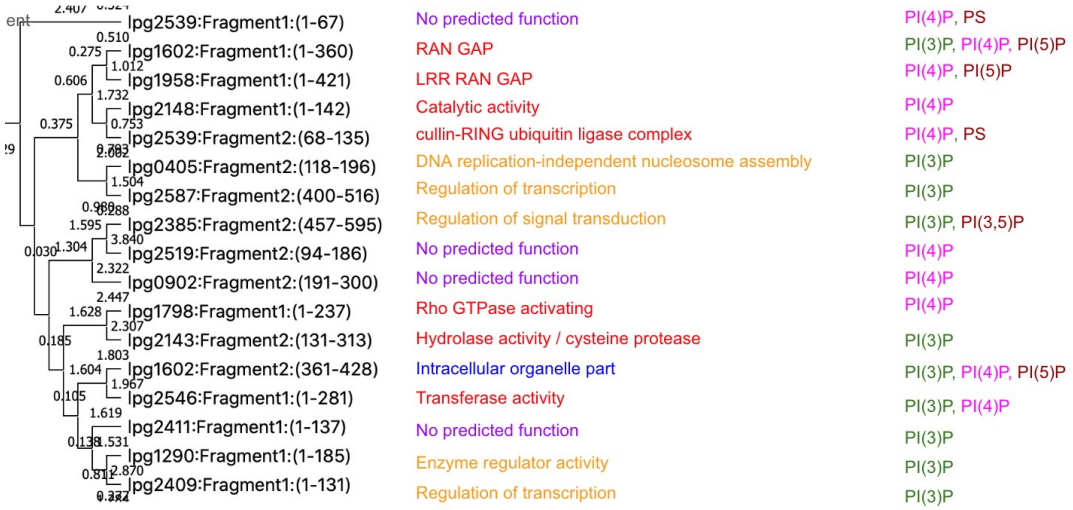


Figure 36 Cluster 3 of the Maximum likelihood tree generated in MEGA of the split protein sequences of 28 effector proteins secreted by *L. pneumophila* identified to bind to phosphoinositides. Directly to the right of the tree are the predicted function of each split fragment and to the right of the predicted function is the PIP species the protein was found to bind. Blue represents functions predicted to be involved with transport activity or an intracellular organelle part. Red represents functions predicted to have catalytic activity, purple represents protein fragments that do not have a predicted function, and orange represents protein fragments with regulatory activity. Green represents an effector protein that binds PI(3)P, pink represents an effector that binds PI(4)P, and maroon represents an effector that binds any other PIP species.

Cluster 3 represents the third cluster down from the phylogenetic tree in figure 34. Again, there appears to be clustering of the different fragments predicted function. Clusters of predicted functions form including a cluster of catalytic predicted functions near the top of cluster 3, and a cluster of regulatory predicted fragments below the catalytic proteins. The bottom portion of cluster 3 shows less clustering of predicted functions, but regulatory and enzymatic functions still tend to cluster together. Conversely to clusters 1 and 2 however, there appears to be some clustering of the PIP species the protein binds to. The top portion of the cluster includes proteins binding to PI(4)P and there is another cluster of proteins binding to PI(4)P in the middle of cluster 3. Proteins that bind to PI(3)P also appear the cluster in the bottom portion of cluster 3.

