8-2023

Characterization of Legionella pneumophila Effector Proteins, LneB and MavA

Kayode Adeyemi

CHARACTERIZATION OF *Legionella pneumophila* EFFECTOR PROTEINS, LneB and MavA.

A thesis submitted in partial fulfilment

of the requirements for the degree

Master of Science

Department of Biology

Western Kentucky University

Bowling Green, KY

By Kayode Adeyemi

August 2023

CHARACTERIZATION OF *Legionella pneumophila* EFFECTOR PROTEINS,

LneB and MavA

Defense Date: 26th June 2023.

Banga, Simran Digitally signed by Banga, Simran
Date: 2023.07.18 13:16:03 -05'00'

Committee Chair: Simran Banga, Ph.D.

Rodney King Digitally signed by Rodney King
Date: 2023.07.18 09:16:10
-05'00'

Committee Member: Rodney King, Ph.D.

Ajay Srivastava Digitally signed by Ajay Srivastava
Date: 2023.07.18 11:54:15 -05'00'

Committee Member: Ajay Srivastava, Ph.D.

Ranjit T. Koodali Digitally signed by Ranjit T. Koodali
Date: 2023.07.18 14:19:23 -05'00'

Dr. Ranjit T. Koodali

Associate Provost for Research & Graduate Education

ABSTRACT

CHARACTERIZATION OF *Legionella pneumophila* EFFECTOR PROTEINS,

LneB and MavA

The crucial virulence factor of accidental human pathogen *Legionella pneumophila* during the course of Legionnaire disease is the over 300 effector proteins secreted from its Dot/Icm secretion system. Eukaryotic host cells usually elicit an arsenal of immune responses against invading *L. pneumophila*. Nonetheless, the bacteria unexpectedly subvert these defense mechanisms to survive and proliferate unhindered in the host. Although some effector proteins have been proposed to play a significant role in this host-pathogen interaction, many still need to be characterized. The LneB and MavA proteins are examples of those effectors that need characterization. Thus, this study aimed to investigate the structural and functional characteristics of LneB and MavA proteins using several bioinformatics predictive pipelines and transcriptomics data supplemented experimentally through cell-based and biochemical assays to support the prediction.

The LneB protein was predicted to have histone acetylation activity (HAT) based on bioinformatics analysis. To investigate the HAT activity of LneB in vitro, the protein was ectopically expressed in the *Escherichia coli* BL21 strain and purified using nickel ion chromatography. The HAT activity assay was carried out on the purified LneB protein and on the nuclear extracts from LneB-GFP transfected 293T cells. Transcriptomics analysis shows that the LneB protein differentially induces upregulation of early growth factor and dehydrogenase (DHRS2) compared to the GFP control.

There was no significant difference between the in vitro HAT activity of LneB protein and the elution buffer (p-value = 0.1137, t-value = 5.537). In vivo, HAT activity was significantly reduced in cells transfected with LneB protein compared to the GFP control (p-value = 0.0025, t-value =

20.08). The HAT activity is not significantly different at a MOI of 10 or 100 when infected cells (Dot/Icm mutant and wild-type *L. pneumophila*) are compared to uninfected U937 cells (p-value = 0.8969 and 0.5384, respectively). However, the HAT activity in cells infected with an *L. pneumophila* Dot/Icm mutant at MOI of 100 was significantly lower than in cells that were not infected (p-value = 0.0236). This result suggests that the effector protein from the wild type plays a significant role in acetylating histone protein in the host. Further investigation is required to understand the HAT activity of LneB and other roles the protein could play in the host.

Our bioinformatics analysis suggested that the MavA protein possesses Ras-GEF domains and potentially binds to GTP. The protein is predicted to possess two coiled-coil domains and also interact with GTP, Ras and actin. The transcriptomic data from cells expressing MavA protein showed significant upregulation of sixteen genes, which are involve in steroid hormone metabolic processes, endocytic recycling, cilia movement among others. The sortilin receptor protein was the only repressed gene in the cell when compared to a GFP protein control. Connecting the bioinformatics finding and the review of literature, we suggested that the MavA protein could be involved in the biological process in the cell such as internalization of *L. pneumophila*, creation of *Legionella*-containing vacuoles in host cells through endosomal remodeling or cytoskeletal reorganization.

Keywords: *Histone acetyltransferase, RasGEF, Bioinformatics prediction, Transcriptomics analysis.*

ACKNOWLEDGMENT

I want to especially appreciate my supervisor Dr. Simran Banga for her support and guidance from the onset till the end of this project, and she is incredible. The immense help and contribution of my thesis committee members, in the person of Dr. Srivastava and Dr. King, cannot be overemphasized. I thank the WKU Graduate school and the Ogden College of Science and Technology for providing support in the form of a Graduate Student Research Grant and the OCSE Graduate Research Fellowship respectively. These opportunities provided some of the resources for this study and the time to engage in two projects. I cannot complete this acknowledgment without thanking the National Institute of General Medical Sciences (NIGMS) of the National Institute of Health (NIH) for the KBRIN IDeA grant # 8P20GM103436 that propels this study, including the Genomic Facility at the UofL for the transcriptomics analysis and Dr. Rinehart's technical support for the bioinformatics analysis. To the big family of the Department of Biology, I am highly grateful for impacting me with adequate knowledge and the thirst for more. I also thank Western Kentucky University for providing such a beautiful and excellent-driven campus community.

I want to thank the Education USA Abuja for making my transition from Nigeria to the USA possible. This appreciation is extended to my mentors for their support and counsel and to my family in Nigeria, for their moral support, even though it is virtual.

To my friends in Kentucky and other parts of the world, whose names are too numerous to be captured in this tiny space, I want to say thank you for all your encouragement.

Table of Contents

List of Table

List of Figures

1.0.INTRODUCTION

Legionnaires disease is a serious type of pneumonia caused by Legionella. The diversity of this family of bacteria is spread across more than 50 species and 70 serogroups, but only a few have been associated with human disease [1]. Legionnaires disease is usually associated with nausea, fever, and blood-stained cough. Other symptoms of Legionnaires include neurological symptoms, vomiting, diarrhea, headache, loss of appetite, ataxia, and pneumonia-associated manifestations [2]. Based on the severity of the clinical signs and symptoms, the death rate in immunosuppressed patients who are not receiving treatment can be as high as 40–80 percent, but it can be lowered to 5–30 percent with the appropriate case management [3].

The first recorded Legionnaires disease endemic led to the death of about 16% of the infected victims who were members of the American Legions that attended the annual convention in Philadelphia in 1976 [4]. The endemic marked the origin of the name of the disease [5]. Since then, the national incidence rate of Legionnaires disease had risen 5.5-fold from 2000 to 2017 [6] and it is the leading cause of death caused by waterborne infections in the USA after nontuberculous mycobacteria and Pseudomonas infections [7][6]. In 2018, the Centers for Disease Control and Prevention (CDC) reported about 10,000 cases of Legionnaire's disease in the United States, representing about 0.003% of the entire population. However, the actual infection rate is about 2.5 times higher due to under-diagnosis [8] which is a significant hindrance to the managing the disease, unlike other globally prioritized pathogens. A detailed understanding of the pathogenesis of *Legionella sp* can help manage these global priority pathogens because of their similar nature of infection and strategies of causing disease [5].

The genera *Legionella* consists of Gram-negative bacteria of the order *Legionellales*, class Gammaproteobacteria and phylum proteobacteria. The order *Legionellales* is classified based on

the presence of four conserved signature indels (CSI) in their members. Such proteins include tRNA-guanine (34) transglycosylase, lipoprotein-releasing system transmembrane protein (lolE) and tRNA (guanosine(37)-N1)-methyltransferase TrmD [9]. *Legionella pneumophila* accounts for more than 90% of Legionnaire disease (L.D.) in humans [3]. The remaining 10% of cases are caused by a group of the minority species that have been repeatedly isolated from hospitalized patients, including *L. micdadei, L. bozmanii, L. longbeachae, L. dumoffii,* and *L. feeleii*, as well as species that are infrequently found in humans, including *L. anisa, L. wadsworthii,* and *L. cincinnatiensis*[10].

L. pneumophila is a facultative intracellular bacterium that is transmitted by aerosols primarily generated from cooling towers of air-conditioning systems. The first endemic reported in Philadelphia of Legionnaire disease was associated with contaminated aerosols from humidifiers, hot tubs, and showerheads because the environmental conditions support the growth of the bacteria [11]. The bacteria naturally inhabit aquatic habitats, as biofilms or as an intracellular parasite of free-living amoebae like *Acanthamoeba castellanii* [12]. Further classification of *L. pneumophila* based on surface heat-stable, pronase-resistant antigen [13] led to the identification of 15 serogroups of which serogroups 1, 3, 4, and 6 are commonly associated with human disease [14]. Pangenomic analysis and clinical investigation identified the medically significant strains of *L. pneumophila* to be Philadelphia (USA), Lens (France), Paris (France), and Corby (England) [15].

In recent years, the use of disinfection, oxidation, and tolerant/indicator colony monitoring system has been an effective approach to manage the bacterial, *L. pneumophila* in water systems [16]. Introduction of oxidizing agent into the water system, such as chlorine or ozone, has been effective in reducing bacterial load. In addition, the water is monitored by testing for the presence of *L. pneumophila* or monitoring the colony-forming units of the bacteria. By tracking the water quality

regularly, it is possible to detect and respond to any increase in the levels of Legionella above 1 CFU/mL quickly and efficiently [17]. This system is a cost-effective way of controlling Legionella and has proven effective in preventing the spread of Legionellosis [16].

Legionella use various virulence factors to infect the lungs and their macrophage to cause lung inflammation and pneumonia. In eukaryotic cells, *L. pneumophila* thrives by forming a Legionella-containing vacuole (LCV), also called a replicative phagosome (R.P.), within the infected host cell to evade the immune system through creating a safe environment for its proliferation [18]. The LCV is secluded in a compartment that prevents *L. pneumophila* from being transported to the lysosomal network, unlike some non-pathogenic bacteria. Since the LCV avoids the phagolysosome pathway, it recruits the host's early secretory apparatus through a variety of effector proteins, and this apparatus has been used as a marker for identifying LCV in the cells [19]. The vesicles around the LCV appeared docked, spread across the surface, and eventually seemed to come from rough endoplasmic reticulum membranes. The bacterium multiplies rapidly in this ER-like compartment and ultimately lyses the host cell [20].

Humans are only incidentally associated with *L. pneumophila* as the final host because the bacteria grow in amoebas as a training ground to proliferate in human macrophages [21]. Human infection occurs in the macrophages because the specialized defensive cells share morphological and mechanistic similarities with amoebas, the natural host. *Legionellae* encode proteins that are important for their metabolism, but they also possess proteins that are secreted for their survival during infection in the host cell [22]. Surprisingly, the bacteria possess over 300 effector proteins that are released extracellularly into the host cells through the Dot/Icm Type IV secretion system (T4SS) in the bacterial cell membrane [3]. These effector proteins facilitate the bacteria's intracellular survival [23]. The effector proteins are virulence proteins in *L. pneumophila*. Some

of them share motifs with eukaryotic proteins and perform similar cellular processes. Examples include mimicking polyubiquitination by F-box effector [24] in the cell to the advantage of the bacteria's growth and survival [25]. The effectors play both positive and negative roles during *L. pneumophila* infection and it is suggested that the repertoire of *L. pneumophila* effectors change based on host specificity. This can significantly impact the bacteria's pathogenicity [26]. The life cycle of *L. pneumophila* and the involvement of the effector proteins in the formation of LCV is depicted in Figure 1 [27].

During infection, *L. pneumophila* is phagocytosed into the macrophage– a similar process occurs in amoeba, but requires pseudopodia, the natural host. Invasion can occur through traditional phagocytosis, coiling phagocytosis and pinocytosis, or multiple receptor-mediated pathways [28]. Immediately after the uptake, the bacteria secrete specific effector proteins depending on the type of infected host [29]. However, not all 300 effectors are secreted at the same time [30]. During the infective period, the bacteria is in its motile phase, but there will be a transition in phase as the bacteria is kept secure in the LCV to establish a nonmotile and less virulence form. The replication rate of the bacteria is higher during this stage leading to nutrient depletion and bacteria switch to a more motile, active, and virulent form in the LCV. At this latter stage, the bacterial lyse the infected cells continue the infection cycle in another healthy host cell [31].

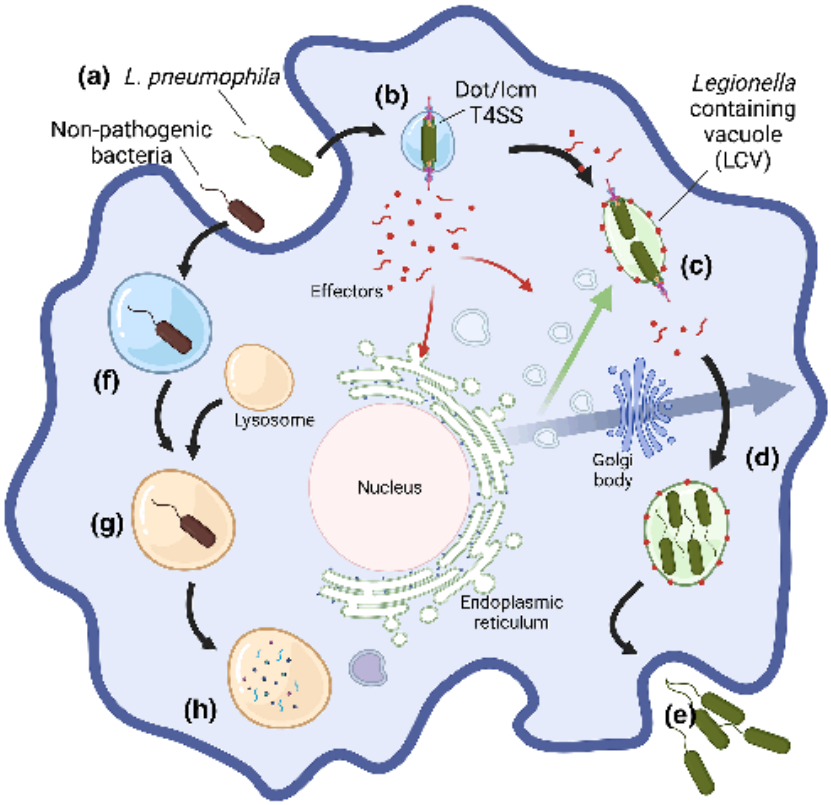


Figure 1: Intracellular life cycle of non-pathogenic bacteria and *L. pneumophila*. *L. pneumophila* infects a host cell through phagocytosis (a) utilizes its Dot/Icm T4SS to secret effector proteins (b) to interfere with host vesicle trafficking to form a protective Legionella containing vacuole (c). The protected *L. pneumophila* multiplies within the LCV (d) and lyse the cells in their motile phase (e). Unlike the *L. pneumophila*, non-pathogenic bacteria are taken up in the phagosome of the host, which merged with the host lysosome (f) to form phagolysosome (g). The hydrolytic activity within the phagolysosome leads to the destruction of the non-pathogenic bacteria (h) [27].

1.1. Bacterial Transport Secretion System

Both bacterial and eukaryotic cells possess membrane transport systems that are involved in vital cellular processes such as the uptake of necessary nutrients from the external environment,

extracellular secretion of metabolites, the expulsion of harmful byproducts, secretion of enzymes/virulence factors, and maintenance of cellular homeostasis by controlling the intracellular concentrations of ions and solutes [32]. These transport systems, also referred to as secretion systems, comprise 3 - 16% of all open reading frames (ORFs) in bacterial genomes [33]. In general, all molecules that cannot cross the cell membrane by diffusion are the substrates of these transport systems. These molecules include metals, sugars, amino acids, peptides, oligosaccharides, and macromolecules like proteins and nucleic acids. Some bacteria have taken advantage of this secretion machinery to affect their host's cellular and molecular processes; thus, the transport system has become one of the essential virulence factors in many pathogenic bacteria. A total of six types of secretion systems have been identified in various genera of bacteria. These transport systems are named Type 1 – 6 based on their morphology and type of specific substrates [34].

1.1.1. Type 1 secretion System

The type 1 secretion system (T1SS) are comprised of an ATP-binding cassette (ABC), an adapter protein, and TolC-like proteins in the inner membrane, intramembranous space, and outer membrane of most Gram-negative bacteria [35]. The transport of specific substrates through the T1SS does not require translocation through the periplasm before complete movement to the extracellular environment from the cytoplasm. Biochemically, the substrates transported by the T1SS are highly acidic, contain glycine/aspartate repeats at the C-terminals, and binds to calcium ions [36]. However, the glycine/aspartate repeat is not the primary signal sequence for all substrates of the T1SS. Apart from sub-family 1, all substrates of the T1SS contain secretion signal at the tail C-terminus which is necessary for secretion [37]. The functionality of the T1SS has also be shown to be through LssB, LssD and TolC dependent mechanism [38].

T1SS and its substrates from selected bacteria are listed in Table 1. The secretion system is also involved in early stage of *L. pneumophila* growth in amoeba and macrophages, by precisely enabling effective entry of *Legionella* into host cell [39].

Table 1: Type 1 secretion system from selected bacteria and their substrates.

Organism	**Substrate**	**Reference**
Pseudomonas aeruginosa	Iron	[40]
Pseudomonas aeruginosa	Alkaline proteases AprA and AprX	[41]
Burkholderia pseudomallei	Hemolysin	[42]
Stenotrophomonas maltophilia	Biofilm promoter Ax21	[43]
Bordetella pertussis	Adenylate cyclase	[44]
Escherichia coli	Colicin CvaC	[45]
Pseudomonas hemolytic	Leukotoxin	[46]
Escherichia coli	Hemolysin	[47]
Legionella pneumophila	Adherence/invasion RtxA	[48]

1.1.2. Type 2 Secretion System

The type II secretion system (T2SS) is mostly found in Gram negative bacteria, and it is important for transporting proteins from the bacterial cytoplasm to the extracellular environment across the lipid bilayer [49]. The T2SS is crucial for bacterial invasion because the majority of the substrates are degradative enzymes such as proteases and lipases [50]. The T2SS requires two steps in secreting its substrates. The first step is aimed at transporting substrates from the cytoplasm to the periplasm across the inner membrane while the second step requires the transport of substrates

from the periplasm to the extracellular compartment across the outer membrane [51]. The first step requires the Sec or Tat secretory pathways[52]. The T2SS is made up of multimeric protein subunits known as general secretory proteins (GSPs), which are encoded in a single operon. One important GSP protein is the secretin GspD, which is used for creating channels for allowing substrate to move out of the bacterial cell. Sixteen GSPs proteins are essential for the function of the T2SS [53].

The T2SS can be further broken down into four components: the outer, inner membrane complexes, the pseudophilus and the ATPase. The inner membrane component of the T2SS is made up of four GSPs; GspC, GspE, GspL and GspM [53]. The role of each Gsp is unique. The GspC forms a complex with GspD to gate the pore through which substrate exit the cells, while the interaction of GspL, GspC and GspM create a shield for them against proteolytic degradation. GspE, being a transmembrane multipass protein, serves as ATPase that drives the assembly and disassembly of the pseudopilus [54]. The pseudopilus is made up of proteins that are similar to pilins, also called pseudopilins. The pseudopilins include GspG, GspH, GspI, GspJ and GspK with hydrophobic tail which enabled the pseudopilins to get inserted into outer leaflet of the inner membrane and also a globular hydrophilic head [55]. The T2SS has been found to be crucial for various human intracellular and extracellular pathogens including *Escherichia coli, Vibrio cholerae, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa,* and *Legionella pneumophila*.

Table 2: Type II secretion systems from selected bacteria and their substrates.

Protein	Organism	Activity	Reference Examples
Toxins	*Vibrio cholerae*	ADP-ribosylation of Gsα subunit leading to increased adenylate cyclase activity and raising cAMP levels.	[56]
Exotoxin A	*Pseudomonas aeruginosa*	ADP-ribosylation of elongation factor 2; Inhibition of protein synthesis in host cells	[57]
Pore forming toxin (aerolysin; cytolysin)	*Vibrio cholerae*	Host cell membrane depolarization and lysis	[58]
Proteases	*Acinetobacter Species*	Cleavage of proteins or peptides; Breakdown of host extracellular matrix; Tissue damage.	[59]
Lipid-modifying enzymes	*Acinetobacter baumannii*	Breakdown of lipids to fatty acids and glycerol.	[60]
Carbohydrate-active enzymes	*Erwinia chrysanthemi*	Breakdown of polysaccharides and Depolymerization of plant cell wall.	[61]
Phosphatases	*Legionella pneumophila*	Dephosphorylation; Phosphate acquisition; Phosphate solubilization	[62]
Nucleic acid targeting enzymes	*Pseudomonas aeruginosa*	Hydrolysis of DNA and RNA; generation of nutrients that include carbon, nitrogen and phosphate.	[63]

Metal reductase	*Shewanella putrefaciens*	Reduction of insoluble metal oxides; Electron transport; Anaerobic respiration	[64]
Others	*Legionella pneumophila*	Adherence; Biofilm formation	[65]

1.1.3. Type 4 Secretion System

The T4SS is made up of a pilus that extends into the extracellular space, as well as a core complex that spans the inner and exterior membranes. Typically located at the C-terminus, secretion signals in substrates are required to transfer effector proteins across the T4SS. These signals are made up of collections of hydrophobic or positively charged residues. The T4SS is also known to transport nucleic acid, essential for the genomic island-associated T4SS (GI-T4SS) for horizontal transfer of genes. In addition, proteins or protein-DNA compounds can move through a channel found in the T4SS [66]. The two major classification of T4SSs are the conjugation systems and the effector translocation systems. Conjugation systems are explicitly designed to exchange mobile genetic components between microorganisms. Systems for effector translocation deliver effector proteins, that change the target cells operations [67]. A collection of cytoplasmic ATPases that may energize significant conformational changes in the translocation complex drives this translocation. The wide range of roles among T4SS family members demonstrates their versatility. In addition to the movement of nucleic acid across the membrane, pathogenic Gram-negative bacteria utilize some T4SSs to translocate several virulence factors into the host cell [68]. Some of the T4SSs used by bacteria are shown in Table 3.

Table 3: Type IV secretion systems from selected bacteria and their substrates.

Organism	**Function**	**Reference**
Burkholderia cepacia	Plasmid mobilization	[69]
Xanthomonas citri	Numerous effectors	[70]
P. aeruginosa	Horizontal gene transfer (HGT)	[71]
H. influenzae	Horizontal gene transfer (HGT)	[71]
H. influenzae	Antibiotics resistance gene	[72]
P. aeruginosa	Biofilm formation and antibiotic resistance	[73]
Burkholderia cepacia	Plasmid mobilization	[69]
Agrobacterium tumefaciens	T-DNA-relaxase complex	[67]

Being a versatile virulence machine, the T4SSs is further classified based on structural components into T4ASS and T4BSS. If the structure of T4SS resembles the VirB/D4 complex of the *Agrobacterium tumefaciens*; a plant parasite, it is classified as T4ASS. The system functions in a similar pattern as that of *A. tumefaciens* [74].

T4SS is classified as T4BSS if it resembles the transfer system of the Inc plasmid. The Defective in organelle trafficking/Intracellular multiplication (Dot/Icm) in T4BSS machinery transports an extraordinary number of effectors—more than 300 per *L. pneumophila* strain—and is crucial for

subjugating protists and human macrophages as niches for replicative growth [27]. The Dot/Icm is a five-protein core complex that traverses inner and outer bacterial membranes. These complexes include DotC, DotD, DotH, DotG, and DotF. DotL is a homohexamer complex consisting of subunits positioned in the N-proximal region of the transmembrane (NTD) and the nucleotide-binding domain (NBD) [75]. The C-terminal extension (CTE) consists of 590 – 783 amino acid residues of the DotL and it is associated with DotN, IcmS, IcmW, and LvgA adaptors.

The DotL Type IV coupling protein (T4CP) recruits hundreds of effector proteins that are both adaptor-dependent and -independent across the T4SS, while the CTE of the DotL only transports adaptor-dependent effectors (Fig. 2) [76]. IcmS and IcmW form heterodimer known as IcmSW which interact with the CTE along with DotN, LygA and IcmI. Through membrane-proximal ATPase and transmembrane helical domains, the DotL interacts with DotM, which is an inner membrane protein, however, the DotL domain alone is not sufficient for processing bacteria effectors[77]. In addition, DotL interacts with the inner membrane protein DotM (IcmP) through its transmembrane helices and its membrane-proximal ATPase domain to form a complex that is able to recognize effector proteins [78].