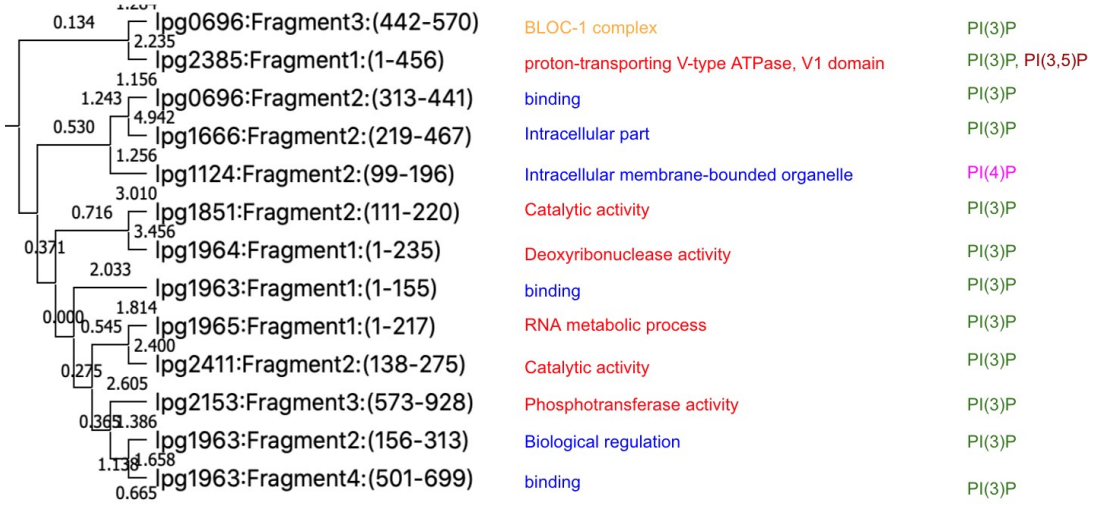


Figure 37 Cluster 4 of the Maximum likelihood tree generated in MEGA of the split protein sequences of 28 effector proteins secreted by *L. pneumophila* identified to bind to phosphoinositides. Directly to the right of the tree are the predicted function of each split fragment and to the right of the predicted function is the PIP species the protein was found to bind. Blue represents functions predicted to be involved with transport activity or an

intracellular organelle part. Red represents functions predicted to have catalytic activity, and purple represents protein fragments that do not have a predicted function. Green represents an effector protein that binds PI(3)P, pink represents an effector that binds PI(4)P, and maroon represents an effector that binds any other PIP species.

Cluster 4 represents the bottom cluster from the phylogenetic tree in figure 34. With this cluster however, there appears to be a mixing of predicted functions. Some catalytic functions cluster together, but overall catalytic predicted domains and domains predicted to be intracellular are well mixed together. However, almost every single protein fragment in cluster 4 binds to PI(3)P.

The phylogenetic tree in figure 33 shows that while many domains predicted were most likely predicted to be accurate due to to the clustering of predicted functions and the PIP species that fragment bound to, many of these fragments did not cluster, suggesting these fragments are not not in fact the true domains of the 28 effector proteins identified by the Neunuebel lab. The bioinformatic pipeline can be improved to split the proteins more accurately. However, these phylogenetic trees indicate that the bioinformatic pipeline we used to predict PIP-binding domains has some merit and can be improved upon even more to predict PIP-binding domains even more accurately.

Chapter 4

CONCLUSIONS AND FURTHER DIRECTIONS

The first main objective of this thesis was to generate HaloTag constructs fused with three *L. pneumophila* PIP-binding effector genes for two different *L. pneumophila* strains. With the help of the Gateway Cloning technology, I was able to clone three *L. pneumophila* effector encoding genes, *lpg1884*, *lpg2511*, and *lpg1958*, into the Gateway Destination plasmid pJB908-HaloTag-DEST. I confirmed correct cloning of gene products with colony PCR and transformed these constructs into *L. pneumophila* strains Lp02 and Lp03. I verified that Lp02 and Lp03 are both correctly expressing the protein products of these constructs with western blotting. With this first aim of the project, three HaloTag fused *L. pneumophila* effector constructs were generated and verified.

In the future, to help answer the question of where these effector proteins are being secreted in the host cell and which membrane are they targeting, *L. pneumophila* expressing these created can be infected into a macrophage cell. The HaloTag protein also allows the possibility of live cell imaging as opposed to fixed cell imaging. Traditionally fixed cell microscopy had been used where a macrophage cell during *L. pneumophila* infection is frozen and different proteins and membrane compartments could be fluorescently. However, live cell imaging allows for the dynamic processes of the cell to be viewed in real time. In the future our goal is to infect macrophage cells with these HaloTag-fused constructs, to understand where these PIP-binding proteins are secreted and what interactions could unfold with these proteins.