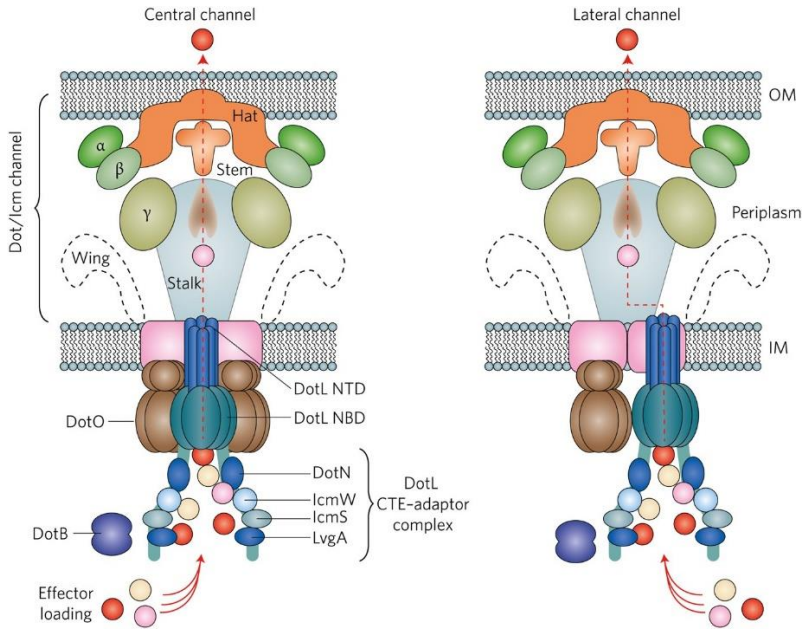


Figure 2: Spatial arrangements for the DotL–Dot/Icm channel. The core channel of DotL transports substrate effector proteins across the inner membrane, while the Dot/Icm T4SS forms a channel across the periplasm and outer membrane (red arrows and effectors). To stimulate substrate protein transfer across the T4SS, DotL cooperates with DotO and DotB [79].

A unique evasion strategy of *L. pneumophila* is through the use of Dot/Icm substrates to modulate host cellular processes to prevent the fusion of the phagosome, created during bacteria internalization, and the host endosome, to inhibit the formation of phagolysosome [80].

1.2.Ras Protein Signaling

The Ras superfamily protein is a critical molecular switch in the eukaryotic host cell. Ras proteins are a family of proteins present in all eukaryotes and it is named after “Ras sarcoma virus” where it was first discovered. They are crucial for signal transduction pathways [81]. The prominent

families of the superfamily, Ras, Rho, Ran, Rab, and Arf GTPases (Fig. 3), are classified based on their structure, localization, and functions [82]. These proteins generally receive signals from the host cell surface and relay them to respective organelles to control cellular processes such as growth, differentiation, and apoptosis of cell [83].

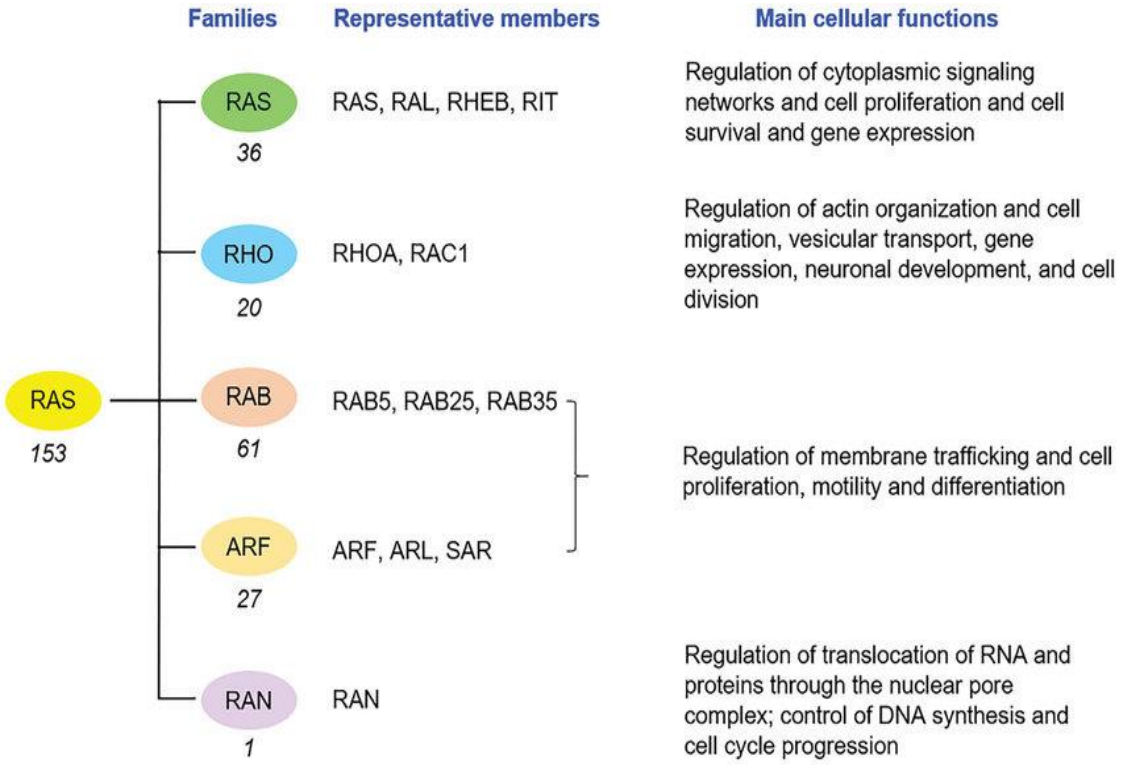


Figure 3: Subfamilies of Ras and their corresponding molecular function in the cell [84]**.**

The Ras protein is activated when it binds to GTP, which is then exchanged for GDP by the GTPase activity of Ras. As a result, GDP is released, and the Ras protein is free to activate various downstream targets. Ras proteins are regulated by multiple proteins, such as guanine nucleotide exchange factors (Ras-GEFs) and GTPase activating proteins (GAPs). Ras-GEF proteins promote the exchange of GDP for GTP. Simultaneously, GAPs stimulate the GTPase activity of Ras, leading to its deactivation (Fig. 4). Activation of Ras can also be regulated by phosphorylation,

which alters its conformation [85]. Ras-GEFs are essential components of the Ras/Raf/MAPK signaling pathway. They are also crucial for the activation of G-protein coupled receptors (GPCR), which are involved in the control of various physiological processes. The multi-domain proteins Ras-GEF and Ras-GAPs can interact with many other proteins, lipids, and regulatory molecules to regulate the amount of active and inactive Ras [86]. Once activated, Ras proteins transmit signals by binding to specific effector proteins, which can activate or repress other proteins, ultimately leading to changes in the cell's behavior. The majority of invading intracellular pathogens utilize Ras protein to infect, and multiply in their hosts and as a means of controlling the small GTPases, the pathogens have evolved various effectors like GEFs, GAPs, and GDIs which imitate GTPase modulators [87].

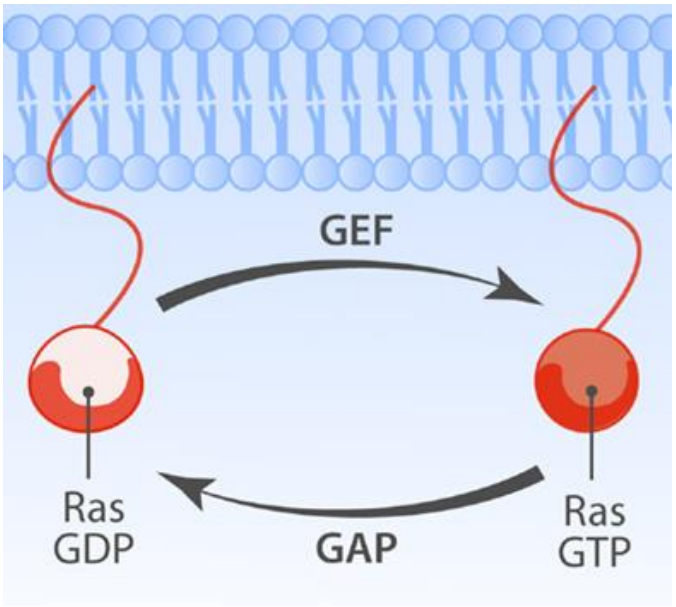


Figure 4: Activation of Ras protein by Ras-GEF [88]. Ras protein is activated when attached with GTP and are in inactive form when bounded to GDP, the GDP-bounded form of Ras protein is changed to active form through the guanine nucleotide exchanged factor (GEF) and the protein is brought back to inactive form through GTPase activating protein (GAP).

1.2.1. The role of modulators of small GTPase in *Legionella pneumophila* Infection

Endocytosis and vacuole-directed biosynthetic trafficking are connected by an essential hub through the endosomal trafficking route. Early endosomes are created when vesicles from the Golgi or plasma membrane combine. These endosomes then go through a maturation process that culminates in the fusion of late endosomes with the lysosome [89].

Ras/Rab regulation, either by activating or deactivating the GTPase, has been recognized as a crucial tactic to remodel the ER to produce the LCV. For instance, the DrrA/SidM effector protein of *L. pneumophila* regulates the Rab-GDI complex, which serves as the storage site for recycling and trafficking Rab1 proteins in various membrane-bound compartments. The transport of Rab1 to cellular membranes depends on the release of GDI from the complex. Once released, the Rab protein associates with the membrane and exchanges its bound GDP for GTP. The DrrA effector protein initiates the exchange of GDP to GTP on Rab1 protein in the host during *L. pneumophila* infection, a function that is similar to the host Ras-GEFs [90].

Other effectors of *L. pneumophila*, such as AnkX and Lem3, were found to work in synergy to regulate the phosphorylation of a small GTPase called Ypt1 in yeast cells to regulate membrane trafficking. The effector protein (AnkX) phosphorylates Rab1 with a phosphorylcholine moiety using cytidine diphosphate-choline as a donor instead of GTP [91]. The PieE effector protein of *L. pneumophila* was found to dimerize with its coil-coil structure at coil-coil3 and coil-coil4. The coil1-coil2 in the PieE protein were crucial for binding the host Rab GTPase 1a, 1b, 2a, 5c, 6a, 7, and 10. The interaction of Rab protein with PieE effector protein was correlated with their localization on the endoplasmic reticulum and the LCV. [92].

The *L. pneumophila* effector VipD binds to the endosomal regulator Rab5 protein and localizes to the endosome, where it causes the removal of the lipid phosphatidylinositol 3-phosphate. Without this essential lipid, endosomes are inhibited, and *L. pneumophila* is protected from their harmful effect. This reduces the exposure of *L. pneumophila* to the endosomal compartment and prevents the surrounding vacuoles from obtaining the Rab5 protein and modifies the protein makeup of endosomes to prevent fusion with vacuoles that contain Legionella [93].

In other bacteria, such as *Shigella flexneri* and *Salmonella typhi*, the Ras-GEFs have become a distinct class of virulence factors and no apparent sequence similarity exists between these GEFs [94], [95]. The same regulation of host Ras protein has been shown in *Mycobacteria tuberculosis*, where the expression level of guanine nucleotide exchange factor-HI, Ras-GEF, and other proteins was elevated during the pathogen infection [97].

1.3.Bacteria Nucleomodulins

Through long-term interactions with their hosts, bacterial pathogens have evolved unique arsenals of effector proteins that interact with specific host targets and reprogram the host cell into a permissive niche for pathogen proliferation. The targeting of effector proteins to the host cell nucleus where they modulate nuclear processes is an emerging theme among bacterial pathogens. Chilton and colleagues discovered some proteins in *Agrobacterium tumefaciens* that modulate processes in the nucleus. These proteins were found to be crucial to tumorigenesis in transfected plant cells; this discovery opened an entire field of research into bacterial nucleomodulins [98]. When certain bacteria proteins localize in the nucleus, they regulate specific nuclear processes, hence, they are called nucleomodulins [99]. In contrast, nuclear-targeting proteins that affect the host cell cycle are called cyclomodulins [100]. The processes influenced

by nucleomodulins include change in chromatin crescents, autophagic processes, modification of histones, direct methylation of host DNA, RNA splicing, and DNA replication. These processes interrupt the host cell signaling pathways, transcription regulation, and promoter of tumor induction [101].

1.3.1. Protein Translocation to Host Nucleus

Nucleomodulins or cyclomodulins could either be secreted to the host extracellular environment by pathogens, endocytosed from host extracellular environment or they could be internalized alongside with bacteria or secreted into the host after infection [102]. Many Gram-negative bacteria pathogens introduce their nucleomodulins directly to the host cell through their T3SS and T4SS [94]. Recently, it was shown that nuclear localization of proteins can be mediated by the nuclear envelope-associated endosomes (NAEs) [103]. Some nucleus-localized proteins escape the lysosomal degradation to the cytoplasm in a retrograde trafficking-independent pathway before finally reaching the nucleus [104].

The dynamic interchange of macromolecules between the nucleus and cytoplasm is highly selective for the nuclear envelope. The exchange is regulated by nuclear pore complexes, in which small molecules pass through the NPC via passive diffusion and larger molecules (>40kDa) pass through via karyopherins [105], [106]. The karyopherins bind to proteins in the cytoplasm that have nuclear localization signals (NLS). The NLS is rich in lysine and arginine; however, the short amino acid sequence varies between proteins. The NLS could be classic - either mono-partite or bipartite sequences or non-classic. These classic NLS are transported into the nucleus by importin-β, a karyopherin transport receptor and non-classic NLS possessing unique sequence, binds to the

importin-α1. The non-classic does not depend on an importin adaptor for nuclear localization [100].

Several effector proteins in *L. pneumophila* have been associated with the modulation of molecular processes in the host cell's nucleus after their successful penetration through the nuclear pore. Most of these activities are directed to either the DNA, the RNA transcript, or associated proteins in the nucleus of the host.

1.3.2. Nucleomodulins in *L. pneumophila*

1.3.2.1.Autophagic Interference

Autophagy is a survival strategy in eukaryotic cells where materials that are deemed for degradation are translocated from the cytoplasm to the lysosome. Such materials include damaged organelles, proteins taken up via phagocytosis and pinocytosis or pathogens. Autophagy could be carried out through chaperone, microautophagy (via pinocytosis) or macroautophagy (via the fusion of phagosome and lysosome) [107]. A selective macro-autophagy that targets intracellular pathogen from the phagosome to the lysosome is called xenophagy.

When viewed from the viewpoint of the host cell, xenophagy is an intracellular supporting mechanism by the innate immune system. The process entails the formation of double-membrane vesicles around damaged content in the cytosol by proteins called autophagic associated proteins (ATGs). *L. pneumophila* causes epigenetic methyladenine alterations (GATC motif to G (6mA) T.C.) in the promoter of Atg12 and Atg8, which are in charge of developing effective autophagic reactions that may attract cellular innate immune responses and cause apoptosis of the bacteria-infected cells [108]. *L. pneumophila* nucleomodulins, such as RavZ (lpg1683), SidE, and SetA, were also discovered to have similar autophagic suppressions in host cell [109]. The most frequent

pathways to autophagy suppression include the LC3 conjugation system (which is an ortholog of Atg8 in yeast), the destruction of ATG proteins, DNA methylation, and glycosylation [110]. The successful adaption of *Legionella pneumophila* in alveolar macrophage and the development of LCV are ultimately aided by a reduced production of LC3-I, an isoform of LC3 and ATGs complex (Atg5-Atg12) [111].

1.3.3. Modulation of DNA Replication

Two important phases in cell cycle in eukaryotic cells are cell division and DNA replication. It was recently shown that *L. pneumophila* inhibits DNA replication of its natural host, *A. castellanii*, during infection[112]. In addition to regulating host DNA replication, effectors of *L. pneumophila* has been shown to affect the bacterial nucleotide metabolism processes. For instance, the ClpP regulated effector protein in *L. pneumophila* mediates factors involving in DNA replication, repair, and recombination proteins. Seven of the ClpP regulated effector proteins were shown to be secreted in the replicative phase (RP) and eleven are important during the transmissive phase (TP) [113].

1.3.4. Modulation of Transcription Process

L. pneumophila effector Lpg2519, also called SnpL, localizes to the nucleus and binds to a conserved transcriptional regulator, SUPT5H, in the host This influences mRNA processing and transcription elongation process during *Legionella* infection [114]. One of the cellular effects is the upregulation of genes that are involved in signal transduction in the host during infection and the death of macrophage cells [99]. However, this protein is not essential for bacterial intracellular survival in bone marrow-derived macrophage or *Acanthamoeba castellanii* cells. SnpL nucleomodulins do not have an identified nuclear localization signal (NLS) motifs or conserved

catalytic domains [114]. Another strategy for reprogramming the host transcription process employed by *L. pneumophila* is interrupting the host Polymerase II-transcriptional elongation process using effectors with ankyrin repeat domain such as AnkH effector. The nucleomodulin (AnkH) sterically interacts with a subunit of the small nuclear ribonucleoprotein (snRNP) complex responsible for polymerase II inhibition; this eventually promotes transcription process in the host [115]. Finally, *Legionella pneumophila* also secrete ankyrin domain protein, AnkX, which plays a role in manipulating inflammatory processes [116]. The AnkX protein does this by localizing in the nucleus while interacting with PLEKHN1 [117]; however, the functional implication of this interaction is not well-known [99].

1.3.5. Histone Modification

The histone is a eukaryotic, chromatin-associated octameric protein that constitutes the structural unit of the nucleosome, consisting of duplicated copies of H3, H4, H2A, and H2B subunits around which the DNA wrapped (Fig 5).

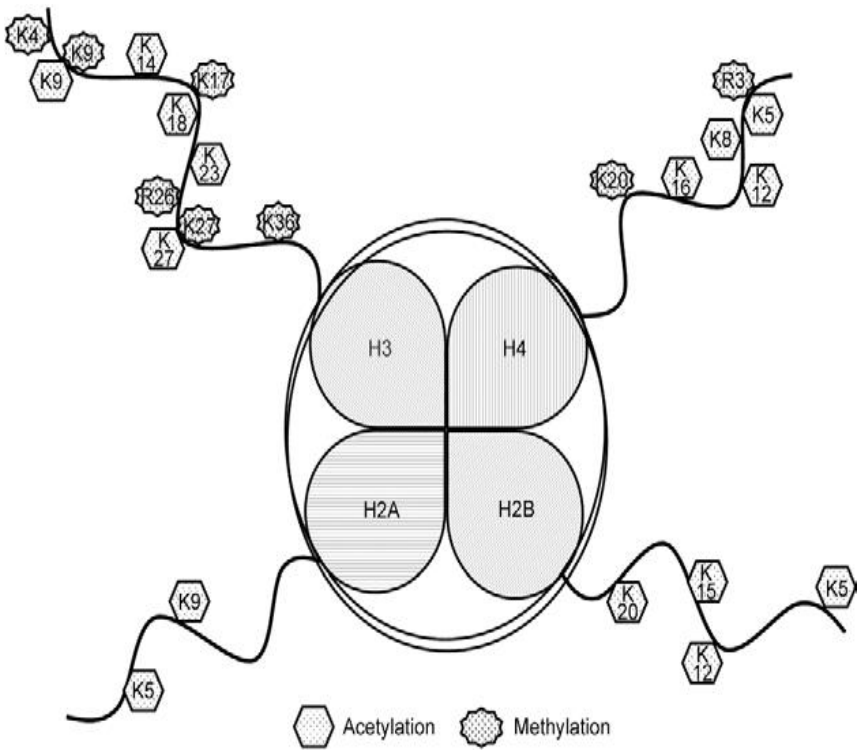


Figure 5: Amino acid residues in the tail of histones protein subunits [adapted from 72]. Four major histone subunits (H2A, H2B, H3 and H4) with individual modifiable N-terminal tails. On the H2A, Lys5 and Lys9 are open to acetylation. On the H2B, Lys5, Lys12, Lys15 and Lys20 are open to acetylation, on H3, Lys9, Lys18, Lys23, and Lys27 are available to acetylation and Lys4, Lys9, Lys17, Arg26, Lys27 and Lys36 are available to methylation. On the H4, Lys5, Lys8, Lys12 and Lys 16 are available to acetylation while Arg3, and Lys20 are available to methylation.

Histone protein is also an epigenetic target in the host by the *Legionella* effector. Most post-translational modifications (PTMs) caused by nucleomodulins on histone protein usually engage the N-terminal tails of the protein in the form of acetylation, methylation, phosphorylation, ubiquitination, sumoylation, carbonylation, and glycosylation [118]. However, the molecular impact of the individual PTMs on gene expression does not have a single outcome. There have been reports of transcription activation when lysine residues in H3 or H4 undergo acetylation as it

was during trimethylation of H3K4. However, methylation is generally known to lead to transcriptional silencing in H3K9 and H3K27 [119].

1.3.5.1.Histone Methylation

Considering the plasticity and redundancy of the *Legionella* effector proteins, the SET domain has been identified as a common motif employed to induce specific epigenetic modifications in host cells via histones methylation that modifies the chromatin landscape. Some identified SET domain effectors include RomA, LegAS4, lmi2895, lmi3093, lfa1741, lfa1445, and lha1431. These also contain an Ankyrin (ANK) domain except for lmi3093 and lha1431 [120]. The sequence identity of these effectors to the human SET domain ranges from 30% - 35%. The RomA methylates H3K14 [121] and LegAS4 methylates H3K4 [122], while the rest are annotated in silico and yet to be established experimentally. The implication of histone methylation in *Legionella* pathogenesis is mostly repression of gene transcription of immunogenic factors. The methylation of H3K15 by RomA leads to direct transcriptional repression in the host cell, while methylation by LegAS4 leads to transcriptional activation of the ribosomal gene in the host [123].

1.3.5.2.Histone Acetylation

Histone acetylation is an epigenetic modification of histone that involves addition of an acetyl group to lysine residue to four of the eight core histone proteins [124]. Due to the –NH2 group in lysine (positive functional group), the amino acid exhibits a positive steric charge within the Cell (Fig. 6). The significance of this is that it links the histone protein to the DNA strand by binding tightly to DNA, which is a negatively charged molecule in the Cell, therefore, regulating gene transcription [125]. Post-translational modification of histone protein such as methylation and acetylation are however important ways of regulating gene expression. Acetylation of the lysine

residue in histone proteins helps neutralize the covalent bond between the negatively charged DNA and histone proteins, eventually leading to the relaxation of the heterochromatin structure, thus giving more access to other transcriptional activators to the DNA strand to increase gene expression [126]. On the histone H3 protein, lysine at the -NH2 terminal on positions 9, 14, 18, and 23 has been reported to be more open to acetylation.

Similarly, on histone H4, K5, 8, 12, and 16 are also conserved targets for histone acetyltransferase enzymes [119]. Due to the role of Acetyl-Coenzyme A in the process of generation of acetyl groups and the modification of chromatin associated with histone acetylation, the impact of metabolism and histone acetylation is fundamental to cellular processes such as gene expression, DNA replication, DNA repair, chromatin condensation, telomeric silencing, proliferation, and differentiation of cells such as the cells of the immune system [127]. The histone acetyltransferases (HAT) are known as transcriptional coactivators because they indirectly aid the transcription process within the cell by acetylating the lysine residues of histone proteins. These enzymes are essential within normal, infected, and oncogenic cells and can also acetylate other non-histone proteins in the Cell [128].

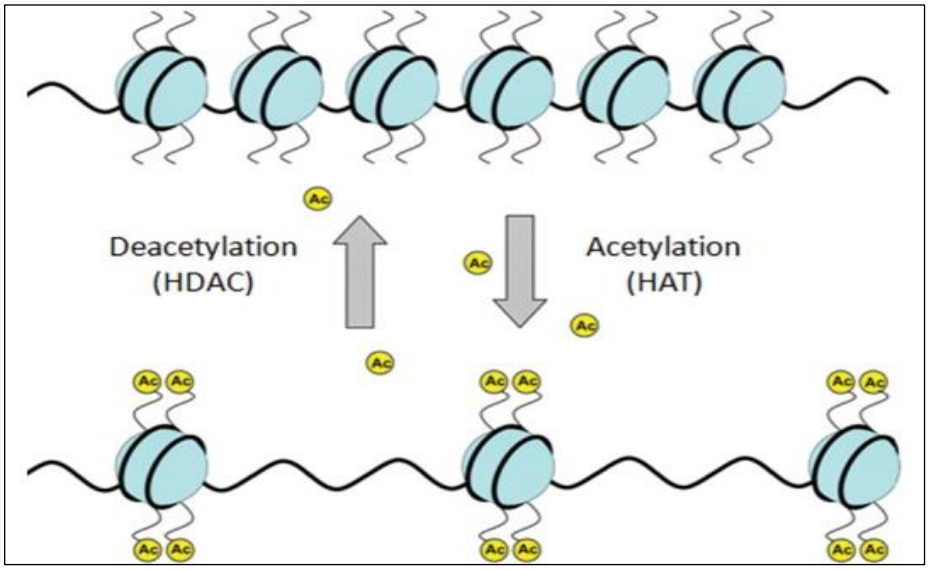


Figure 6: Effect of histone acetylation on chromatin structure. Histone acetylation in the nucleus relaxes the chromatin structure to activate gene expression [129]. Heterochromatin is a tightly condensed structure of DNA and histone protein, which is maintained by removal of acetyl group from the chromatin by histone deacetylation (HDAC), and histone acetylation (HAT) aid addition of acetyl group to relax the chromatin into euchromatin.

Nucleomodulins with histone acetyltransferase activity catalyze the addition of acetyl groups to the N-terminal lysine residue of the histone proteins; this epigenetic event usually changes the heterochromatin to a relaxed euchromatin structure that increases the localization of transcriptional factors and the accessibility of gene promoters [130]. To this effect, flagellin-dependent and T4SS-dependent global-wide histone acetylation was reported in *L. pneumophila* to enrich certain gene sets, including genes of cytokine activity, cytokine receptor binding, TNF receptor binding, and immune responses in the host through the TNFAIP2 promoter [131]. Flagellin protein was also found to contribute to the acetylation of histone H4 at Lysine 14 residue, leading to increased transcription due to enriched recruitment of RNA polymerase II at the IL-8 promoter, a

phenomenon that depends on p38 kinase and NF-κB pathways [132]. In a recent study, the localization of p300, a known histone acetyltransferase to the IL-8 promoter following infection with *Legionella pneumophila,* was shown to be a time-dependent event [133]. It has also been indicated that the methylation of H3K14 on residue 82 can tremendously inhibit the acetylation of specific histone residues [134].

Monica and Carmen 2014 performed an analysis on the genome of *L. pneumophila* to investigate effector proteins involved in the acetylation of histone H3 and H4 during infection. The author identified the SET domain as indicative of histone modification [135]. The first report that *L. pneumophila's* infection led to histone acetylation was published in 2008, and it was shown to be partly dependent on the presence of flagellin [136]. From a study by Bernd and colleagues, infection of lung epithelial A549 cells from a human with *L. pneumophila* induces genome-wide modification of histone with no specific reference to any effector protein. Significantly, the infected human cell line had elevated acetylation of histone H4 and H3 Lys (14). This modification is important for p38 kinase and NF-kappaB dependent transcription, which is essential for regulating proinflammatory gene expression in lung epithelial cells infected with *L. pneumophila* [136].

1.4. Protein of Study

LneB and MavA are hypothetical effector proteins in *L. pneumophila* [137]– [139] and the goal of this study was to characterize the two proteins using bioinformatics analysis. This analysis guided further examination of the biochemical properties of the LneB protein.

The LneB protein is encoded by the lpg1273 gene, has 354 amino acid residues, and predicted to have a molecular weight of 40.7kDa. Recent studies have shown that the LneB protein is regulated

at the transcriptional level LetAS-RsmYZ-CsrA regulatory cascade, expressed at higher levels during the stationary phase which is also responsible for the expression of other *Legionella pneumophila* /Dot/Icm effectors such as CegC1, LegA7, RavH, MavT, RavR, Lem11, LegL3, MavQ, Lpg0375, Lpg0963, CegL2, and Lpg2461 genes during infective growth phase [140].

The MavA protein is a member of the 'more regions allowing vacuole colocalization' (Mav) subfamily. The Mav subfamily was first identified by a study that showed that the subfamily proteins co-localize with the Legionella-containing vacuole during infection [141]. Members of the Mav protein family were named alphabetically from MavA to MavW. MavA is a 47-kDa Dot/Icm effector and a putative protein encoded by the lpg1687 gene of the *L. pneumophila,* Philadelphia strain [141].

1.4.1. Rationale for Hypothesis

The LneB and MavA proteins were investigated for their possible molecular roles in the bacterial pathogenesis using bioinformatics and transcriptomics analysis. Previous work in our laboratory showed that the LneB protein localizes in the host cell's nucleus. Based on our bioinformatics analysis, we hypothesized that the LneB may acetylate histone proteins in the human cell. We also hypothesize that MavA is a Ras-guanine exchange factor. The LneB protein may have a dual role in removing acetate from a substrate and adding it to another. This was strengthened by a prediction that LneB is predicted to have nucleosome complex-binding capacity with a sequence similarity with NAC domain-containing protein. The NAC (NAM, ATAF1/2 and CUC2) domains is a DNA-binding nuclear excision protein on which LneB shared 24% sequence similarity, which is, however, lower than the 30% threshold rule of thumb for sequence similarity [142]. Analysis through MotifScan also predicted HAT activity in LneB protein. The molecular processes of the

genes upregulated by LneB protein and LneB 33% homology with nucleoside diphosphate-linked moiety X (Nudix) protein. Nudix protein is known to be involved in the direct metabolism of acetyl-CoA and histone modification in the cell[143] and histone acetyltransferase (H4) activity [144]. Other predictions also showed that LneB protein might be involved in lysine histone demethylase activity. The demethylation of histone plays similar role as the histone acetylation in cell by leading to the activation of gene expression [146]. Transcription of genes is enhanced by histone acetylation/demethylation, which could lead to upregulating genes involve in hyper-inflammatory responses in the host in response to *L. pneumophila* infection [147].

1.5.Objective and hypothesis

The aim of this study is to contribute to the characterization of LneB and MavA effector proteins of *L. pneumophila.*

The null hypothesis (H_0) in this study is 'the LneB protein does not have histone acetyltransferase activity either as a stand-alone protein or when analyzed in vivo, $H_o: HAT_{(neg\ control)} > or = HAT_{(LneB).}$ The alternative hypothesis (H_a) is that the LneB protein has histone acetyltransferase activity, $H_a: HAT_{(neg\ control)} < HAT_{(LneB).}$

METHODOLOGY

2.1.Transcriptomics Analysis

We aimed at understanding the molecular pathways affected by LneB protein in the host cell, therefore, Human embryonic kidney (HEK) 293T (batch # 70032512), cultured in Dulbecco's Modified Medium (DMEM) containing 10% fetal bovine serum, was used as model. The pEGFP.C1 plasmid constructs previously constructed by Dr Banga were transfected into the cell line using Lipofectamine (Invitrogen). After 24 hours, the cells were collected, and RNA was extracted from the cells using the RNEasy kit (Qiagen). Afterward, the RNA samples were sequenced at the University of Louisville's Genomics core facility to identify differentially expressed genes by comparing mRNA transcripts from LneB-GFP, MavA-GFP, and GFP only transfected cells. Cluster Profiler was used to perform a functional annotation based on the differentially expressed genes followed by the gene ontology (G.O.) of the proteins coded by the genes [148].

2.1.1. Pathway Analysis of Differentially Expressed Genes

The differentially expressed genes were obtained, and a principal component analysis was conducted to show the uniqueness of the transcripts recovered from LneB-GFP, MavA-GFP, and GFP-transfected cells. The genes with p-value (statistical probability) less than 0.05 and q-value (false discovery rate) less than 0.05 are the significant differentially expressed genes (DEGs). The DEGs were analyzed and displayed as downregulated and upregulated genus using a volcano plot. The DEGs were also analyzed for the enriched gene ontology of the corresponding protein products for determining the biological process affected by the LneB and MavA proteins after

removing GFP influence. Furthermore, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was carried out to understand the potential molecular role of the two proteins.

2.2.Predictive Bioinformatics Analysis of LneB and MavA Proteins

Transcriptomics analysis, sequence and structural-based bioinformatics predictive analysis were carried out to characterize the structure and function of the LneB and MavA proteins using the pipeline described in Fig 7.

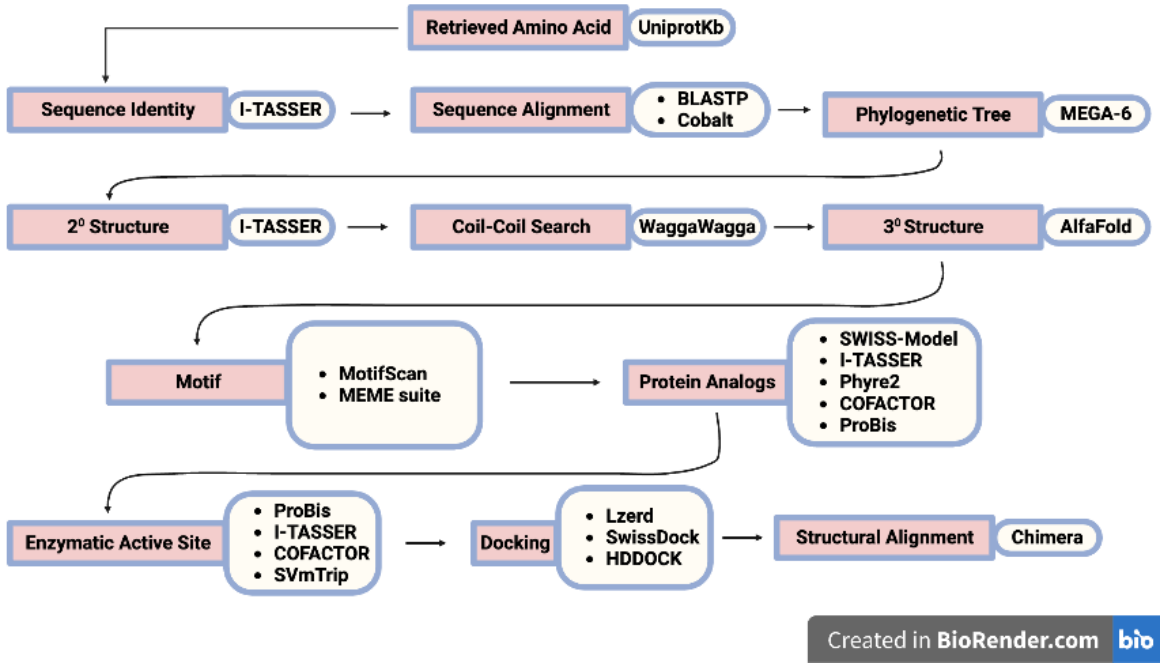


Figure 7: Bioinformatics pipeline for structural and functional characterization of LneB protein.

2.2.1. Evolutionary Analysis

To understand the evolutionary relationship between proteins on the NCBI database that shared sequence similarity with LneB and MavA protein. The amino acid sequence of the LneB and MavA proteins were retrieved from the Uniprotkb database with the gene I.D., lpg1273, and lpg1687, respectively, as a FASTA file. Similarly, matched protein sequences were identified using the local alignment (BLASTP) on the NCBI server to obtain the evolutionary neighbors of

the two proteins. Next, a neighbor-joining phylogenetic analysis was constructed using MEGA-11. Finally, OrthoDB was used to check the hierarchical catalog of orthologs of the most recent and distanced homologs from the phylogenetic tree.

2.2.2. Modeling the Structure and Function of LneB and MavA Protein

Using five intrinsic pipelines to combine the fast homology search of MMseqs2 with ColabFold (a server-based version of AlphaFold2) [149], an optimized prediction of LneB and MavA tertiary structures was made. ColabFold is MMseqs2 profile databases that generate diverse multiple sequence alignments to predict protein structures [150]. First, distinct models were searched for the LneB protein using a model-building probe on SWISS-MODEL that also shows predicted protein homologs [151]. Additionally, to locate structural templates from the PDB database, I-TASSER modeling LneB protein using LOMETS (a meta-server threading technique with ten template alignments). The LOMETS gave a quick predictions of protein tertiary structures and spatial constraints of the structure. The LneB and MavA proteins' tertiary structures were also modeled using Phyre2, a web-based remote homology server, with heuristics that maximized confidence, percentage identity, and alignment coverage [152]. Finally, the top-ranked structural model of LneB and MavA proteins from the ColabFold was used to predict the function via COFACTOR and the ProBiS servers, respectively. To predict likely functions, COFACTOR combines the properties of the LneB proteins, such as the amino acid sequence, likely protein-protein interactions, and predicted structure, and identifies proteins with similar structures, gene ontology, biological processes, cellular components, and protein homology [153]. The predicted functional motifs of the LneB protein were identified using MotifScan. Also, ProBiS found potential binding sites for small molecules, proteins or nucleic acid in each of LneB and MavA proteins in the database [154]. Next, the predicted characteristics from MotifScan, I-TASSER,

COFACTOR, ProBiS, and Phyre2 were combined. Finally, the possible interaction of MavA and LneB protein with some predicted ligands from the bioinformatic and transcriptomics data was probed with protein-protein docking tools, including ProteinPlus, and SWISSDOCK.

2.3.Biochemical Experimentation

2.3.1. Expression and Purification of LneB Protein

Previously, Dr Banga tagged the LneB protein with 6x histidine at the C-terminal and cloned downstream of the IPTG inducible promoter of the pET-21a plasmid vector. Through a heat shock protocol, the BL21 strain of *Escherichia coli* was transformed with the plasmid prior to my joining the lab. Successful transformant were recovered on a plate containing 100mg/L of ampicillin antibiotics. Transformed *E. coli* was grown in the presence of ampicillin to an optical density of 0.6 at 600nm wavelength, followed by induction of LneB expression with IPTG (Invitrogen UltraPure IPTG, Lot 27950522). Cells were incubated on a shaker (100rpm) in 1 Litre conical flask at 18^{o}C and overnight incubation. The bacterial culture was pelleted by centrifuging at 3000 rpm (1549 RCF) for 10 min at 4^{0}C, and the pellet was kept at -80^{o}C for further analysis. (See detailed protocol in Appendix 01).

2.3.1.1.Purification of LneB Protein

The pellet of bacterial culture expressing LneB-HIS protein was placed on ice throughout purification steps to avoid protein denaturation. The bacterial cells were lysed with previously prepared lysis buffer containing lysozyme and protease inhibitor (2 mM sodium azide, 500 mM NaCl, 5 mM imidazole and 20 mM Tris-Cl, pH 7.9). After vortexing, the lysed cells, were subjected to sonication at a pulse-on of 10 seconds and 20 seconds rest at an amplitude of 50%, probe size 12mm. The cell lysate was centrifuged at 15000g for 40 minutes at 4^{o}C to remove the

cell debris (pellet) and retrieve the crude LneB-HIS protein. The crude lysate was filtered through a 0.22µm filter from which a small fraction (50µL) was taken at each purification step for subsequent analysis by SDS-PAGE. The filtered crude protein sample was passed through a column containing Ni-NDA resin beads (HisPur Ni-NTA purification kit, Lot XB333025) for purification by affinity chromatography. The column was washed thrice to remove unbounded protein on the beads. LneB-HIS protein was eluted thrice using elution buffer (2 mM sodium azide, 500 mM NaCl, 500 mM imidazole and 20 mM Tris-Cl, pH 7.9) (see Appendix 02 for detailed steps and buffer constitution).

2.3.2. SDS-PAGE

After purification, protein samples were separated by mass using sodium dodecyl-sulfate polyacrylamide gel electrophoretic analysis. A stock solution of 30% acrylamide was used to create a12% resolving gel and 5% stacking gel and run at 100 V for 15 min and 130 V for 60 minutes. Briefly, an equal volume of protein sample was boiled at 100^{o}C for 5mins with 2X SDS-loading buffer and centrifuged at 14000 rpm for 3mins to pellet cell debris before loading 5µL of the denatured negatively charged protein into the wells of the stacking gel. Protein molecular marker was loaded on a separate well. Next, the gel was stained with 10mL of Coomassie blue (BIO-RAD Coomassie Brilliant Blue R-250, Cat. #1610436) for 4 hours on a shaker at room temperature. The gel was de-stained twice each for 10 minutes with de-staining solution containing methanol and glacial acetic acid to observe the protein bands on the gel. Finally, the de-stained gel was viewed under the white translucent light of the gel imaging machine (Protein simple) at the Biotechnology center of the Department of Biology.

2.3.3. Western Blot

Following protein separation by weight on 12% SDS PAGE gel, the gel was placed on the BIO-RAD nitrocellulose 0.45µm, 7 × 8.5 cm (Cat. #1620145) that was pre-wet with transfer buffer (5.8% w/v Tris Base, 2.9% w/v Glycine, 0.37% w/v SDS, and 20% v/v methanol). The setup of the blotting process was done as outlined in Appendix 04. After protein transfer from the gel to the membrane at 100V for 1 hour on ice, the membrane was blocked with blocking buffer (2% non-fat milk, PBS, 0.5% Tween20) at room temperature. The anti-HIS tag primary antibody was diluted at ratio 1:5000 in an appropriate volume of blocking buffer. The specific antigen on the membrane was detected by the diluted primary antibody for 1hr at room temperature on a rocker. The membrane was washed five times with washing buffer (10%PBS.1% Tween20) before the addition of diluted secondary antibody (Anti-mouse IgG, H + L, horse-radish peroxidase-conjugated antibody, Lot XH359897) for one hour at room temperature (see Appendix 04 for detailed protocol). Secondary antibody dilution was done at a ratio of 1:20000 in blocking buffer. Then, the membrane was washed in washing buffer for 5 mins and incubated at room temperature in respective ECL substrate before viewing with a chemiluminescent imager at the Biotechnology center of the Department of Biology, Western Kentucky University.

2.3.4. LneB Protein Concentration and Estimation

Purified LneB protein was concentrated using a Pierce protein concentrator, PES (Lot XC346197), 10K grade to remove proteins of lower molecular weight than a 40kDa and 30k grade to remove proteins higher than 50kDa in the eluent. Next, the protein concentration was estimated using the Pierce BCA protein assay kit (Lot XE345845). Finally, the protein was aliquoted for western blot and HAT activity (see Appendix 05 for a detailed step).

2.4.Infection Assay of *L. pneumophila* in U937 Cell

The effector proteins are essential factors for the survival of *Legionella pneumophila* in human cells. An alteration in the Dot/Icm system transport mechanism may hinder the release of the effectors during infection, consequently reducing bacterial intracellular multiplication. Therefore, the effect of the Dot/Icm substrates (effector proteins) was investigated through an infection assay in U937 cells.

Initially, *L. pneumophila* mutant LP03 with a defective Dot/Icm transport system and the wild type, LP02, were grown in AYE broth at 37^0C for 24 hrs to an optical density of 0.3 – 0.4 at 600nm. The U937 cell line was commercially purchased and stored in a lyophilized form under liquid nitrogen). Before performing cell-based assays the biological safety cabinet was sterilized with ultraviolet light for 15 minutes, and the media was warmed to 37^0C. The lyophilized cells were thawed and revived in Roswell Park Memorial Institute (RPMI) medium (RPMI 1640, 1X with L-glutamine and 25 mM HEPES, Lot 14117005 from Cellgro) with 10% fetal bovine serum, FBS. The reconstituted cells were incubated under 5% CO_2 at 37^0C. The cells were passed thrice each on a 3-day passage. The U937 cells were differentiated into adhering cells using 5µL 12-O-tetradecanoylphorbol-13-acetate (TPA). Differentiated cells, which becomes adherent were detached and counted before plating the cells. 500µL of the cell suspension containing 1 x 10^6 cell/mL was transferred to each well of sterile 24-well plates and incubated in a 5% CO_2 incubator at 37^0C for 24 hours. The following formula was used to obtain the cell count/mL.

- Average cell count (A) = $\frac{C1+C2+C3+C4}{4}$.
- Cell count per mL (Cc) = $\frac{1}{5}$(A x 10^6)

To obtain the proper bacterial count for the desired multiplication of infection (MOI) of 0.05, both mutant and wild-type *L. pneumophila* were grown to an O.D. range of 0.3 to 0.4. Bacterial count was derived with; bacteria cell count (Bcc) = D x 10^{10} where D is the optical density.

The suitable volume of bacterial culture corresponding to a cell density of 2.5 x 10^4 cells/mL and multiplicity of infection (MOI) of 0.05 bacterial per infected cell was calculated and transferred to each corresponding labeled 24 well plate. Upon centrifuging the plate at 1000 g for 5 minutes at room temperature, the infected cells were incubated for 2 hours at 37^0C under CO_2 to encourage bacteria uptake. The wells were washed with warm DPBS (Gibco, DPBS 1X without calcium chloride and magnesium chloride, Lo 2235070) thrice, each time with a volume of 500µL. Finally, at 24, 48 and 72 hours, 500 uL of fresh RPMI media (with 10% FBS) was added to remaining wells. The cells were collected from each well of each timepoint and lysed with 2% saponin and plated on sterile CYE plates for bacteria colony counts after incubation at 37^0C for 48 hours.

2.5.Histone Acetylation in Infected U937 Cells

Based on the bioinformatics prediction that LneB has histone acetyltransferase activity, the investigation of histone acetylation due to Dot/Icm substrates (effector proteins) became part of the objective. U937 was infected with Legionella at a MOI of 10. Infection was followed by extracting nuclear proteins from the infected cells, upon which the HAT assay was measured. The same procedure in section 2.4 was followed, except for changing the infection time points from 2, 24, 48 and 72 hours to 2, 6, and 18 hours. The nuclear extraction was done using the NE-PER Nuclear and Cytoplasmic extraction kit (Thermo Scientific, Lot UH288241A), according to the manufacturer's instructions. Simultaneously, bacteria uptake was measured on a separate culture plate with the same bacteria and cell culture using a MOI of 0.05 in U937 cells and allowing infection to proceed for 2 hours before lysing the cells. The uptake analysis was done to ensure

proper uptake of bacterial cells by the cell. Upon nuclear extraction, the concentration of the protein extracts was determined before HAT analysis. The nuclear extract samples from the infected cells were diluted to a 1.25µg/µL concentration to meet the specification of the kit. The EpiQuick HAT Activity/Inhibition Assay Kit by Epigentik was used to carry out the histone acetyltransferase activity assay plates were read at 440 nm wavelength. HAT activity was expressed relative to the O.D. value per µg or nmol/min/µg of protein tested. The histone acetylation activity of the nuclear extracts was further analyzed using the ANOVA test at a 95% confidence level to determine the overall significance difference among the sample and subsequent multiple pairwise comparisons between samples using the Tukey multiple comparison tests in R studio.

2.6.Histone Acetylation in Transfected HEK 293T Cell (See Appendix 06)

To understand the effect of LneB protein in human cells, HEK 293T cells were transfected with a plasmid containing LneB-GFP protein or GFP only as a control. (The complete technique is in Appendix 06). Briefly, Dulbecco's modified minimum Eagle's media (DMEM) supplemented with 10% fetal bovine serum was used to cultivate human embryonic kidney (HEK 293T) cells in tissue culture dishes (100 mm × 20 mm) at 37 °C and 5% CO2 (ref # 10013CV). Transfection was performed using 5 x 10^6 cells/mL of HEK 293T in a 100 mm x 20 mm culture dish. The plasmid vector was cloned with LneB-GFP and GFP which was transfected into 70 to 90% growth confluent HEK293T cells aided by Lipofectamine 3000 (Invitrogen) (ref# L3000-015) following the manufacturer's instructions. LneB-GFP and the GFP-only constructs were utilized for transfection separately, and a fluorescence microscope was used to visualize the results after 24 hours. First, the cells were collected, then nuclear extraction was performed, followed by histone acetylation of the extracts.

RESULTS

Project One: LneB Protein

3.1.Transcriptomics Analysis

3.1.1 Principal Component Analysis of Differentially Expressed Genes

HEK 293T cells were transfected with a plasmid vector containing the LneB protein gene tagged with GFP or the vector containing GFP alone. RNA was extracted and sequenced to identify differentially expressed genes. The identities of the transcribed genes were transformed from ENSEMBL ID to Gene ID of *Homo sapiens* genomic library using Biotools [155] and analyzed for Principal Component Analysis (PCA) using Qlucore Omics Explorer 3.5 (Lund, Sweden). The three-dimension PCAs (PCA1, PCA2, PCA3) account for more than 90% of the total variation of the genes expressed in LneB-GFP and GFP transfected cells (Fig. 8a).