The second main objective of this thesis was to run the first bioinformatic analysis of newly identified PIP-binding effector proteins secreted by *L. pneumophila*. We also aimed to predict the possible PIP-binding regions of these newly discovered PIP-binding proteins. With the utilization of the MEGA software, we generated a fulllength protein sequence maximum likelihood tree and discovered that we would have to predict PIP-binding domains and could use phylogenetic trees as a method to validate whether these predicted domains were accurate. We used a combination of HHpred, Alpha Fold, ProteInfer, and CONSURF, to predict how best to split the fulllength protein sequences and predict the functions of these fragments. We were able to generate a maximum likelihood tree that showed much improved clustering of not only proteins bound to a specific PIP species, but also clustering of functions. However, this clustering was much less apparent than the phylogenetic tree generated from known PIP-binding domains. These results indicate that while the bioinformatic pipeline we generated was a good start, it can be significantly improved to generate more accurate predictions of PIP-binding domains.

In the future, we would like to improve upon the bioinformatic pipeline we generated by searching for more atypical domains. We heavily relied on HHpred, which takes advantage of homology searches through predicting protein sequences could be similar to known protein domains. However, it is becoming clear that *L. pneumophila* effector proteins could contain pseudokinases or other atypical enzymatic domains that have not been identified yet[38]. We hypothesize that utilizing bioinformatic pipelines, already shown to be effective in identifying atypical

enzymatic domains, in combination with the bioinformatic pipeline we developed will be the most accurate in predicting PIP-binding domains.

REFERENCES

1. Winn WC. Legionella. In: Baron S, ed. *Medical microbiology.* 4th ed. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. http://www.ncbi.nlm.nih.gov/books/NBK7619/.

2. Legionnaires disease history, burden, and trends | CDC. https://www.cdc.gov/legionella/about/history.html. Updated 2022.

3. Legionellosis (legionella) https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/Legionellosis%28Legionella%

29.aspx.

4. Legionnaires' disease-legionnaires' disease - symptoms & causes. Mayo Clinic Web site. https://www.mayoclinic.org/diseases-conditions/legionnaires-disease/symptomscauses/syc-20351747.

5. Lettinga KD, Verbon A, Nieuwkerk PT, et al. Health-related quality of life and posttraumatic stress disorder among survivors of an outbreak of legionnaires disease.

Clin Infect Dis. 2002;35(1):11-17. doi: 10.1086/340738.

6. Jonkers RE, Lettinga KD, Pels Rijcken TH, et al. Abnormal radiological findings and a decreased carbon monoxide transfer factor can persist long after the acute phase of legionella pneumophila pneumonia. *Clin Infect Dis*. 2004;38(5):605-611. Accessed. doi: 10.1086/381199.

7. Rahimi B, Vesal A. Antimicrobial resistance properties of legionella pneumophila isolated from the cases of lower respiratory tract infections. . 2017;10(1):59-65. https://biomedpharmajournal.org/vol10no1/antimicrobial-resistance-properties-oflegionella-pneumophila-isolated-from-the-cases-of-lower-respiratory-tract-infections/.

8. D'Ancona J. What is legionnaires' disease? https://www.inquirer.com Web site. https://www.inquirer.com/philly/health/What_is_Legionnaires_disease.html. 9. Escoll P, Rolando M, Gomez-Valero L, Buchrieser C. From amoeba to macrophages: Exploring the molecular mechanisms of legionella pneumophila infection in both hosts. *Curr Top Microbiol Immunol*. 2013;376:1-34.

10. Conner SD, Schmid SL. Regulated portals of entry into the cell. *Nature*. 2003;422(6927):37-44. Accessed Apr 30, 2024. doi: 10.1038/nature01451.

11. Newton HJ, Ang DKY, van Driel IR, Hartland EL. Molecular pathogenesis of infections caused by legionella pneumophila. *Clin Microbiol Rev*. 2010;23(2):274-298. doi: 10.1128/CMR.00052-09.

12. Research interests | infection-biology. https://sites.fct.unl.pt/infectionbiology/pages/research-interests.

13. Mondino S, Schmidt S, Rolando M, Escoll P, Gomez-Valero L, Buchrieser C. Legionnaires' disease: State of the art knowledge of pathogenesis mechanisms of legionella. *Annu Rev Pathol*. 2020;15:439-466. doi: 10.1146/annurev-pathmechdis-012419-032742.

14. Ninio S, Roy CR. Effector proteins translocated by legionella pneumophila: Strength in numbers. *Trends Microbiol*. 2007;15(8):372-380. doi: 10.1016/j.tim.2007.06.006.

15. 2009 annual report of the division of intramural research, NICHD | unit on microbial pathogenesis. . . https://annualreport.nichd.nih.gov/2009/ump.html.