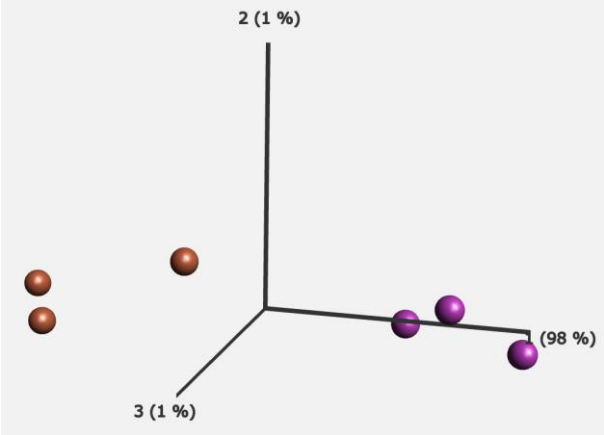


Figure 8: Variation in DEGs by LneB. Principal Component Analysis (PCA) of genes expressed in triplicate from HEK293T cells by LneB-GFP (purple) and GFP-control (orange). The highest

linear possible variance is shown on PC1, which is 68%, without the effect of the other PCs. The remaining variance is represented by PC2, with a value of 20%, and PC3 with a value of 8%.

3.1.1.1.Differentially Expressed Genes by LneB-GFP Protein

Differential expression of genes by LneB compared to the GFP only control in HEK 293 cells showed the upregulation of genes encoding for dehydrogenase/reductase and early growth response by LneB protein (Fig. 9). The genes are involved with energy and hormone metabolism, respectively (Table 3).

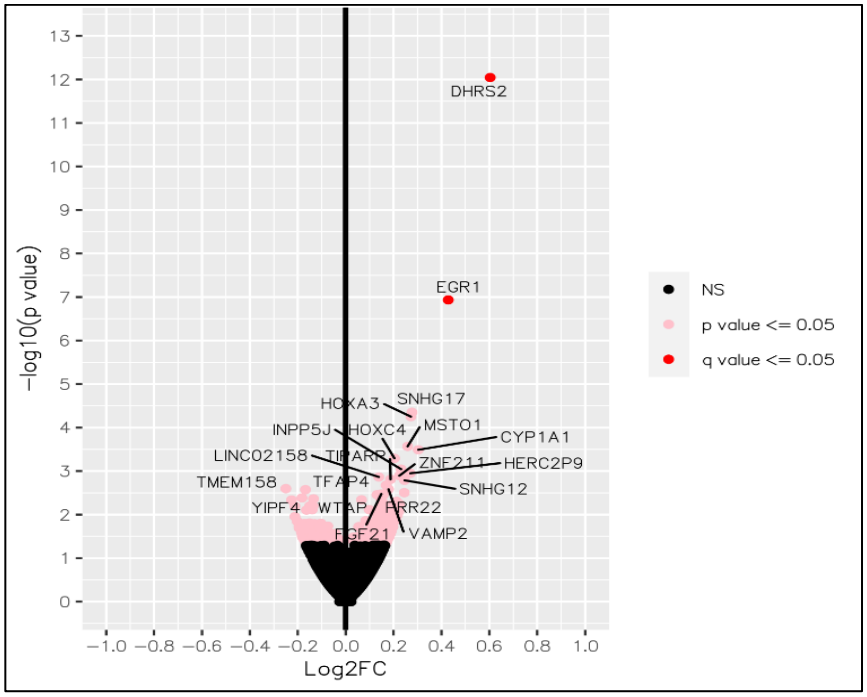


Figure 9: Volcano plot of the transcriptome analysis. This figure represents differentially expressed genes HEK 293T cells expressing LneB GFP-tagged protein compared to those expressing GFP alone. The x-axis represents fold change on a logarithmic scale of 2. The upregulated genes show a positive fold change (the right side). The y-axis represents the -log10 of the p-values. Genes with significant q-values (false discovery rate) are shown in red, and genes

with low significance are shown in pink. Fourteen genes showed significant p-value. DHRS2 and EGR1 had significant false discovery rates (q-value).

Table 4: Differentially Expressed Genes by LneB protein in HEK 293 cell.

Ensembl ID	Gene Symbol	Gene Description	Log_2FC	P-value	Q-value
ENSG00000100867	DHRS2	Dehydrogenase/reductase 2	1.52 ↑	9.0E-13	2.09E-08
ENSG00000120738	EGR1	Early growth response 1	1.35 ↑	1.16E-07	0.001344

3.1.1.2.Pathway Analysis of Differentially Expressed Genes

Pathway analysis of the implicated genes showed that the LneB protein potentially affects the metabolism of steroids and other hormones in the cells in addition to leukocyte differentiation. The pathways potentially affected by the LneB protein in the cell also suggests that the protein is likely toxic to human cells (Fig. 10) or could be involved in regulating protein kinase, histone acetyltransferase, and oxidoreductase [156]. According to the predicted gene ontology, the significant pathways that may be impacted are the biosynthesis of hormones and essential cellular immune biomarkers such as interleukins and chemokines (Figure 10).

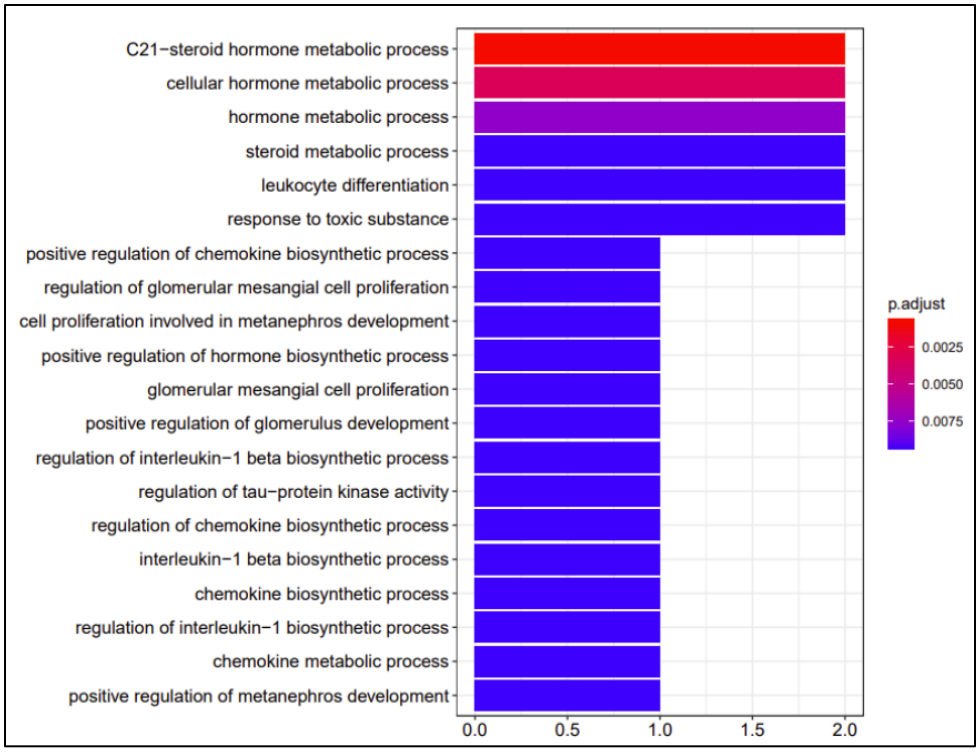


Figure 10: Pathway analysis of differentially expressed genes by LneB protein in HEK 293T cells. The top 20 identified by gene ontology of the biological processes affected by LneB protein. The hormone metabolic process is the highest suggested pathway. Other processes, including protein kinase and biosynthesis of other immune responses such as chemokines and interleukins are also represented.

Other potential molecular interactions of the LneB protein in the transfected HEK 293T cell were also identified on KEGG pathways, which indicate that LneB affects hypothalamic gonadotropin-releasing hormone (GnRH) in the host (Figure 11).

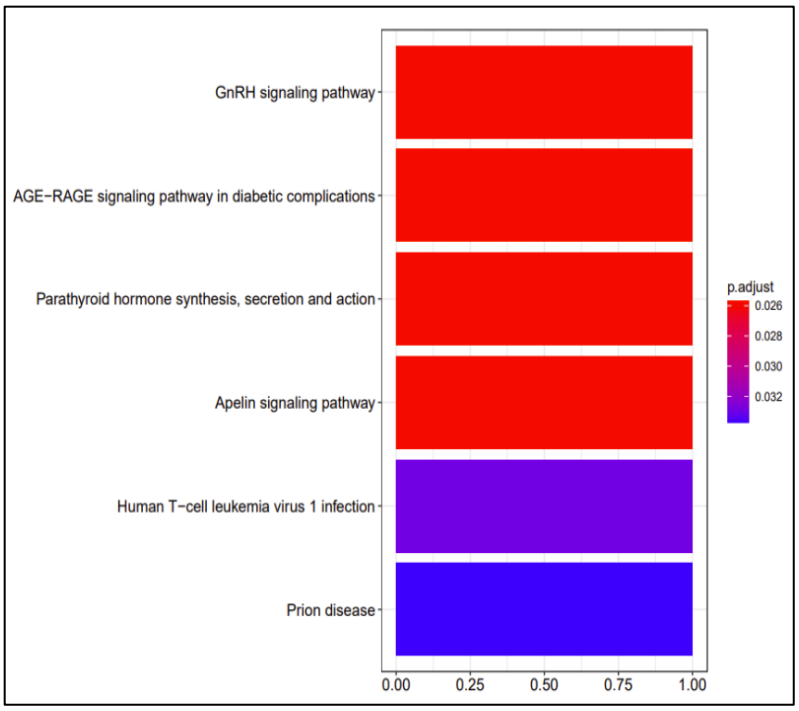
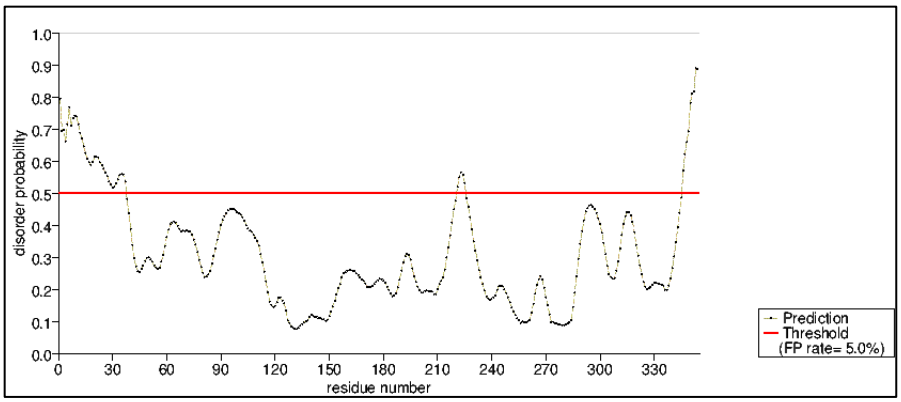


Figure 11: Top 6 enriched significant KEGG pathways for LneB protein. The endocrine system was highly affected by LneB in HEK293T cells with secretion and activity of gonadotropin-releasing hormone (GnRH) and parathyroid hormone system. Other pathways include diabetes complications and T-cells implicated in HIV infection.

3.2.Bioinformatics Prediction of the Properties of LneB Protein

3.2.1.1.Sequence Analysis

The amino acid sequence of LneB protein was retrieved from Uniprotkb and analyzed with free open-source bioinformatics servers for potential annotation of the protein's properties. The outcome of the analysis from PredictProtein [157] shows that the LneB protein is predicted to bind to DNA using its N-terminal domain. The disordered regions of LneB protein was predicted to be at the C and N-terminal by PrDOS at false discovery rate of 0.05 and disorder probability greater than 50% from residue 1 – 47, 221 – 225 and 346 – 354 as shown in Figure 12.




```
1   MKIFKGLKKN  TGFRKQSNTQ  KISNSSIQSS  ITSLSGDASQ  PRFDFFPINN  50
51  RDVLLKSPSF  NHIGSEFQPM  VSKRDVNSIS  ITLPIKHNVH  RNHNSNLPVD  100
101 SISVVDSSNQ  QDHISYLHIC  FTHEEGKWYL  FLKCFTTPDV  FNYDFAPLLK  150
151 ILHKPDIRTY  PMSFGYPGVI  DRTELIDKDR  KIYARQFIIH  GPNDEAKILQ  200
201 DVHDFNLNCL  RRFEELTYGR  DNSAYRNSQA  LGRILHKIVT  AYQEVTVNEY  250
251 DIALNRLKIV  FQDQVFQNDP  QICLSLTHIV  NALIVLSQKK  ELNNKGIEMD  300
301 EHQLHLILKE  LAKHEGKKFK  NNLWFPVGTV  ETISQEVDEL  FDLKQTTAAA  350
351 SIMS                                                      400
```

Figure 12: Predicted disordered region on LneB protein. The amino acid sequence of LneB protein was used to predict the disorder regions on the LneB protein using PrDOS. The disordered region is predicted to be at the tail ends of the C and N terminals and at residue 221 – 225.

However, the protein is not predicted to bind to other proteins or RNA, but the N-terminal is likely to bind to DNA (Fig. 13). The LneB protein is predicted not to have disulfide interaction. Highly conserved regions are predicted across the protein from its N to C-terminals, but the C-terminal region has more predicted conservation score (Figure 13). In addition, using the PROTEUS2 tool [158], the LneB protein has no suggested transmembrane domain nor signal peptide. The LneB protein is predicted to have more helix secondary structure at the C terminal. More of the protein residues are predicted to be exposed to the protein surface.

Figure 13: Secondary structural analysis prediction of LneB protein by PredictProtein. LneB amino acid sequence was analyzed with PredictProtein. RePROF was used for secondary structure, and solvent accessibility, TMSEG for topology, ProNA for protein, RNA and DNA binding and

DISULFIND for predicting disulfide bond. The annotation with high confidence is colored yellow, red is for intermediate confidence and blue for low confidence.

3.2.2. Structural Analysis of LneB Protein

The tertiary structure of the LneB protein was predicted using Alfafold2-ColabFold [159] server from the amino acid sequence of the protein retrieved from the Uniprotkb server [160]. The best model has a predicted local distance difference test (pLDDT) score of 0.658 and an expected error pTM score of 0.518. The pLDDT score of 0.658 predicted that the superposition score of the differences between the local distance of all atoms in the model is good. The 3D structure comprises seven alpha-helices and six antiparallel beta-sheet structures, with coil-coil joining individual beta sheets (Fig. 14).

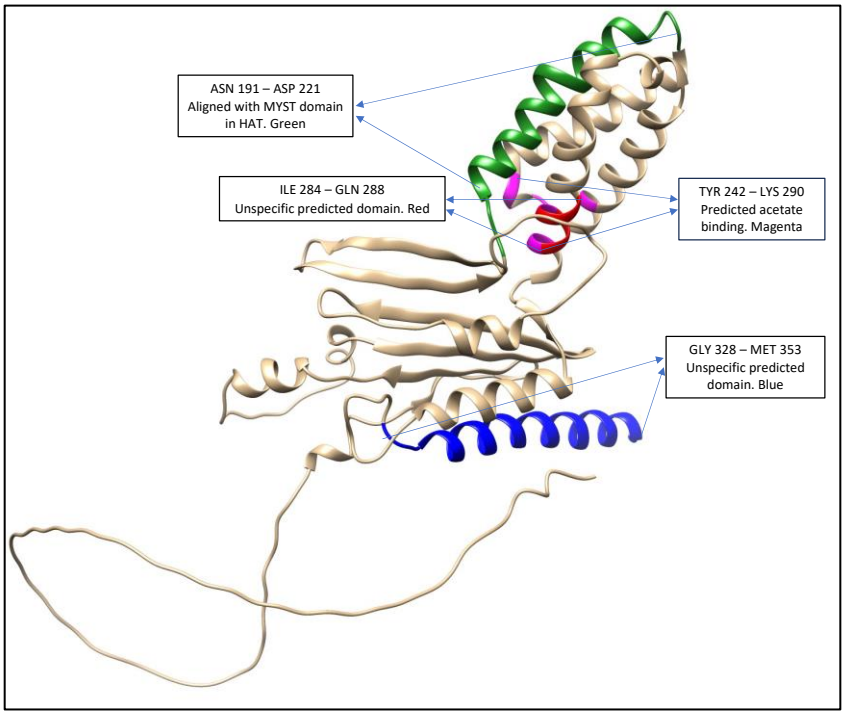


Figure 14: Predicted Conserved regions on LneB Tertiary Structure. Labelled prediction of the conserved regions on the LneB protein structure from sequence alignment, Phyre2 and MotifScan.

Aligned region with MYST domain in human HAT enzyme is shown in green, the region that is predicted to bind to acetate is shown in magenta and two other unspecific domains are predicted as red and blue.

Subcellular Localization of LneB Protein

The LocTree3 program of the PredictProtein server predicted sub-cellular localization of LneB protein in the nucleus of the host cell with a prediction confidence score of 52% and a G.O. term ID of 0005634 using the protein amino acid sequence (Fig. 15). The confidence score shows the level of confidence of the subcellular localization prediction, which is slightly above average. The nuclear localization of LneB is supported by previous work in our lab.

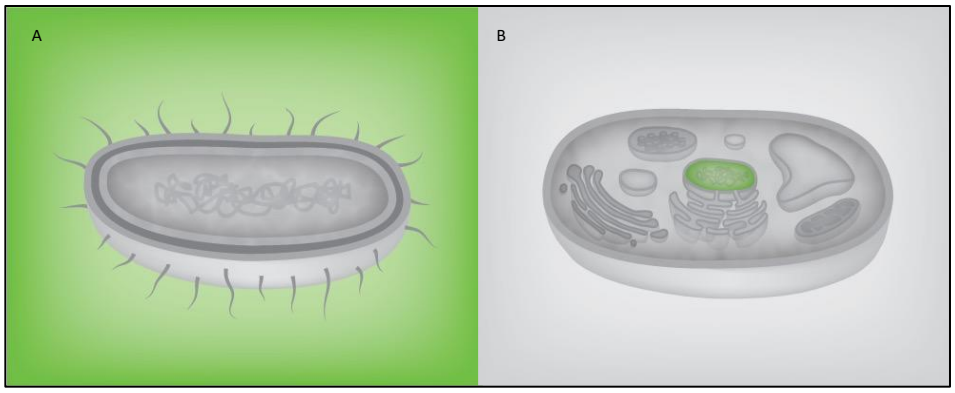


Figure 15: Predicted subcellular localization of LneB protein in (A) Bacteria extracellular space. (B) Eukarya host nucleus.

3.2.2.1.Function and Domain Prediction of LneB Protein

A functional motif of the LneB protein was predicted using MotifScan serve. The motif is a histone acetyltransferase activity repeat at the C-terminal of the protein (position 174 – 206) (Fig. 16). The program Phyre2 also predicted the same terminus to be responsible for transcription by LneB protein in cells.

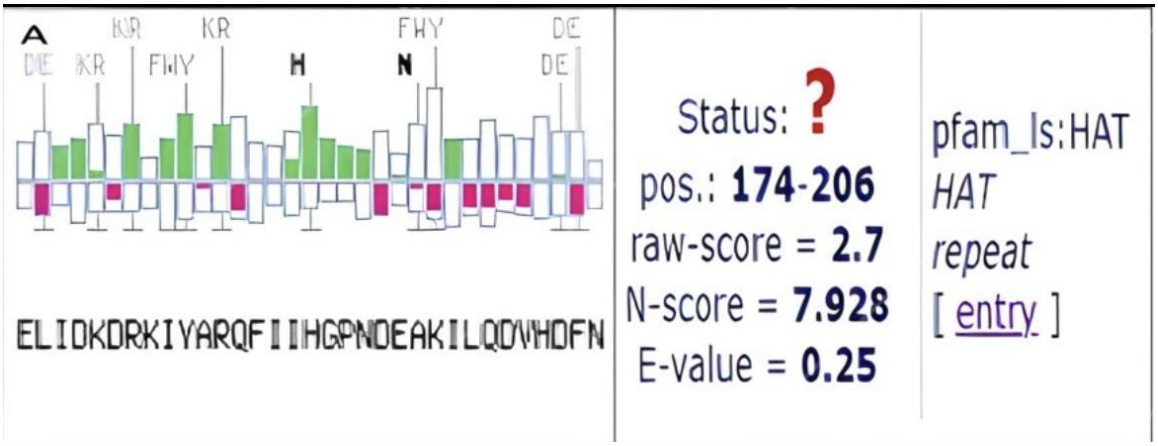


Figure 16: Predicted histone acetyltransferase motif (HAT) in LneB protein. The MotifScan server predicted a HAT motif in the LneB protein at position 174 – 206 with a raw score of 2.7, reflecting low sequence similarity with known proteins. However, the expectation value (0.25) is higher than 0.1, showing less evidence for the prediction.

The functional predictions of LneB protein using six servers were merged with the findings from the transcriptomics analysis for robust probable annotations for LneB protein. These annotations are classified under six categories because each server gave a different prediction, though nucleus and HAT activity are related. The most common annotation is the acetylation of histone and other molecular events surrounding the histone such as histone acetylation, acetyltransferase, histone dimerization, transcription initiation factor, acetate ion binding, and lysine histone demethylase (Fig. 17).

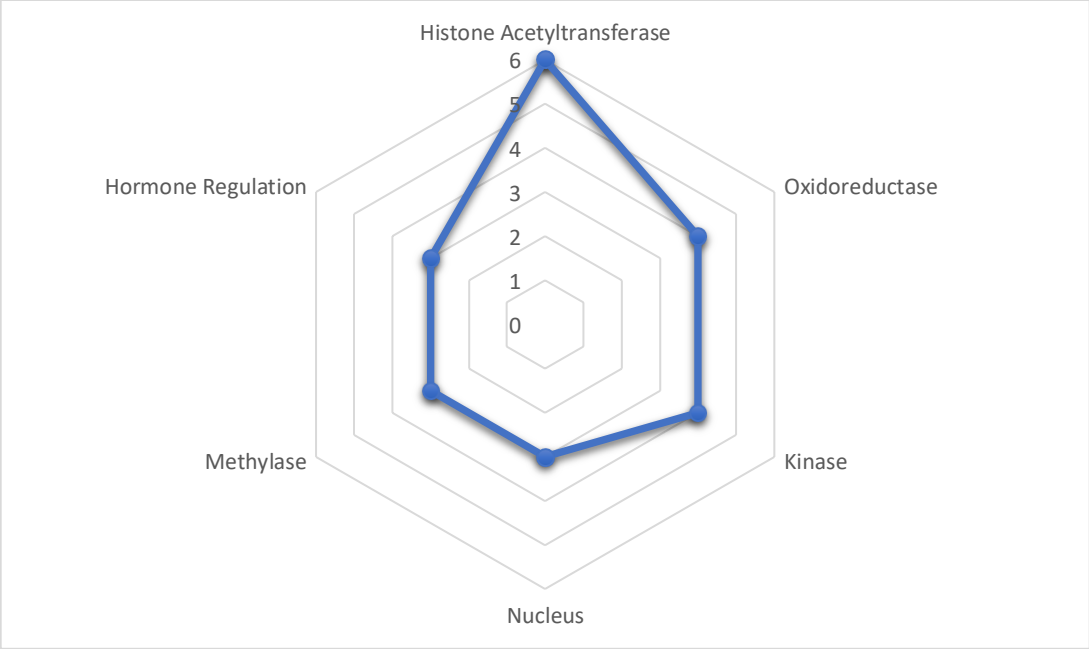


Figure 17: Radar Frequency showing combined annotation of LneB protein by six bioinformatics tools at each chat vertices and transcriptomics data. The result show that LneB protein is more likely to be involved in post-translational histone protein modification, primarily acetylation.

3.2.2.2.Multiple Sequence Alignment of LneB Protein with HAT Enzymes

Based on the prediction of histone acetylation of LneB, further investigation was carried out on the similarity between the amino acid sequences of LneB protein and human histone acetyltransferase using the ClustalW multiple sequence alignment tool [161]. This analysis showed there are 67 out of 354 (19%) conserved residues in LneB compared to aligned with the MYST domain of histone acetyltransferase enzymes in humans with a coverage of 3.2%, as shown in Fig. 18.

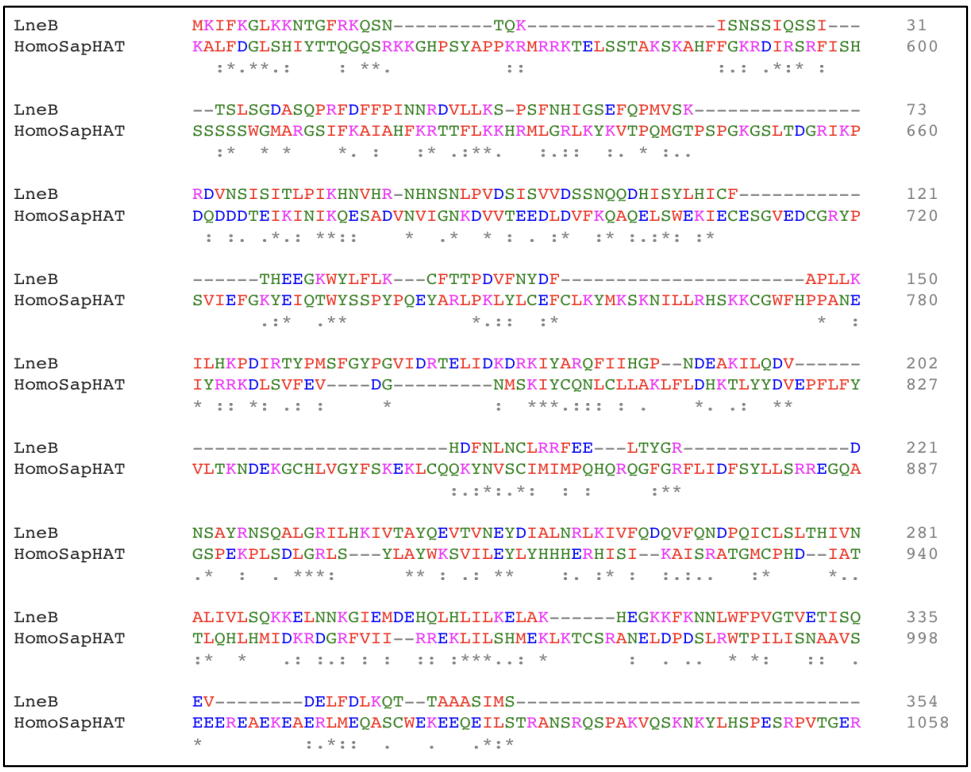


Figure 18: Sequence Alignment of LneB Protein and HAT enzyme of *Homo sapiens.*

Evolutionary Analysis was also carried out with amino acid sequences of LneB protein and other known histone acetyltransferase proteins. Five histone acetyltransferase proteins were selected from *Homo sapiens, Dictyostelium discoideum, Mycobacterium tuberculosis,* and *Mycobacterium pseudoshottsii* and were aligned with LneB protein. The neighbor-joining phylogenetic tree shows that the histone acetyltransferase from the *Dictyostelium discoideum* (Amoeba) is the closest ancestor to LneB protein, while the HAT from *Mycobacterium tuberculosis* is the most distant ancestral to the LneB protein (Fig. 19).

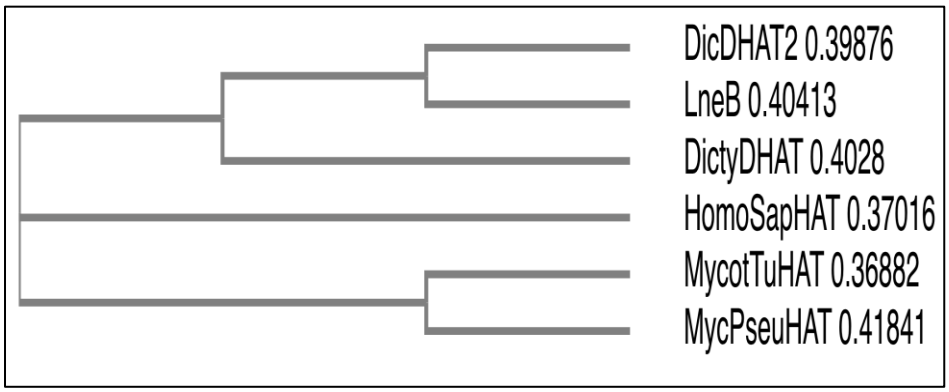


Figure 19: Neighbor-Joining Phylogenetic Tree of LneB Protein and Other Histone Acetyltransferase Protein. Amino acid sequence of HAT proteins from amoeba, *M. tuberculosis* and human was aligned using multiple sequence alignment tool.

To further analyze the relationship between LneB and HAT protein in Amoeba, pairwise sequence alignment was carried out. The identity between the protein sequence is 69/475 (14.5%), similarity is 128/475 (26.9%) and there is a 217 (45.7%) gap in the alignment (Fig. 20).

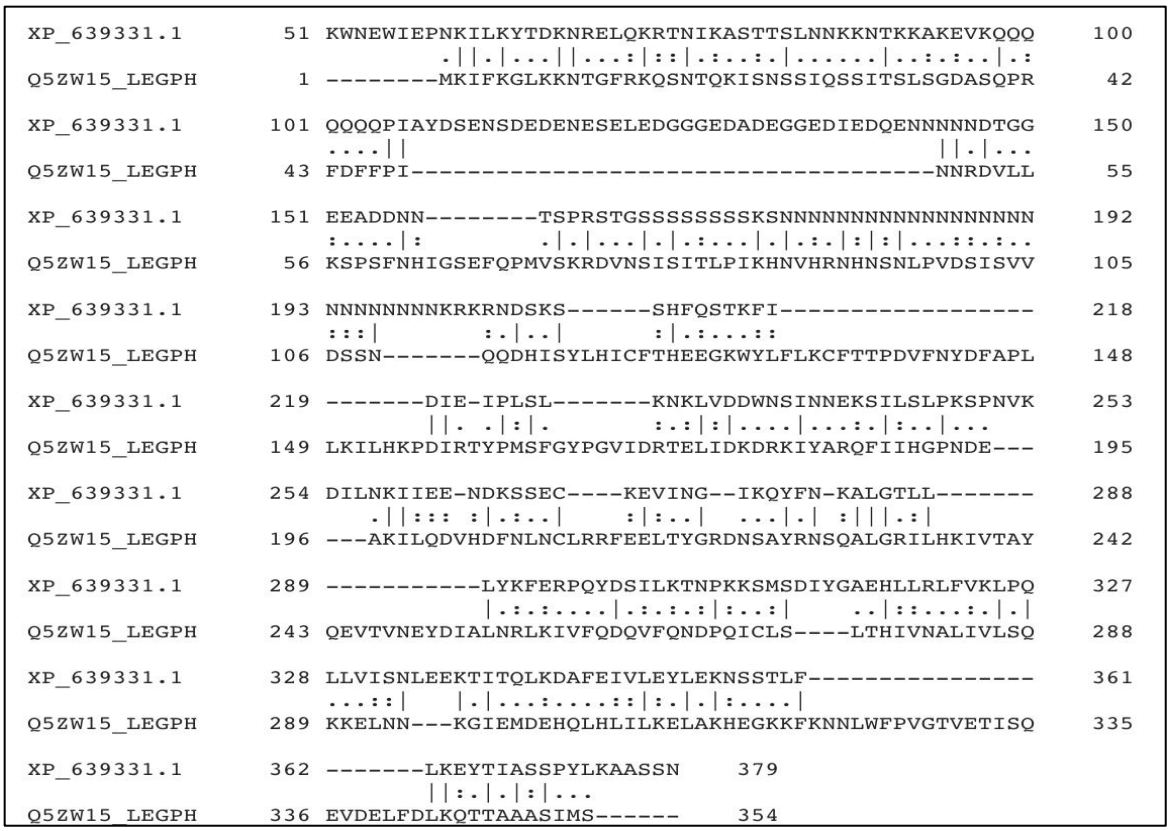


Figure 20: Pairwise sequence alignment between LneB and histone acetyltransferase of Amoeba retrieved from Uniprotkb. Stroke | means aligned residue; double dot means a more conserved properties than the single dot.

3.2.3. Molecular Docking

3.2.3.1.LneB Interaction with Histone

To simulate the possible molecular interaction between LneB protein and histone proteins in humans, molecular docking of histone 2A/2B, 3 and 4 complexes from *Homo sapiens* with the best-ranked tertiary structure of LneB protein was carried out. This was supported by docking protein that is known to interact with histone, human nucleophosmin (HPM) as positive control and green fluorescent protein (GFP) as negative control [162]. The structures of the histone, HPM and GFP proteins were retrieved from the protein data bank, RCSB [163]. The Lzerd molecular

docking server [164] was used for the protein-protein docking, the parameters of the server were set at default. The lower the value of the GOAP, DFIRE and ranksum in the reported model, the better is the model. The GOAP is a generalized orientation-dependent all-atom potential is a measure of statistical potential that depends on the relative orientation of the associated planes of the heavy atoms between interacting structures [165]. The DFIRE is distance-scaled, finite ideal-gas reference state score that predict the potential of mean force of all the atoms in two structures from a nonhomologous database of protein structures less than 2 Å [166]. The smaller the rank-sum score, the better is the predicted interaction.

The score reported between LneB, and histone is lower than the score from HPM and higher than the scores from GFP interaction with histone (Table 5).

Table 5: Statistical scores of predicted interactions of histones with GFP, LneB and HPM proteins.

	GFP		**HPM**		**LneB**		
	GOAP Score	**DFIRE Score**	**GOAP Score**	**DFIRE Score**	**GOAP Score**	**DFIRE Score**	**Ranksum**
Histone 2A/B	-66706	-38057	-97741.26	-60375.64	-70791.91	-43848	786
Histone 3	-51431.64	-29903.95	-82417.58	-52244.16	-55681	-35783.67	68
Histone 4	-49914.05	-28175.04	-80468.69	-50753.76	-54528.43	-34159.5	73

The GOAP and DFIRE scores of the interaction between HPM and histone is greater than the interaction between GFP with histone. The best model of the interaction was predicted between LneB protein and histone 2A/2B (Fig. 21a).

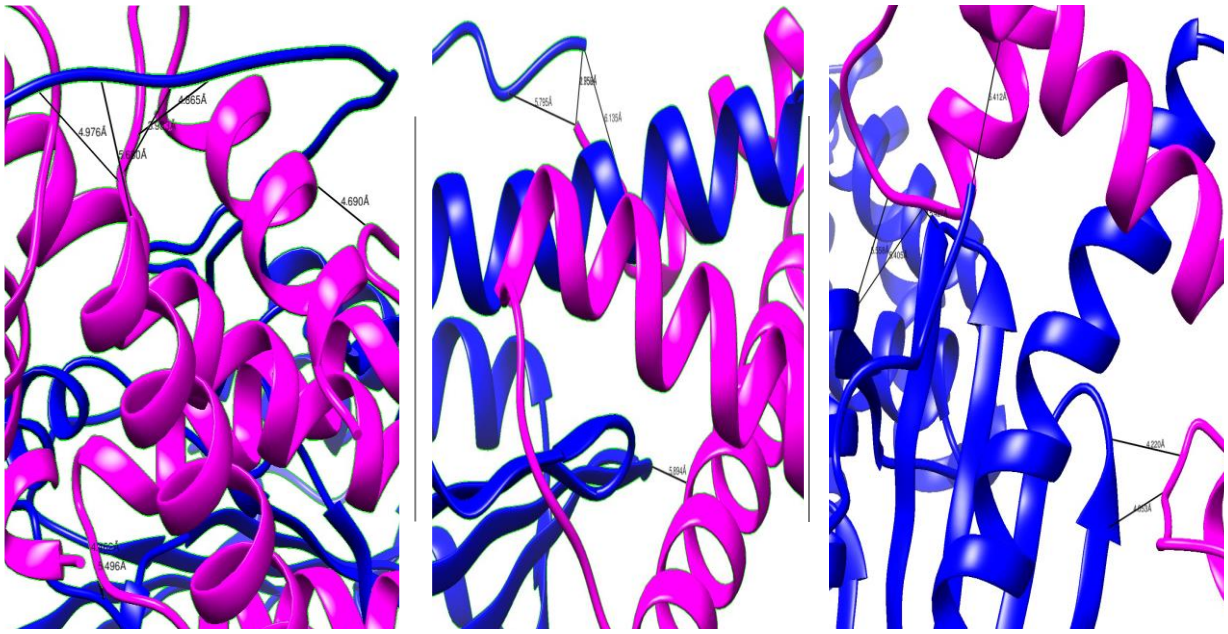


Figure 21: Simulated interaction between LneB protein and human histone 2A/B, 3 and 4. A: docking LneB (blue) with histone complex 2A and 2B (magenta). Interaction is predicted to be at residues ILE 54, PHE 45, SER 55, SER 56, ILE 102 at RMSD 3.983, 4.865, 4.976, 5.68 and 5.496 respectively. B: Docking LneB (blue) with histone complexes 3 (magenta). Interaction is predicted to be at residues PHE 4, and GLU 125 at RMSD 5.795 and 5.894 respectively. C. Docking LneB (blue) with histone complexes 4 (magenta). Interaction is predicted to be at residues GLY 43, VAL 76, TYR 99, GLU 301 with RMSD 5.427, 4.228, 4.053, and 5.405.

3.2.3.2. LneB Interaction with Acetate

We also examined the potential interaction between LneB protein and acetate ion using the CB-Dock server [167]. The simulation showed potential interaction between acetate and LneB protein, GFP (negative control), and acetate kinase (positive control). The acetate ion is predicted to bind to LneB protein via hydrogen bond most probable at the protein's C terminal. The best interaction

model is shown in Figure 22a, through positions TYR242, GLN243, VAL245, TYR250, LEU286, LYS289 and LYS290 of the LneB protein.

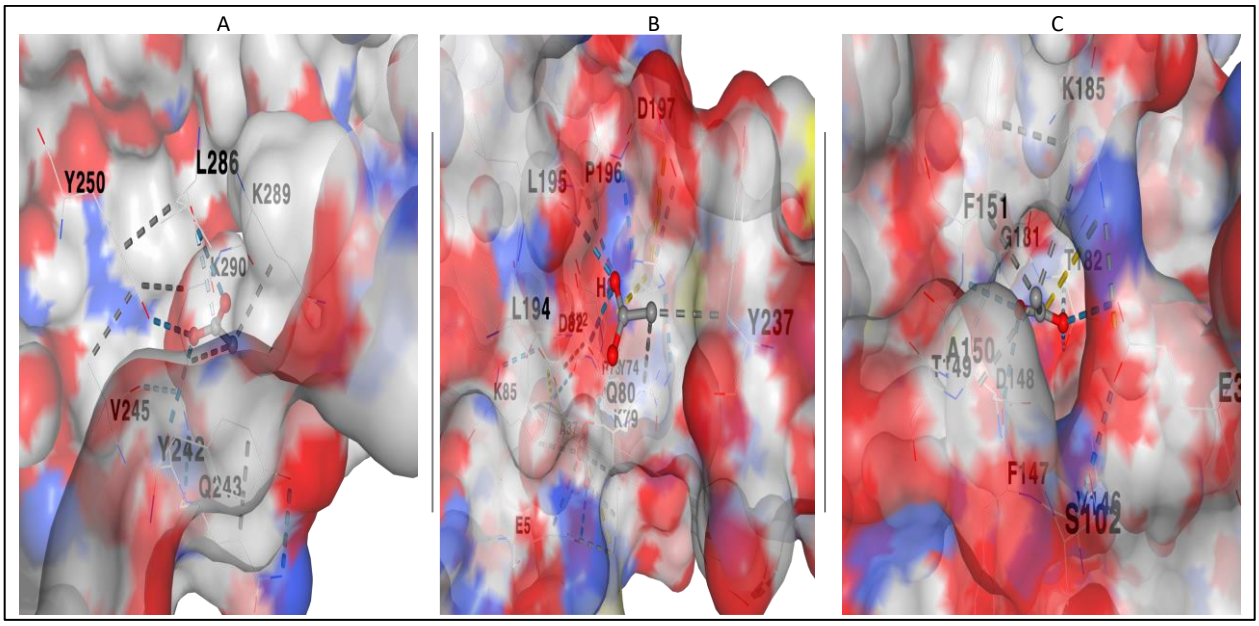


Figure 22. Predicted Molecular Interaction between proteins and Acetate Ion. A: Pocket with the best model of interaction between LneB and acetate ion. The best model has the lowest Vina score of -3.0 and volume (Å^3) of 1865 (CB-dock2). The best-modeled pocket is at the C terminal. The acetate is predicted to form a hydrogen bond with LneB protein at residues TYR242, GLN243, VAL245, TYR250, LEU286, LYS289 and LYS290. B. Pocket with the best model of interaction between GFP and acetate ion. The best model has the lowest Vina score of -1.8 and volume (Å^3) of 151 (CB-dock2). C. Pocket with the best model of interaction between acetate kinase and acetate ion. The best model has the lowest Vina score of -3.0 and volume (Å^3) of 915 (CB-dock2).

3.2. Expression of LneB Protein from cloned copy of the gene

LneB expression was induced from transformed *Escherichia coli* using IPTG (see appendix 1 for details). To obtain the expressed protein, the cells expressing LneB were collected and lysed, and

the pellet was analyzed for purification using Ni-ion column chromatography (see Appendix 02). Purified samples were treated with SDS sample buffer for SDS-PAGE analysis (see Appendix 03).

3.2.2. SDS-PAGE

LneB protein was overexpressed in BL21 *Escherichia coli* cells and SDS-PAGE was used to analyze purity, as shown in Figure 23.

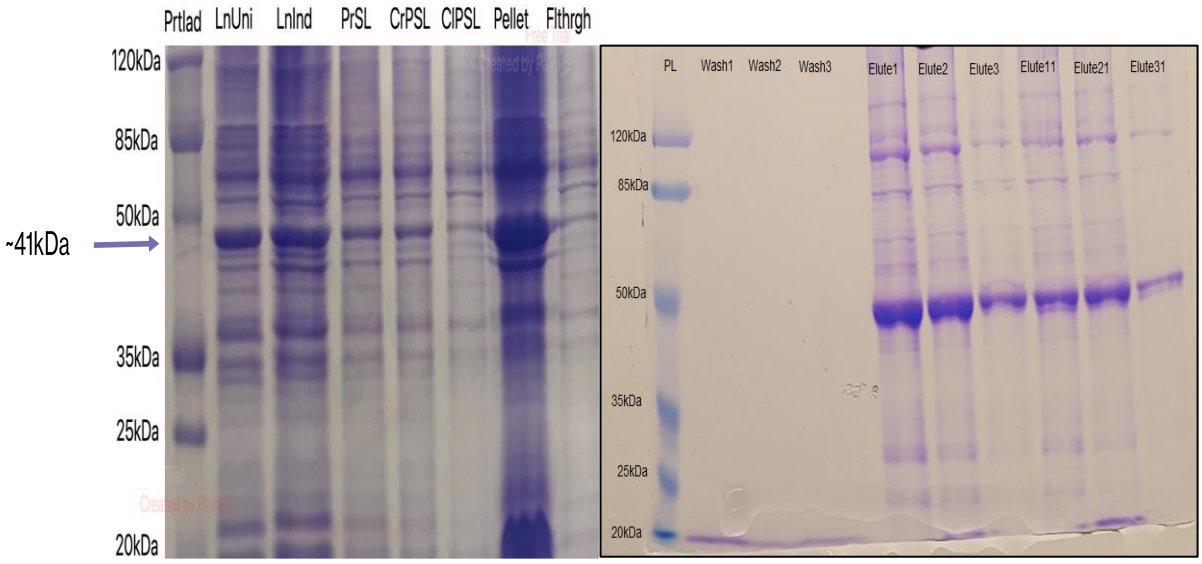


Figure 23: SDS-PAGE gel of induced LneB protein in BL21 *Escherichia coli.* On the SDS-PAGE gel, the lanes show protein estimation from cell lysate treated with SDS-loading dye. Lanes are arranged from protein ladder (PL), uninduced, pre-sonicated lysate (PRSL), crude-post sonication lysate (CrPSL), clear post-sonication lysate (ClPSL), pellet, flowthrough (Flthrgh), wash 1, 2 and 3 were filtrate of corresponding washes, elute 1 was the first elution, followed by elution 2 and elution 3. The label lysate represents samples that were treated with lysozyme alone.

The LneB protein (size 41kDa) appeared to be induced (with a bigger band) compared to the uninduced cells when the IPTG concentration was 200mM, and the incubation condition was at 18^0C overnight.

3.2.3. Western blot

Western blot was carried out using an anti-HIS primary antibody, and the LneB protein was verified in the eluted samples using the corresponding protein molecular ladder (Fig. 24).

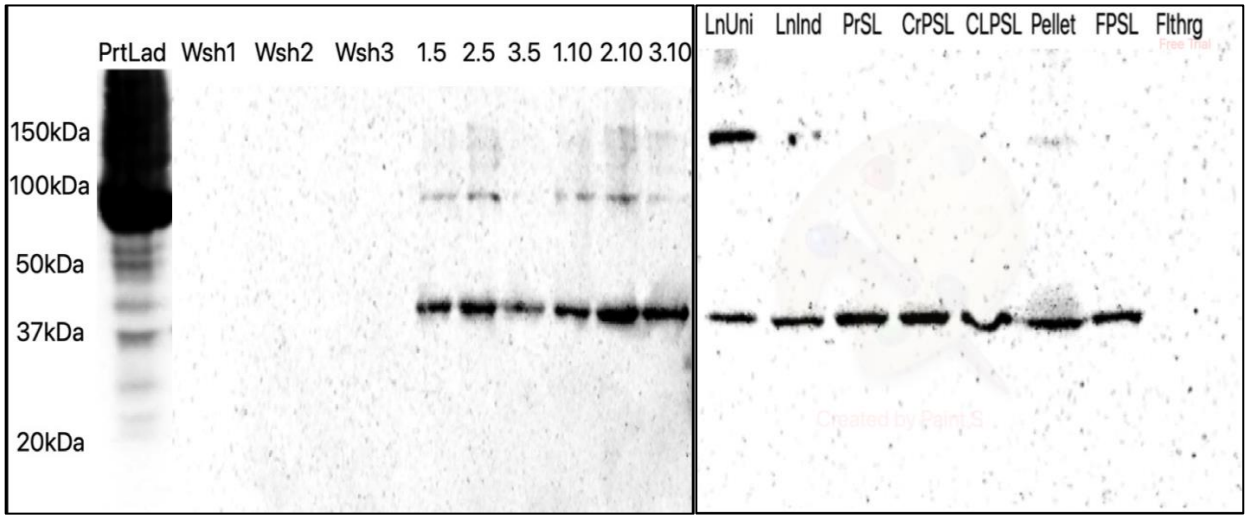


Figure 24: Western blot showing the presence of LneB protein in each purification fraction. The lanes represent the protein ladder (PL), wash 1-3, elute (each elution 1, 2, and 3 loaded with in duplicate at 5µg and 10µg respectively), 5µg of the other samples were loaded including cell lysates from uninduced, pre-sonicated lysate (PRSL), crude-post sonication lysate (CrPSL), clear post-sonication lysate (ClPSL), pellet, filtered post-sonication lysate (FPSL), and flowthrough (flThrg).

The western blot analysis suggests that His-tagged protein is present in the tested eluents, and the band size correspondent with the amount of the protein loaded into each well. However, some proteins of size about 100kDa were seen faintly on the membrane.

3.2.4. LneB Protein Quantification

The Bicinchoninic acid (BCA) assay was used to quantify the concentration of the eluted LneB protein through extrapolation from the linear regression formula obtained from the standard curve (Figure 25).

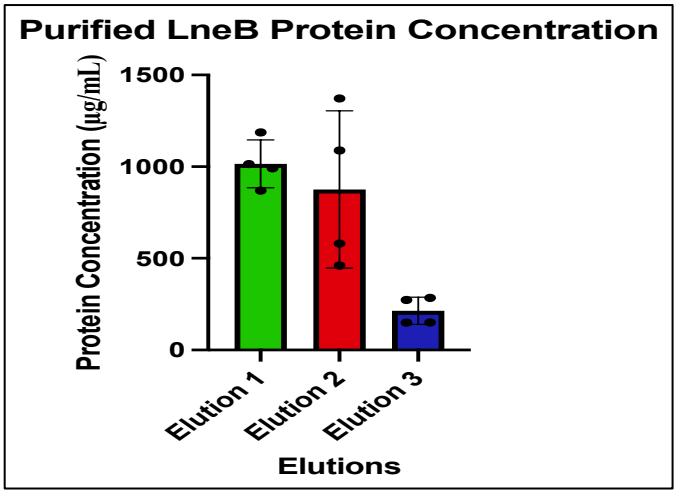


Figure 25: Estimating purified LneB protein concentration using BCA Assay with a standard deviation. The eluent concentration reduces from elution1 to 3 and ranges from 1000µg/mL to 150µL.