16. Swart AL, Hilbi H. Phosphoinositides and the fate of legionella in phagocytes. *Front Immunol*. 2020;11:25. doi: 10.3389/fimmu.2020.00025.

17. Nakamura S, Yoshimori T. New insights into autophagosome-lysosome fusion. *J Cell Sci*. 2017;130(7):1209-1216. doi: 10.1242/jcs.196352.

18. Montaño-Rendón F, Walpole GFW, Krause M, Hammond GRV, Grinstein S, Fairn GD. PtdIns(3,4)P2, lamellipodin, and VASP coordinate actin dynamics during phagocytosis in macrophages. *J Cell Biol*. 2022;221(11):e202207042. doi: 10.1083/jcb.202207042.

19. Anquetil T, Payrastre B, Gratacap M, Viaud J. The lipid products of phosphoinositide 3-kinase isoforms in cancer and thrombosis. *Cancer Metastasis Rev*. 2018;37(2-3):477-489. doi: 10.1007/s10555-018-9735-z.

20. Zhu W, Hammad LA, Hsu F, Mao Y, Luo ZQ. Induction of caspase 3 activation by multiple Legionella pneumophila Dot/Icm substrates. *Cell Microbiol*. 2013;15(11):1783-1795. doi:10.1111/cmi.12157

21. Pike CM, Noll RR, Neunuebel MR. *Exploitation of phosphoinositides by the intracellular pathogen, legionella pneumophila.* IntechOpen; 2019. undefined/chapters/69108. 10.5772/intechopen.89158.

22. Bishé B, Syed G, Siddiqui A. Phosphoinositides in the hepatitis C virus life cycle. *Viruses*. 2012;4(10):2340-2358. doi: 10.3390/v4102340.

23. Szlasa W, Zendran I, Zalesińska A, Tarek M, Kulbacka J. Lipid composition of the cancer cell membrane. *J Bioenerg Biomembr*. 2020;52(5):321-342. doi: 10.1007/s10863-020-09846-4.

24. Halotag. https://www.ptglab.com/news/blog/halotag/.

25. PIP strips - lipid-protein interaction assay. Echelon Biosciences Web site. https://www.echelon-inc.com/product/pip-strips/.

26. Nachmias N, Zusman T, Segal G. Study of legionella effector domains revealed novel and prevalent phosphatidylinositol 3-phosphate binding domains. *Infect Immun*. 2019;87(6):153. doi: 10.1128/IAI.00153-19.

27. IJMS | free full-text | characterization of a PIP binding site in the N-terminal domain of V-ATPase a4 and its role in plasma membrane association. . . https://www.mdpi.com/1422-0067/24/5/4867.

28. Genomic analysis of 38 legionella species identifies large and diverse effector repertoires | nature genetics. . . https://www.nature.com/articles/ng.3481.

29. Nachmias N, Zusman T, Segal G. Study of legionella effector domains revealed novel and prevalent phosphatidylinositol 3-phosphate binding domains. *Infect Immun*.

2019;87(6):153. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6529665/. doi: 10.1128/IAI.00153-19.

30. Soriano M. Plasmids 101: Gateway cloning. https://blog.addgene.org/plasmids-101-gateway-cloning.

31. Western blot analysis | biocompare.com. . . https://www.biocompare.com/26776-Western-Blot-Analysis/.

32. Home. https://www.megasoftware.net/.

33. HHpred | bioinformatics toolkit. . . https://toolkit.tuebingen.mpg.de/tools/hhpred.

34. Highly accurate protein structure prediction with AlphaFold | nature. . . https://www.nature.com/articles/s41586-021-03819-2.

35. Sanderson T, Bileschi ML, Belanger D, Colwell LJ. ProteInfer, deep neural networks for protein functional inference. . 2023;12:e80942. https://doi.org/10.7554/eLife.80942. doi: 10.7554/eLife.80942.

36. ConSurf | evolutionary conservation profiles of proteins. . . https://consurf.tau.ac.il/consurf_index.php.

37. Multiple sequence alignment - CLUSTALW. https://www.genome.jp/toolsbin/clustalw.

38. Methods for discovering catalytic activities for pseudokinases - ScienceDirect. . . https://www.sciencedirect.com/science/article/pii/S0076687922001380?via%3Dihub.