Purification was done twice and at each purification, samples were run in duplicate. One-way ANOVA testing showed that the purification is reproducible with no significant difference in protein concentration between replicates (p-value of 0.1160).

3.2.5. HAT Activity of Purified LneB Protein

The in vitro HAT assay of the purified LneB protein was measured and compared to the HAT activity of the elution buffer alone. The result shows no significant difference in the HAT activity

between the pure protein and the elution buffer at a confidence level of 95% with a p-value of 0.1137 and a t-value 5.537 (Fig. 26).

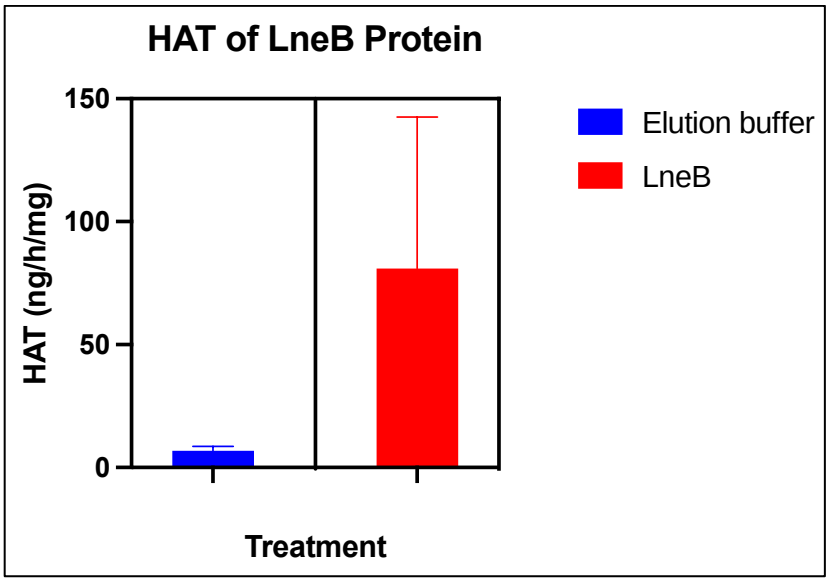


Figure 26: Histone acetyltransferase activity of purified histidine tagged LneB protein. The elution 3 of the purified LneB protein has higher histone acetyltransferase activity HAT than the elution buffer.

3.3. Infection Assay of *L. pneumophila* in U937 Cell.

To reaffirm the importance of effector proteins in the intracellular growth of *L. pneumophila* in the host cells, U937 cells were infected with the wild type and mutant form of *L. pneumophila* at a MOI of 0.05 in triplicate. The Lp03 mutant (ΔDot/Icm) of Legionella, has a dysfunctional Dot/Icm transport system, and has a growth defect compared to the wild-type strain Lp02. However, mutants missing up to three effectors are as viable and have average fecundity as the wild type, as shown in Figure 27. Deleting LneB, MavA, Ceg10 and RavQ does not affect the survival of *Legionella pneumophila* in mammalian cells (U937). However, a defective Dot/Icm

transport system significantly hinders the secretion of effector proteins and hinders intracellular growth (Fig. 27).

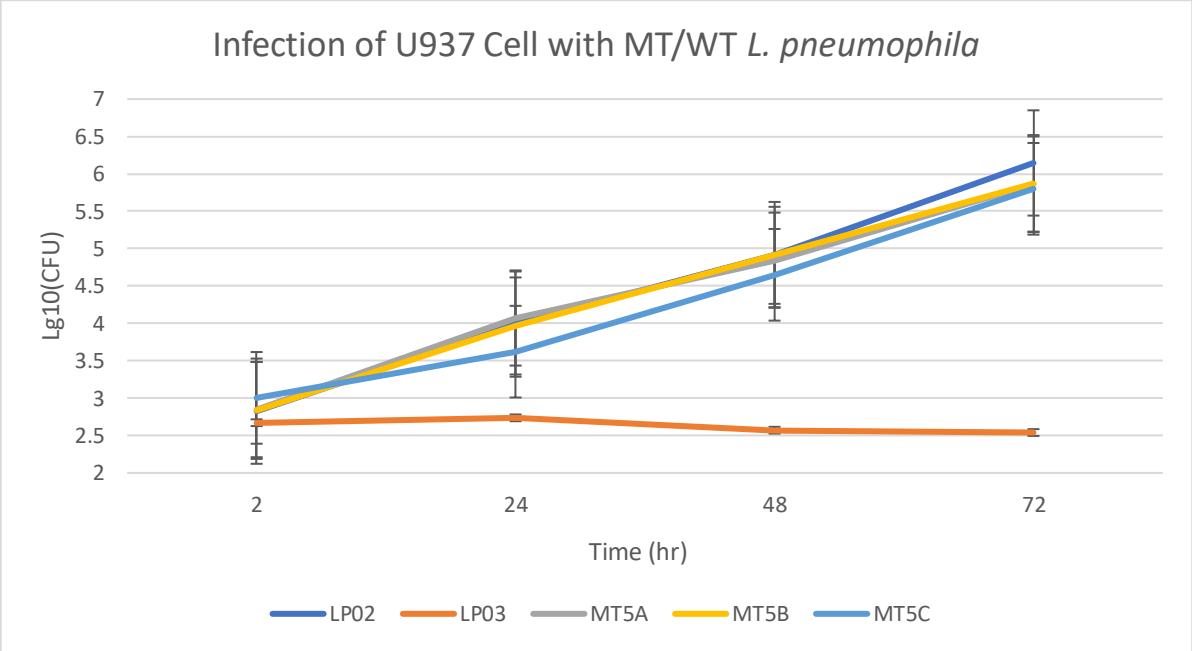


Figure 27: Growth curves depicting replication of wild type (Lp02), ΔDot/Icm (Lp03) with defective Dot/Icm system, and mutants (MT) missing up to three effectors (MT5A, MT5B, and MT5C) *L. pneumophila* in human myeloid leukemia cell line, U937. At the specified time points, non-infective bacteria were washed with warm PBS thrice and infected U937 cells were lysed, and bacterial multiplication was measured by plating serial dilutions of the lysates. The ΔDot/Icm (Lp03) has a defective growth pattern in the cell, and the colony-forming unit remains the same from 2hrs, 24hrs, 48hrs, and 72hrs.

3.3.1. *L. pneumophila* Uptake into U937 Cells

The histone acetyltransferase activity in the infected mammalian cells was also investigated. First, the uptake of the *L. pneumophila* in the host cells following infection was verified at a

multiplication of infection 10 and 100, followed by infection of the host cells and histone acetylation assay. Bacteria that could not be taken up by cells were washed off thrice with warm PBS before cell lysis. Bacterial count in the lysate showed increased colony-forming unit from 2 hours to 24 hours in U937 cells, as shown in Fig. 28.

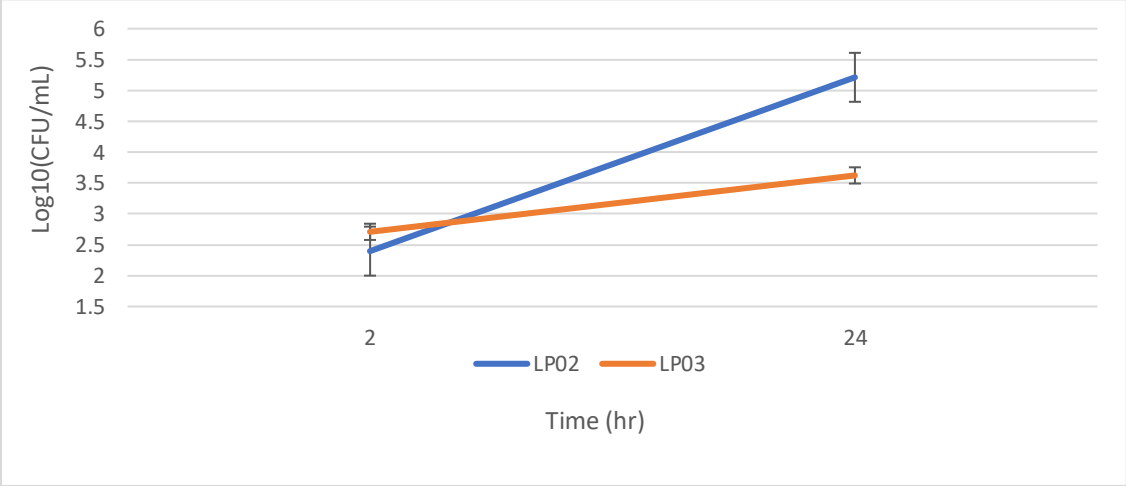


Figure 28: Uptake analysis of *L. pneumophila* wild type (Lp02), ΔDot/Icm (Lp03) in U937 cells at MOI 0.05. The uptake of the Δdot/Icm (Lp03) is not different from the uptake in the wild type at time two hours, but the growth of the wild type increased 44% more than the mutants, as expected.

3.3.2. Nuclear Extraction of U937 Cells Infected with *L. pneumophila.*

As stated in the procedure described above, the infection assay was carried out in triplicate at a MOI of 10 and 100, followed by nuclear extraction of the infected cells. The standard curve for the BCA assay for protein estimation and the concentration of the nuclear extracts are represented in Fig. 29.

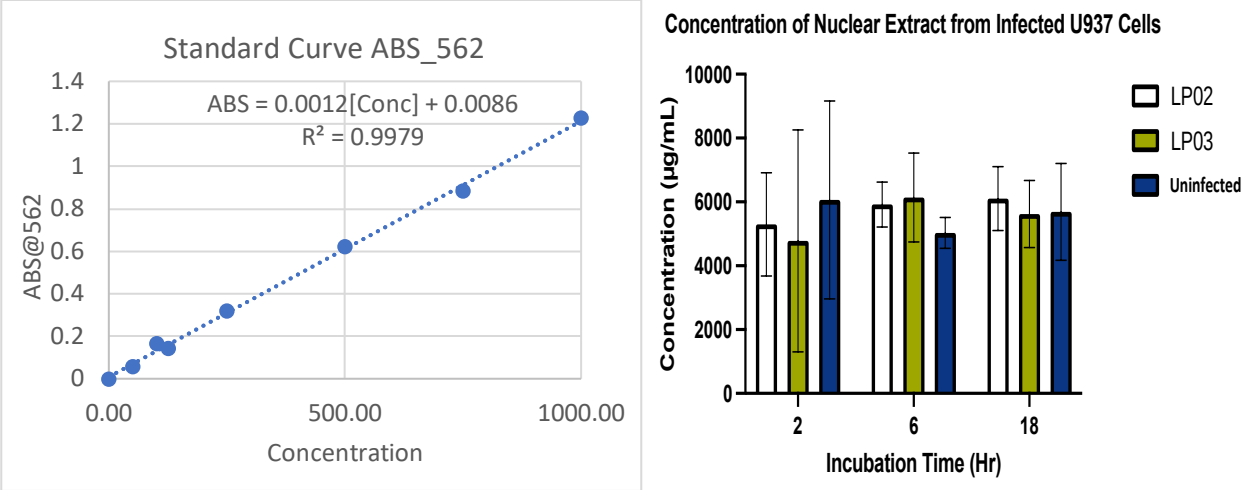
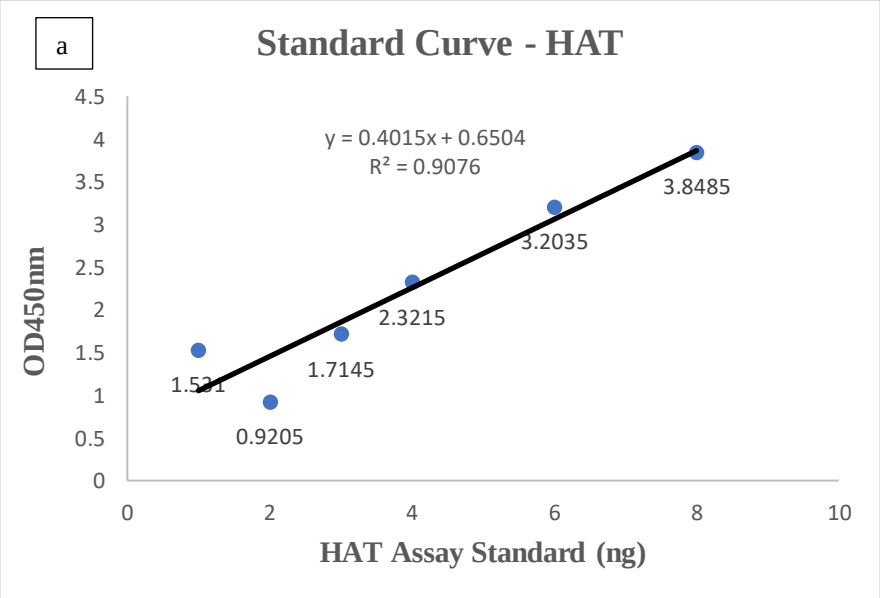


Figure 29: Extracted nuclear protein from infected U937 Cells. A: Standard curve for BCA assay. B: Average protein concentration of nuclear extracts from U937 cells infected with W.T. (Lp02), ΔDot/Icm (Lp03), and uninfected cells at 2, 6, and 18 hours in triplicate with data standard deviation.

One-way ANOVA testing was carried out on the concentration of the nuclear extracts from three treatments to determine statistical differences at 95% degree of freedom using Prism-GraphPad. The p-value between treatment (LP02, LP03, and PBS) is 0.8431. This shows that infection of either wild type, Dot/Icm defective mutant, or the uninfected cells does not significantly affect nuclear protein expression in U937 cells, thus, the cell harvest, lysing and recovery or protein are repeatable. Even though there is no statistically significant difference between the groups, the nuclear extraction from Lp02 infected cells increased from time point 2 to 18 hours (Fig 29). In contrast to this, the LP03 increased from 2 to 6hrs and decreased at 18hrs, while the uninfected cells showed a dip in the nuclear extract at 6hrs.

3.3.3. HAT in Nuclear Extract of Infected U937 Cells

A Histone acetylation assay was performed on the nuclear extract of uninfected, Lp02, and Lp03 infected U937 cells as described below using the commercially available ELISA-based HAT Epigentik Assay kit as described in Appendix 8. The estimation of histone acetylation is based on a linear regression formula (Fig. 30a). The HAT activity between uninfected, LP02 and LP03 at 10 MOI is not significant at any given time (Fig. 30b). However, for LP03, the HAT activity at 6hrs decreased significantly compared to the 2hrs with a p-value of 0.0360, which suggest a reduced need for host cells to initiate transcription because the virulence factors, the effector proteins, are not released by the bacteria. The HAT significantly increased afterward at 18hrs at a p-value of 0.0319 (Fig. 30b). The experiment was carried out in triplicate.



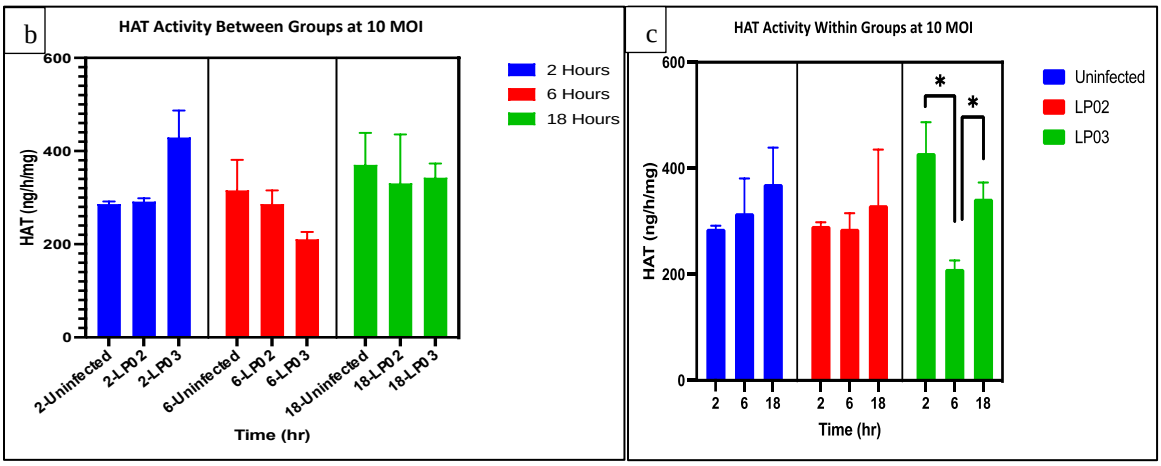


Figure 30: Histone acetylation of U937 cells infected with *L. pneumophila* at MOI 10. (A) Standard curve of histone acetyltransferase assay. U937 cells were infected with Legionella at a MOI =10 and analyzed for HAT activity at 2-, 6- and 18-hours post-infection with LP02 and LP03, uninfected cells served as the control. (B) HAT activity between groups at the same time. (C) HAT activity within each group at different time.

At MOI 100, activity increased from 2 to 18hrs in uninfected cells and highest at 6hrs in cells infected with the wild type (LP02). In the wild type (LP02), the activity reduces significantly at 18hrs compared to 2hrs and 6hrs. Cells infected with LP03 have a reduced activity at 6hrs and increased insignificantly at 18hrs (Fig. 31a).

There is no significant difference in the HAT activity between infected and uninfected cells; p-value of HAT activity at a 95% confidence level is 0.5384. There is a significant increase in HAT activity from 2hrs to 18 hrs. in LP02 infected cell; p-value 0.0314 and a significant drop-in HAT activity from 6hrs to 18hrs in LP02 infected cells; p-value 0.0238. Between the groups, the HAT activity has a significant reduction at 18 hours in the wild-type (LP02) compared to the uninfected (p-value = 0.0016) and the mutant (LP03) (p-value = 0.0217). Cells infected with mutant also has

significant reduction in HAT activity when compared with the uninfected cells at 18 hours (p-value = 0.0236).

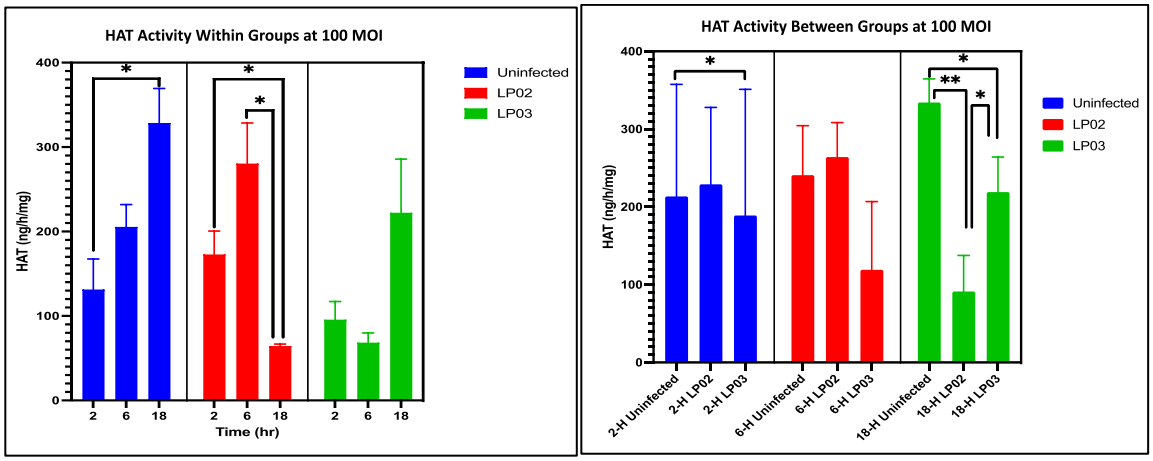


Figure 31: Histone acetyltransferase activity of nuclear extracts in uninfected cells and cells infected with LP02 (wild type), LP03 (Dot/Icm mutant) at a MOI of 100. (A) HAT activity within group at the different time. (C) HAT activity between groups at the same time.

3.4. HAT in Nuclear Extract of Transfected Cells

The nuclear extract of A549 cells transfected with a plasmid vector containing LneB-GFP or the GFP only control increased nuclear protein concentration from 24 to 48 hours, showing a time-course increase in total protein concentration; translation, in the cell. The concentration of the nuclear extracts is expectedly generally lower than the cytoplasmic extracts. The protein concentration is lower in cells transfected with LneB protein compared to GFP control (Fig. 32).

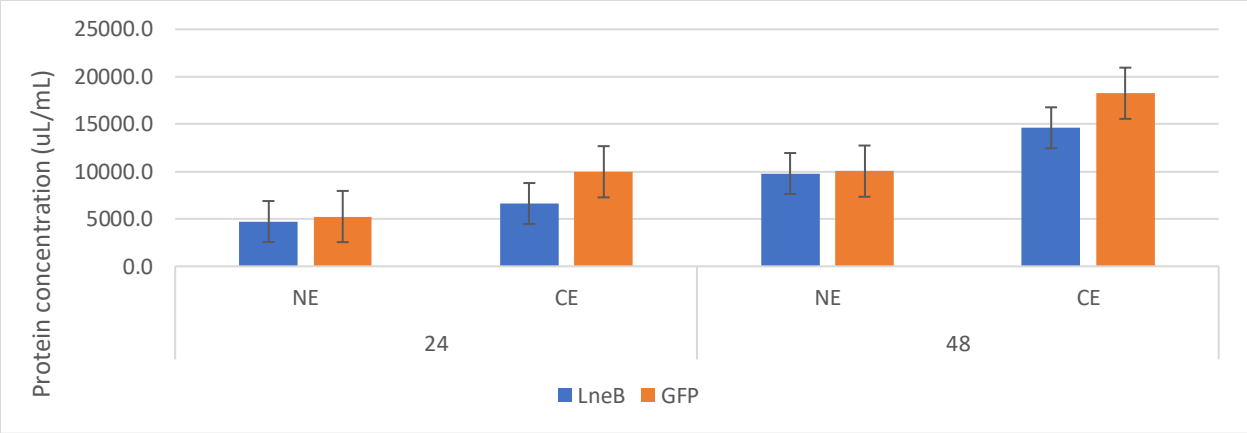


Figure 32: Protein concentration of nuclear extracts from transfected A549 cells. The concentration of nuclear extracts (NE) and cytoplasmic extracts (CE) were estimated 24 and 48 hours after transfection. The NE is generally lower than the CE while the GFP is higher than the LneB in both NE and CE.

Histone acetylation in transfected A549 cells was significantly reduced in LneB-GFP transfected cells compared to GFP control with a p-value of 0.0025 and a t-value of 20.08 (Fig. 33).

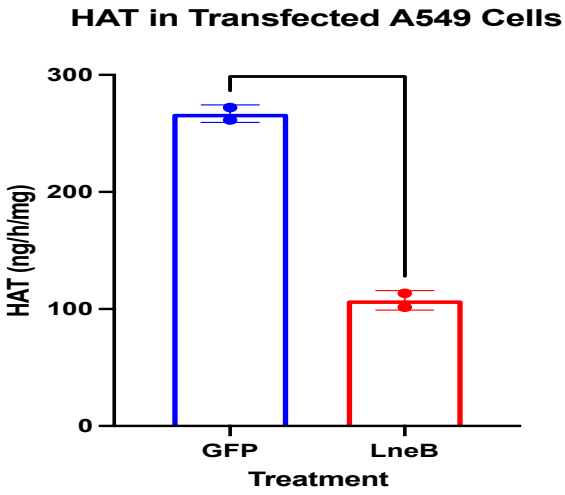


Figure 33: Histone acetylation in transfected A549 cells, comparing GFP control with LneB-GFP vector. The HAT activity in LneB transfected cells is significantly lower than in GFP-controlled cells.

DISCUSSION

It was not until 1976, during the 58th American Legion convention in Philadelphia, that Legionnaires disease and its causative gram-negative bacillus, *L. pneumophila,* were discovered. The endemic caused by the bacteria led to the death of 15.6% of the victims. The water-borne transmission of the causative agent created awareness of the environmental and medical relevance of the bacteria. As such, there is categorization into hospital-acquired and environmental—acquired Legionnaire diseases [168]. *L. pneumophila* naturally resides in amoeba in the environment but accidentally infects humans via contaminated aerosolized water particles containing infected amoeba [169], [170].

L. pneumophila, like other intracellular bacteria, has unique virulence factors in addition to its flagella. The flagella are usually lost when the bacteria are in the reproductive phase or retained when they are in the transmissive phase. The additional virulence factors in *L. pneumophila* are the effector proteins which are more than 300 in number [171]. As a single protein, para-effector or meta-effector, the effector proteins affect various cell processes in the host [172]. One unique pathogenic behavior of *L. pneumophila* within the host is the ability to reside and make a "comfortable" environment for itself within the host's innate immune cells (macrophage). This evading strategy is achieved by creating a Legionella-containing vacuole (LCV), which is important for escaping lysosome degradation in macrophages [173]. Furthermore, some of the effector proteins have been implicated in the development and maturation of the LCV within host cells [174]. Still, the significance of each of the effector proteins in the pathogenesis of the bacteria is unclear. This knowledge gap informed this study to investigate the function of one of the uncharacterized proteins, LneB protein, in the host cell.

Previous work in our lab showed that LneB protein localized in the nucleus. In this study, further investigation on LneB protein was carried out using transcriptomics analysis of differentially expressed genes and bioinformatics exploration of the structure and function of LneB protein to build a testable hypothesis towards understanding the protein's role. Structurally, the LneB protein is predicted to have seven alpha helices, and six anti-parallel beta sheets and binds to nucleotide and acetate rather than with protein. Post-translational modification of histone through acetylation is the dominant predicted function of LneB protein from the bioinformatics and transcriptomics analysis as well as previous work in our lab on the nuclear localization of the protein.

There is generally little or no homology among HAT proteins [175], but the predicted functional domains on LneB protein include regions that showed 19% dispersed similarity in sequence alignment with the MYST domain of human HAT enzyme. The dispersed nature of the aligned sequence weakens the argument of similarity between the two proteins. Further analysis through pairwise sequence alignment was done between the Legionella LneB protein and histone acetyltransferase enzyme from amoeba. This comparison showed about 24% similarity, and a coverage of 93%. Thus, the evolutionary relationship between LneB of *L. pneumophila* and the HAT enzyme in the bacteria's natural host, amoeba, reflects the evolutionary closeness between the LneB protein and HATs of amoeba than with human HAT enzyme. Furthermore, the prediction of acetate binding and similarity with MYST domain of HAT supports the concept of the bi-substrate nature of most HAT proteins, where the HATs bind to Acetyl-CoA and lysine-rich substrates [176] This is supported by the predicted probability of an interaction between LneB and Acetyl-CoA in the simulated molecular interaction software.

The predicted interaction between nucleic acid and LneB protein rather than interaction with either protein or RNA is supported by the literature showing that some histone acetyltransferase enzymes

such as MOF protein in *Drosophila melanogaster* only bind to nucleic acid rather than histone proteins [177]. The MOF do not bind to free histones but to chromatin and carry its histone acetyltransferase activity through chromobarrel domain [177]. In contrast, other HAT enzymes such as human TIP60 and homologous yeast NuA4 complexes, require other associated cofactors to form native complexes before accessing chromatin [135].

Through transcriptomics analysis, the molecular activities differentially induced by the LneB-GFP protein in 293T cells support the interaction of the LneB protein with acetyl-CoA. There are relationships between the upregulated genes; dehydrogenase reductase 2 (DHRS2) and early growth response (EGR) genes and acetyl-coA and transcription process. In similar lung infection caused by *Pseudomonas aeruginosa*, the Egr-1 was shown to be rapidly expressed in the presence of the bacteria which lead to host's hyperinflammation and increased mortality [178]. The DHRS2 is highly expressed at transcription and translation levels when histone deacetylase is inhibited in ovarian cancer cells [179], suggesting that histone acetylation may enhance the expression of the DHRS2. The DHRS2 activity is known to depend on NADPH and reduces dicarbonyl compounds used for modifying cellular components [180]. DHRS2 also regulates hormone levels by hydroxysteroid dehydrogenase activity through substrates such as ursodeoxycholic acid (isoUDCA) and ursodeoxycholic acid (UDCA). The DHRS2 has also been linked to increased cell transcription through p53/TP53 stabilization because it can attenuate MDM2-mediated p53/TP53 degradation [181]. [179]

Implicated early growth response protein 1 (Egr-1) upregulated by LneB-GFP is a zinc finger protein 268 (ZNF 268). It regulates the transcription of genes responsible for differentiation and mitogenesis which is in line with the outcome histone acetylation. The Egr-1 protein increases transcription by recruiting the TET1 protein to aid the demethylation of DNA [182]. The

recruitment of RNA polymerase II S5ph to the promoter of the Egr-1 protein is known to be influenced by phosphorylation through the p38, and MEK1/2 pathways [183]. This increases transcription through histone H3 acetylation at the +1 nucleosome [184]. However, the activation of the promoter and corresponding histone acetylation was said to be likely due to the CREB-binding protein [185]. Inferences from possible increased transcription by DHRS2 coupled with Egr-1's role in activating histone acetylation suggested the role of LneB in the processes.

Experimental analysis showed that LneB protein does not have statistically significant effect on the histone acetylation in vitro as a purified protein when compared to elution buffer. However, the LneB protein significantly reduced histone acetylation in vivo in transfected A549 cells. This result is surprising because it is in contrast to the predicted activity of the protein. This may be due to some reasons that are needed to be investigated further, such as if LneB protein does not acetylate free histone, or its activity requires other supporting effectors.

The epigenetic modification of histone is a phenomenon whereby histone proteins are modified by addition of certain groups such as acetyl, methyl, or phosphorus groups which has overall effect on the processes that are related to chromatin such as compaction of chromatin, nucleosome related processes, and transcription [186]. This study showed that histone acetylation by Dot/Icm substrates in wild type *L.* pneumophila in infected U937 cells at 10 MOI is not significantly different from the histone acetylation in uninfected cells. These results suggested that at 10 MOI, the effector proteins could not affect significant HAT activity, or the host response to HAT by the effectors reversed the process through histone deacetylase (HDAC) enzyme.

More significant changes in HAT activity were seen when the MOI was increased to 100. At 100 MOI, the HAT activity increased 2.5-fold from 2 hours to 18 hours in uninfected cells but in cells

infected with the wild type (LP02), the HAT activity dropped 4-fold from 6 hours to 18 hours, showing that the Dot/Icm substrates has significant effect on HAT activity in the host cells. This result suggested that at 100 MOI, there are more concentration of effector proteins that will be released; thus, the proteins' HAT activity was more evident. extension, comparing HAT activity at 18 hours between wild type-infected cells and uninfected cells, there is a significant reduction in the activity, the same is true when comparing the mutant and the infected cells. However, the HAT activity reduces in wild type infected cells at 18 hours when compared with mutant infected cells. The molecular mechanism of the influence of Dot/Icm substrates on host histone requires further investigation to understand host response to the activity (through HDAC) and at which minimum MOI does such activity become significant.

Further questions need to be addressed in future study as whether LneB protein binds to acetyl-coA as predicted in this study or binds to histone protein. Further investigation into the molecular mechanism the Dot/Icm substrates reduces HAT activity in U937 and A549 cells respectively need to be carried out. I, therefore, suggest further investigation on whether LneB requires other supporting effector, or if the LneB protein acetylates different types of histone protein other than H3 used in the assay kit. For instance, it was previously shown that histone 3Lys12 was deacetylated by a combination of LphD and RomA effector proteins in *L. pneumophila.* [187].

Project Two: MavA Protein

4.1.Transcriptomics Analysis

4.1.1. Differentially Expressed Genes

The gene profiling of the RNA transcripts from cells transfected with MavA showed that it altered the expression of seventeen genes in the cell compared to cells expressing the GFP protein alone from which only one was downregulated (Fig. 34). The MavA protein induced expression of a TBP-2-like inducible membrane protein called alpha-arrestin 3 (ARRDC3) with a significance -Log10(p-value) of 8.3 and Log2FC of 0.4472 compared to GFP only control. Also, an essential player in the metabolic energy pathway, dehydrogenase/reductase 2 (DHRS2) was significantly upregulated by MavA protein in the HEK 293T cell at a -Log10(p-value) of 7.1 and Log2FC of 0.4545 (45.45%) differential upregulation.

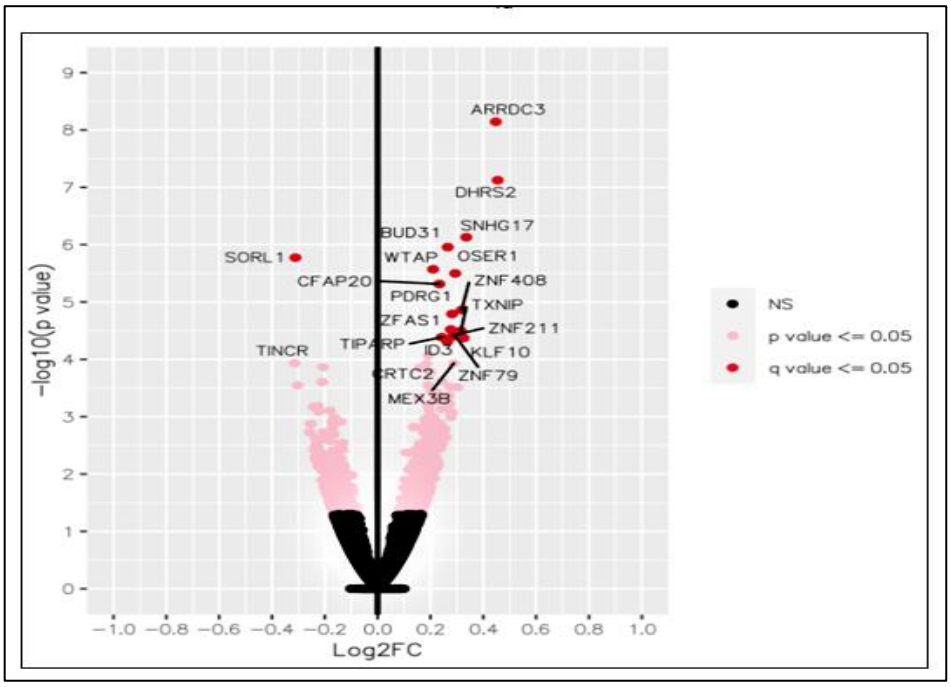


Figure 34: Volcano plot of the transcriptome analysis. This figure represents differentially expressed genes HEK 293T cells expressing MavA GFP-tagged protein compared to those expressing GFP protein alone. The x-axis represents fold change on a logarithmic scale of 2. The positive or negative fold changes mean the gene are upregulated or downregulated. The y-axis represents the -log10 of the p-values. Genes with low q-values (more relevant) are shown in red, and genes with low significance are shown in pink.

Other upregulated genes are non-coding genes; small nucleolar RNA host gene 17 SNHG17, functional spliceosome-associated protein, BUD31, OSER1, and TXNIP. These genes are associated with post-transcriptional modification [188], pre-mRNA splicing [189], oxidative stress mediator [190] and transcript repressor gene families [191] respectively. In this regard, a recent study has shown that intracellular growth of *L. pneumophila* activates more stress-related effectors when grown U937 cells than in natural protozoan host, *Acanthamoeba polyphagia* [192].

The only significantly downregulated gene is sortilin-related receptor 1 (SORL1). Sortilin protein binds directly to the receptor. The sortilin receptor is important in directing several proteins in the cell to their specific targets. The sortilin protein is also crucial for proper sorting of proteins to the lysosome for degradation [193]. The downregulation of this receptor by MavA protein is statistically significant with a p-value and q-value less than 0.05. The molecular process affected by the differentially expressed genes are shown as KEGG annotation in Table 6.

Table 6: KEGG Pathway Analysis of Differentially Expressed Genes in MavA-expressed HEK 293T Cells.

GENE ID	**DESCRIPTION**	**KEGG Annotation**	**Log2FC (MavA/GFP)**	**p_value**	**q_value**
DHRS2	Dehydrogenase/reductase 2	Cellular response to oxidative stress	0.454500528	7.50E-08	0.000648
ARRDC3	Arrestin domain containing 3	Protein binding, Lysosome, Early endosome	0.447182244	7.17E-09	0.000124
SNHG17	Small nucleolar RNA host gene		0.335591675	7.40E-07	0.004259
KLF10	Kruppel like factor 10	Nucleic acid binding, cell-cell signaling	0.325273397	4.24E-05	0.045809
ZNF408	Zinc finger protein 408	Nucleic acid binding	0.317692001	1.38E-05	0.026399
TXNIP	Thioredoxin interacting protein	transcription, DNA-templated	0.3149697	3.17E-05	0.045683

ZNF79	Zinc finger protein 79	Nucleic acid binding	0.293910974	3.92E-05	0.045809
OSER1	Oxidative stress-responsive serine-rich 1		0.293486498	3.16E-06	0.007805
ZNF211	Zinc finger protein 211	Nucleic acid binding	0.286702719	3.80E-05	0.045809
PDRG1	p53 and DNA damage regulated 1	Protein folding, Protein binding	0.281697622	1.61E-05	0.027862
ZFAS1	ZNFX1 antisense RNA 1		0.276926574	3.02E-05	0.045683
BUD31	BUD31 homolog	mRNA processing	0.26549211	1.10E-06	0.004754
ID3	Inhibitor of DNA binding 3, HLH protein		0.263811623	4.85E-05	0.049307
TIPARP	TCDD inducible poly (ADP-ribose) polymerase	Transferase activity	0.242642575	4.10E-05	0.045809
CFAP20	Cilia and flagella associated protein 20	Cytoskeleton, Microtubule, Cilium assembly,	0.233448021	4.85E-06	0.010473

		Positive regulation of cell motility			
WTAP	WT1 associated protein	mRNA processing	0.209835601	2.69E-06	0.007728
SORTL	Sortilin receptor protein 1	Protein targeting to the lysosome, mediated endocytosis, Golgi vesicle transport.	-0.3113	1.68e-06	0.005806

4.2.Bioinformatics Predictions of the Properties of MavA Protein

4.2.1. Sequence Analysis and Evolutionary Analysis

Alignment of the amino acid sequence of MavA protein with similar proteins from NCBI protein-protein BLAST was carried out and the FASATA sequence of the aligned proteins were retrieved for further analysis with MEGA-11, with which a neighbor joining phylogenetic tree was drawn. The analysis showed that the closest protein homolog is a 378aa uncharacterized protein of *L. waltersii*, WP_058481790.1 (Fig 35a). This protein is 46.37% identical to MavA, with alignment coverage of 99% and an expectation value of 1e-^{107}. In comparison, the most distant ancestral

protein is WP_058493972.1 (402 amino acid residues), a hypothetical protein of *L. worsleiensis* with a coverage of 99%, e-value of $1e^{-107}$, and degree of identity of 43.28% (Fig. 35a).

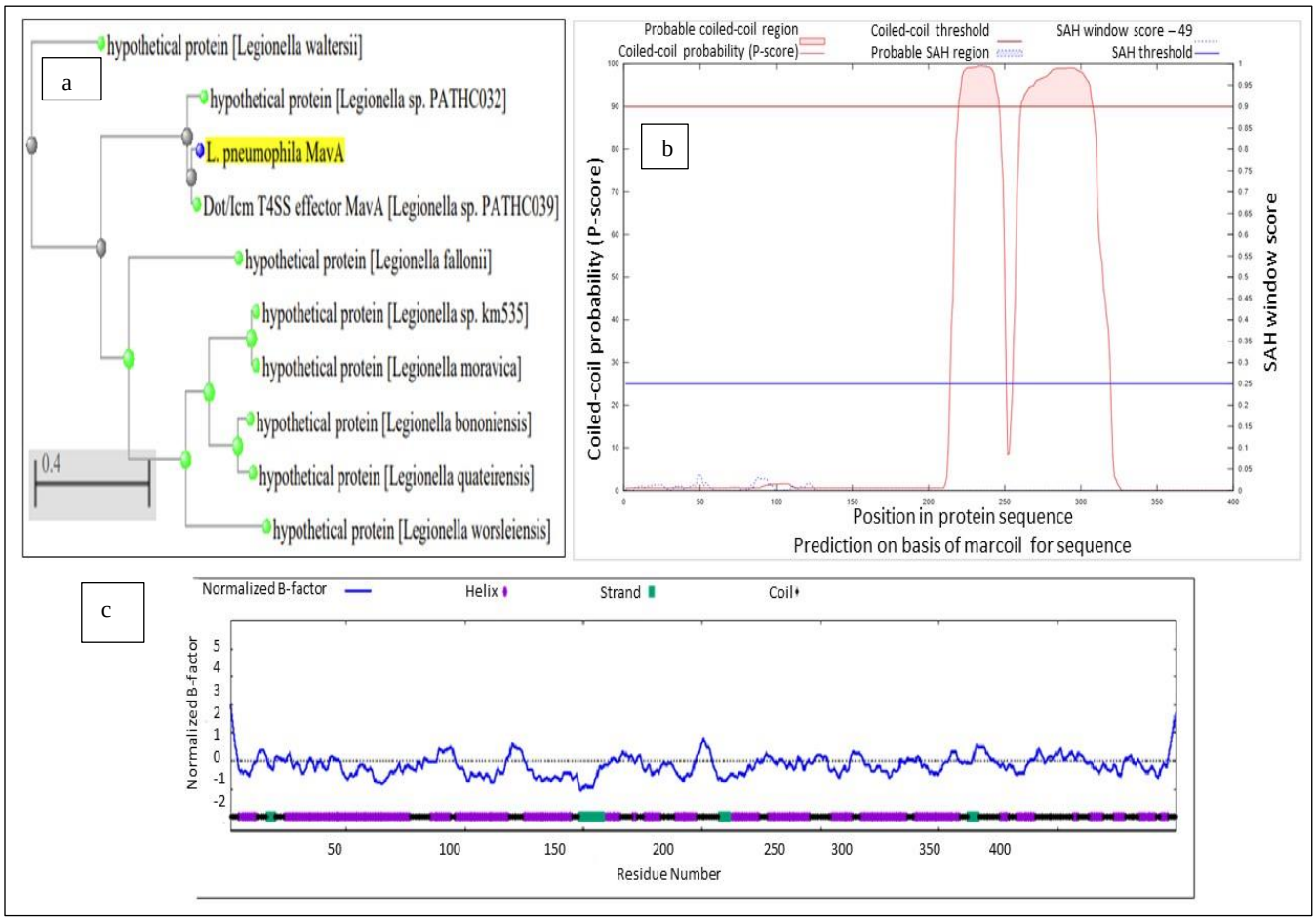


Figure 35: Evolutionary and Sequence Analysis of MavA. A: Neighbor-joining phylogenetic analysis of MavA protein and sequence-aligned proteins. The amino acid sequences of the MavA protein were analyzed with the protein-protein BLASTP program on the NCBI on the full length of 400 amino acids using the standard database of non-redundant protein sequences. B: Normalized B-factor profile of MavA generated from the I-TASSER server showing the types of secondary structures predicted for each amino acid residue. Residues with BFP values higher than 0 (about 29.5% of the residues) are predicted to be less stable when analyzed experimentally C:

Coil-coil domain prediction in MavA secondary structure from ASP 204 to LYS 319 using WaggaWagga server.

[194][195]. The predicted structural conformation the of the amino acid residue in MavAare shown against their normalized B-factor. The B-factor indicates the atomic thermal mobility of the residues (Fig. 35c). Due to the numerous helix structures on MavA protein, specific probe for coil-coil domains was carried out using the Waggawagga tool [197], which showed two random coil-coil interactions at positions 220 - 246 and 261 – 307 (Fig. 35b), which are regions of superhelical structures.

4.2.2. **Structural Predictions of the MavA Protein**

One region of the MavA prediction of coil-coil domains was supported by I-TASSER with residues LESLKDKEVLLKKQLKHLNKK at positions 222-242 by the I-TASSER server. There are twenty-one predicted helices distributed across the protein's monomeric chain, as well as four sheet structures. In contrast, the most continuous coil configuration and the most continuous sheets lie between residues 341 to 356 and 149 to 156, respectively, as shown in the predicted I-TASSER secondary structure Fig. 36.

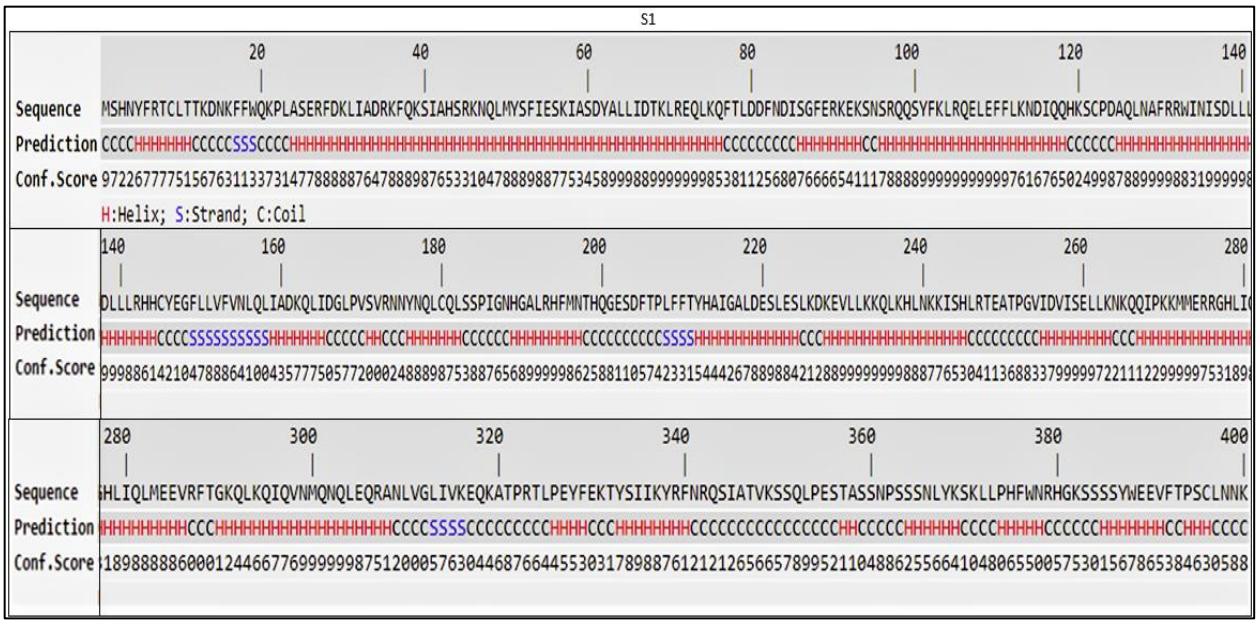


Figure 36: MavA Predicted Secondary Structure with I-TASSER. The Cs, Hs and Sc shows Coils, Helixes and Strands respectively.

The I-TASSER server was set at a default setting to obtain the predicted secondary structure, including confidence scores. C represents residues that potentially formed coil structure, H for helix, and S for sheet conformation. Helix > Coil > Sheet. The confidence score ranges from 0 (minimum) to 9 (maximum).

The Colab-Alfafold was also used to predict five ranked tertiary structural models of MavA protein. The comparison of the five predicted structures of MavA protein by AlfaFold is represented as a heatmap (Fig. 37b). The best model (Fig. 38) is monomeric and is primarily composed of alpha helix structure with a minor beta-sheet and loops, including their aligned error profile (Fig. 37a). In addition, the alignment of the best model with other proteins in the COFACTOR and the I-TASSER servers gave the predictions with the highest TM-score as Ras-activator and Ras-specific exchange factor, respectively (Fig. 37b-c).

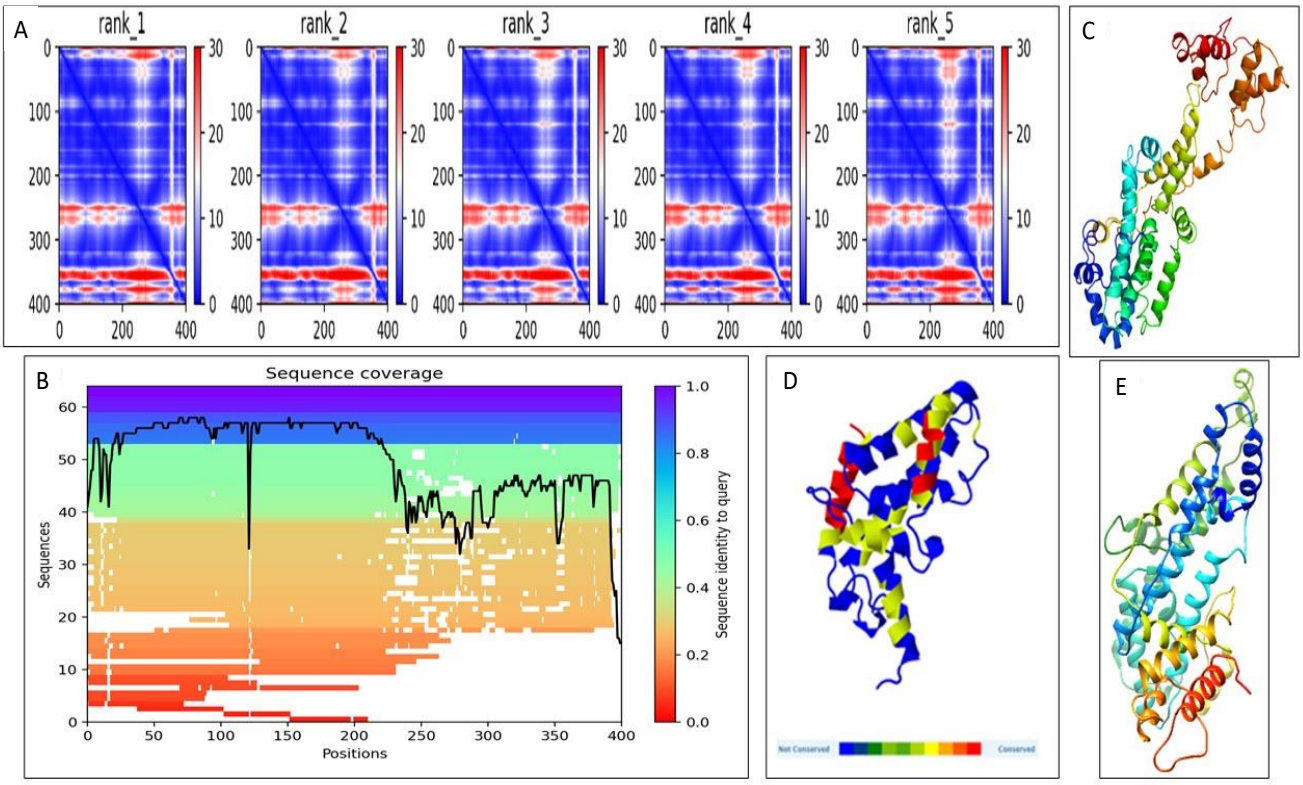


Figure 37: Structural Analysis of MavA. (A): The 5-model estimated PAE plot derived from an AlphaFold2 run. PAE score shown at the right-hand side of each plot determines the confidence of the predicted relative positions of each atom. Regions with lower PAE are well-defined. (B): Heatmap summarizing the multiple sequence alignment. Sequences are ranked from the top (highest identity) to the bottom (most minor identity) according to the identity score indicated by the color scale (lowest identity). Using the total number of aligned sequences, the relative coverage of the sequence is indicated by the black line. This figure displays the expected alignment error between each model residue. The blue and red colors are the most and least identical to the query. The confidence scores of sequence alignment of MavA and the template used range from 0.7 to 0.8. (C-E): Predicted tertiary structures of MavA protein using Phyre2, SWISS-MODEL, and I-

TASSER, respectively. The blue color shows non-conserved regions while the conserved regions are shown with red color.

The full length of the MavA amino acid sequence was uploaded to the individual servers, and the structural prediction was carried out using default parameters.

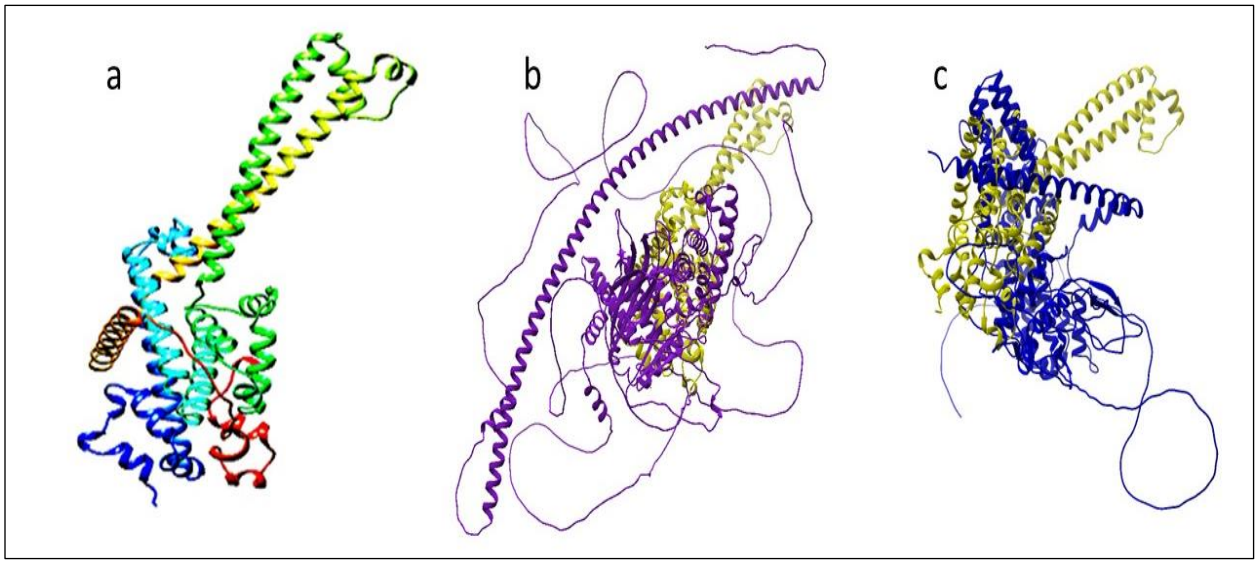


Figure 38: Structural Analysis by AlfaFold2. A: Tertiary Structure of MavA Protein. The full-length amino acid sequence was uploaded to the Google Colab AlfaFold server at a default setting. The analysis generated five models in order of prediction confidence. The best structure with the highest pLDDT score of 96.6% is shown. The model confidence band colors the predicted region as a spectrum based on pLDDT values of the alpha carbon-Cα atoms from red: higher pLDDT values to blue: lower pLDDT values. B: Using Chimera, Ras activating protein (color purple) aligns with the first-ranked MavA (color yellow) model. The best structural model of MavA was uploaded into the COFACTOR server to be threaded by local and global structure matches through the BioLiP protein function database to identify functional sites and homologies. The Homolog with the highest TM-score (0.523) was the Ras-activating protein. C: Homolog model (RasGRP)

with the highest TM score from the I-TASSER server (blue) was aligned with the MavA protein from Chimera (yellow). Rethreading 3D models created the functional model through the BioLiP protein database. The highest-ranked model, as shown, is a Ras-specific exchange factor RasGRP1 protein.

4.2.3. Function and Domain Prediction of MavA Protein

The potential active sites and domains of MavA protein were predicted using I-TASSER, COFACTOR, SMVTriP, Mmseqs2, and NCBI domain search. The I-TASSER predicted active sites for guanine exchange factor along the protein's N-terminal at positions 53 – 151, from which the highest C-score is at positions 136 and 140. The I-TASSER further identified a potential binding site for GTP on the C terminal of the protein at positions 281, 284, 356, 359, and 360. The SVMTriP predicted eight binding sites with other unspecified proteins, mainly on the C-terminal of the MavA protein, from which the top three sites are at positions 208 – 227, 274 – 293, and 376 – 395. However, COFACTOR predicted binding sites mainly in the N-terminus at positions 3 – 58, with residues SER 25, GLU 26, GLU 55, and ILE 58 assigned the highest Bscore of 0.67 among the PDB hits.

The predicted tertiary structure of MavA obtained from the Alfafold2 was superimposed with known proteins in the database using the Local Distance Difference Test. The LDDT indicated seventeen potential domain sites in MavA protein, and positions with IDDT scores greater than 90 are 7-12, 30-47, 60-70, 100-110, 130-220, 255-265, 280-305, 320-330, 360-370 and 375-380. Following the acquisition of the predicted tertiary structure of MavA, the predicted ligands of the homologous enzymes including GTP, non-polymeric [(Z)-octadec-9-enyl] (2R)-2,3-

bis(oxidanyl)propanoate (MPG) and N-[(4-aminophenyl)sulfonyl] cyclopropane carboxamide (RVI) with their respective RMSD value Table 7).

Table 7: Predicted Enzyme Homologs, with ligands and binding site of MavA.

Bioinformatic Tool	**Predicted Enzymatic Activity**	**Predicted Ligand**	**PDB Hit**	**Cscore**	**Predicted binding site**	**TM-Score**	**Coverage**
I-TASSER	Guanyl-nucleotide exchange factor activity	RV1	4uryS	0.04	57,61,142,143	*	*
	Phosphoribosyl-transferase GTP complex	GTP	1jlrD	0.04	102,105	*	*
	phosphotransferase	MPG	4o6*	0.04	126,130	*	*
COFACTOR	Guanyl-nucleotide exchange factor activity	*	*	*	*	*	0.8
	Acetyl-CoA carboxylase	*	*	0.588	26,27,29,30,31	0.352	0.01

In table 7, the greater the C-score (from -5 to 2) of a model, the more the confidence of its prediction. The second criteria to shown is the TM-score, which when greater than 5, the model is considered reliable while TM-score less than 0.17 is considered a random model.

In addition to amino acid sequence alignment, protein analogs to MavA protein predicted structure were predicted using ProBiS, Phrey2, I-TASSER, and COFACTOR. These programs predicted a guanine nucleotide exchange factor with identity ranging from 5 – 15% and confidence ranging from 42 – 97% (Table 8). The percentage of similarity is however quite low, but the unified prediction of the programs suggests more confidence on the result of the analysis.

Table 8: Predicted Protein Analogs of MavA protein according to their percentage identity, coverage, and statistical confidence using four different bioinformatics tools.

Tool	**PDB ID**	**Predicted Protein Analog**	**Z-score**	**Identity**	**Coverage**	**Confidence**	**TM-Score**	**RMSD**
ProBis	4f7z	Guanine Exchange Factor	1.69	*	*	*	*	*
Phyre2	3cf6E	Guanine Exchange Factor	*	15	*	96.8	*	*

I-TASSER	4l9mA	Guanine Exchange Factor	*	11.2	0.94	*	0.91	1.53
	1tr2B	Vinculin	*	5	0.62	*	0.42	6.09
COFACTOR	2ijeS	Guanine Exchange Factor	*	14.7	0.56	*	0.5	2.98

In the ProBis database, the Z-Score indicates the standard deviations from the mean in each alignment result [198], such that a Z-Score of 1.69 shows an alignment result that is in top 1.69% of all alignments. A good model on the I-TASSER and cofactor has a TM-score >0.5, while a random model has a TM-score < 0.17; “Identity” is the percentage of similarity between the query and template proteins. * The asterisk means the parameter was not reported by the tool.

In the consensus functional analysis, the Ras-GEF is the primary functional annotation for MavA made by 80% of the predictive tools (Fig. 39a). The Ras-GEF was predicted at the highest rate, followed by transport, cytoskeletal protein, and kinase. Two of the tools predicted other properties, including isomerase, oxidoreductase, and signalosome. The similarity in prediction by these tools is shown as a Venn diagram which indicates that Ras-GEF activity of the MavA protein is the only shared prediction among the four explorational tools (Fig. 39b). The dominant annotation for the MavA protein in all the tools is the Ras-GEF function. Two common elements in "I-TASSER" and "Phyre2" are Kinase and Cytoskeletal protein, and one common factor in "I-TASSER" and

"COFACTOR" is Carboxylase. Oxidoreductase is one common element in "COFACTOR" and "Phyre2," and transport protein is the only common element in "Phyre2" and "ProBiS."

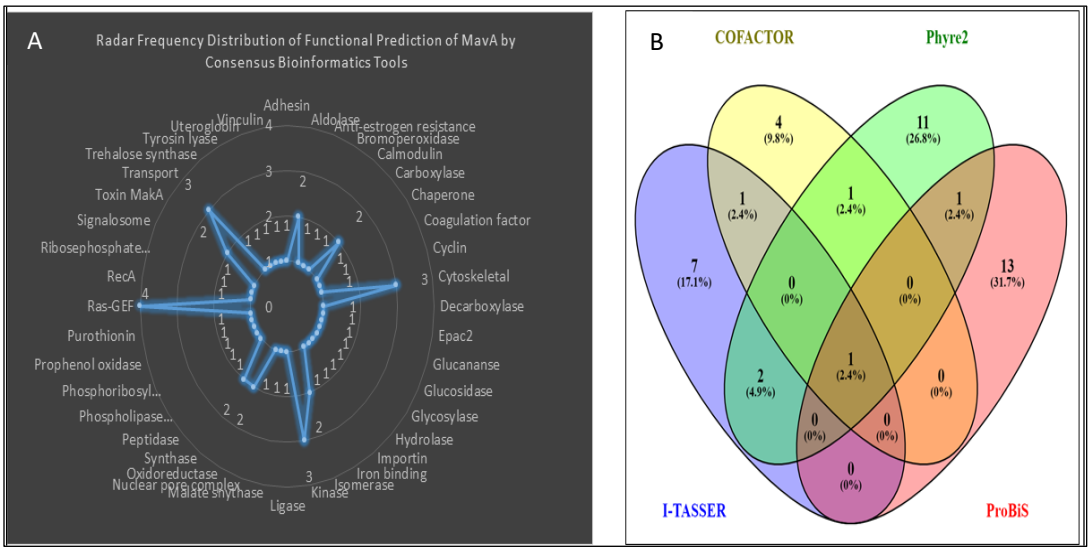


Figure 39: Functional Consensus Prediction of MavA Protein: A. Radar frequency distribution of all functional predictions by five bioinformatics tools. Four properties were predicted for MavA by at least three independent bioinformatics tools. B. Venn diagram showing the intersection of the predictions made by four bioinformatics tools. The share functional prediction for MavA by the four tools (COFACTOR, I-TASSER, Phrey2, and ProBiS) is RasGEF.

4.2.4. Sequence Alignment of MavA Protein with other RasGEF Proteins

The potential functional domain of the MavA protein was also examined using likely homologs and orthologs in the protein database. The OrthoDB defines orthologs (which are descendant genes from a single gene of the ancestor) by speculating the function of the genes based on the related function performed in the ancestor [199]. The hierarchical catalog of orthologs shows that the most recent and distant protein homolog to MavA is a Ras guanine nucleotide exchange factor (Ras-GEF) superfamily member. Domain search in the MavA amino acid sequence on the Uniprotkb

predictive annotation utilize the properties of characterized proteins in its database to predict that MavA protein has the potential for Ras-GEF activity, although the annotation score is 20%.

To further analyze MavA protein on its prospective RasGEF property, other *L. pneumophila* proteins that were known to possess RasGEF property were retrieved as FASTA files and sequence alignment using NCBI BLAST was carried out. The analysis showed that MavA protein is orthologous with *Legionella moravica*'s RasGEF (NCBI reference sequence ID: WP_028383452.1) at 47% identity and *Legionella worsleiensis*'s RasGEF (NCBI reference sequence ID: WP_058493972.1) at 43% identity, both with 400 and 402 amino acids residues respectively. CLUSTAL O (1.2.4) was used for multiple sequence alignment of the three proteins, as shown in Fig. 40.

```
MavA_Legionella_pneumophila       MSHNYFRTCLTTKDNKFFWQKPLASERFDKLIADRKFQKSIAHSRKNQLMYSFIESKIAS   60
RasGEF_Legionella_moravica        MRINKFQDWLNRNGNFQFLSNPISSEQIDTVIADKSVQKSIASSKKNQVIYSYIENKIVT   60
RasGEF_Legionella_worsleiensis    MPSNVFLEWLSQRGNLQFLFDPLCSEQLDKAIVSTSVQKSIAGSKKNQALYGYIENRIVT   60
                                  *  * *   *. ..*  *  .*:.**::*. *.. ..***** *:*** :*.:**.:*.:

MavA_Legionella_pneumophila       DYALLIDTKLREQLKQFTLDDFNDISGFERKEKSNSRQQSYFKLRQELEFFLKNDIQQHK   120
RasGEF_Legionella_moravica        DYASVIDFKLREVFYKLSLDDFRDPRGFESKNKSSYNQCNYFELRNELEFFLKQDILQHR   120
RasGEF_Legionella_worsleiensis    DYATLIDFKLREAFVRLTQEDFKDPKGFECKDKSSAHQNKYFTLRNELEYFLKKDILQHQ   120
                                  *** :** **** : ::: :**.*  *** *:**. .* .** **:***:***:** **:

MavA_Legionella_pneumophila       SCPDAQLNAFRRWINISDLLLRHHCYEGFLLVFVNLQLIADKQLIDGLPVSVRNNYNQLC   180
RasGEF_Legionella_moravica        ARPDAKLNAFRRWIEVSDVLLRRHCYEGFMLVFTNLQLIASPNLVSSLPQSVQKNYHELC   180
RasGEF_Legionella_worsleiensis    DSPDAKLNAFRRWIEIADVLLSRHCYEGFLLVFTNLQLIASPDLVRGLPQNIQNSYNELC   180
                                    ***:********:::*:** :******:***.******. :*: .** .:::.*::**

MavA_Legionella_pneumophila       QLSSPIGNHGALRHFMNTHQGESDFTPLFFTYHAIGALDESLESLKDKEVLLKKQLKHLN   240
RasGEF_Legionella_moravica        QLNSPNKNHFALRDFIKTHSNETDFSPLIFSYHAIAVINESIVHIREQDLLLNNRKKEVR   240
RasGEF_Legionella_worsleiensis    HLNSPSKNHLALRSFIKNRTNEADFSPLILIYHAIAMINESIVHIREQDIVLKSRKKQVR   240
                                  :*.**  ** *** *::.: .*:**:**:: ****. ::**:  :::::::*:.: *.:.

MavA_Legionella_pneumophila       KKISHLRTEATPG-VIDVISELLKNKQQIPKKMMERRGHLIQLMEEVRFTGKQLKQIQVN   299
RasGEF_Legionella_moravica        KEIHRLQNEIDPD-DFKTLVKLVGKHEHIPNDLNLENRYLISLLRENKRIPKNLRNNHKI   299
RasGEF_Legionella_worsleiensis    REISRLQKELDDDSTFRDLFQCVDHHQSVPGDLVSGNRYLISLLQENKRISKDLEQNQEI   300
                                  ::* :*:.*   .  :  : : : ::: :* .:   . :**.*:.* :   *:*.: :

MavA_Legionella_pneumophila       MQNQLEQRANLVGLIVKEQKATPRTLPEYFEKTYSIIKYRFNRQSIATVKSSQLPESTAS   359
RasGEF_Legionella_moravica        LCEQIEQRTELLNSITKEQNYRVKSLPSYLENTYNRIRFRFNKHNLERTMGIVPSEPARS   359
RasGEF_Legionella_worsleiensis    LSEQVKQRSDLLDTITKQQQYKPRPLSSHLEKSYHLIHCRYYKHNRDLKMGLISDERFSQ   360
                                  : :*:::**::*:. *.*:*:   : * .::*::*  *: *: ::.    .   *  .

MavA_Legionella_pneumophila       SNPSSSNLYKSKLLPHFWNRHGKSSSSYWEEVFTPSCLNNK-  400
RasGEF_Legionella_moravica        QDIAPSRLYSHSLLPSFWYRKGKSAEQYWKETLTPSCLPSR-  400
RasGEF_Legionella_worsleiensis    AYSLPSRLYSESLLPSFWNRRGSSAEDHWRAMMKPSCLEPAQ  402
                                      *.**. .*** ** *:*.*:..:*.  :.****
```

Figure 40: Aligning amino acid sequences of Ras-GEFs in other Legionella species and MavA protein using CLUSTAL O (1.2.4) multiple sequence alignment. The * shows positions in the three proteins with a single, fully conserved amino acid residue. For example, a colon: shows strongly conserved similar amino acid residues scoring > 0.5 in the Gonnet PAM 250 matrix, and a period'.' indicates weakly identical residues between the protein sequences scoring =< 0.5 in the Gonnet PAM 250 matrix. Residues shown as red are small and hydrophobic, those colored blue are acidic, magenta is primary, green are residues with hydroxyl, sulfhydryl, and amine groups, and grey has at least an imino group.

4.2.5. Potential Biological Process and Molecular Function of MavA Protein

Five (62.5%) of the individual models from Phyre2 and ProBiS are guanine nucleotide-releasing proteins representing a 62.5% probability rate.

The best-ranked model of the MavA protein in PDB file format was uploaded to the COFACTOR server to predict the biological processes affected by the protein. Figure 40 depicts the gene ontology for each biological process indicated with color-coded Cscore which range from 0.4 to 1.0 in increasing order of confidence. In addition, COFACTOR was used to evaluate the molecular function of the protein by predicting terms within the Gene Ontology hierarchy associated with known Molecular Functions with color-coded Cscore.

The biological processes prediction with COFACTOR shows that the protein primarily affects the regulation of the tricarboxylic acid cycle (Fig. 41a). In contrast, the predicted molecular function of the protein primarily involves molecular regulation and guanyl-nucleotide exchange factor activity in the cell (Fig. 41b).

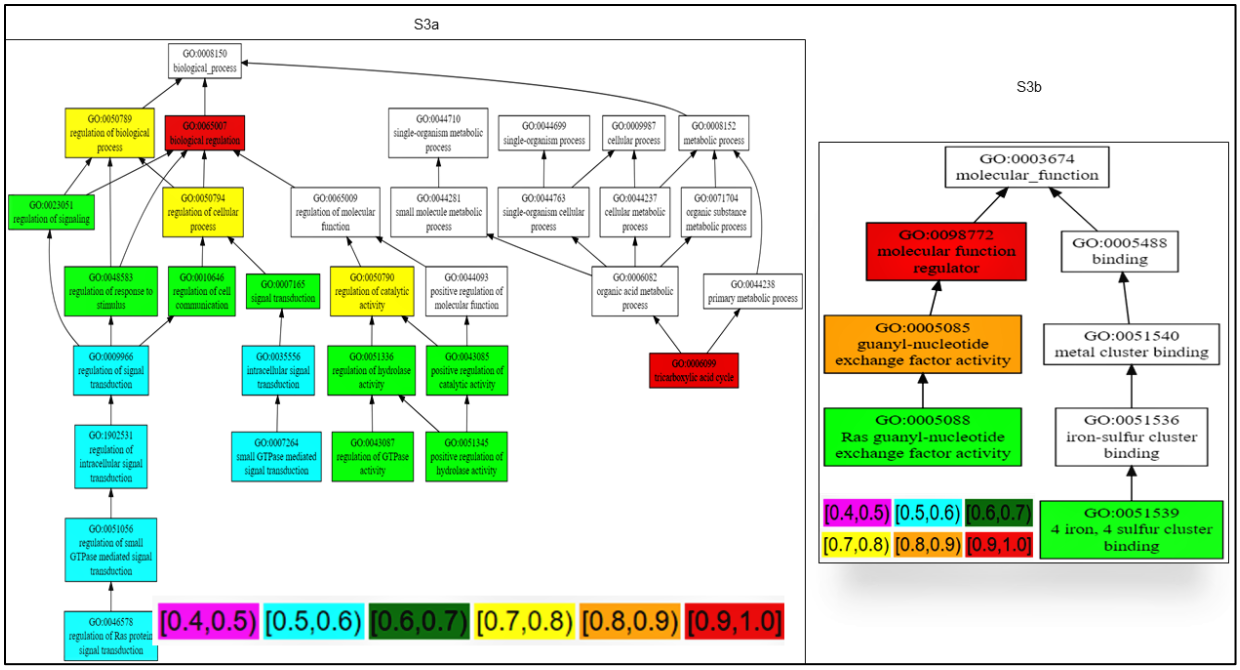


Figure 41: Gene Ontology Analysis of MavA Protein. (a) Biological Process of MavA. (b) Molecular Function of MavA. The confidence scores of sequence alignment of MavA and the template used range from 0.7 to 0.8.

4.2.6. Simulated Molecular Interaction of MavA with Selected Ligands

Based on the findings from the structural predictions and transcriptomic data, the interaction between tertiary structures (PDB files) of MavA protein with Ras protein, GTP, and carboxylase were simulated using Lzerd [164], CB-Dock [167] and HDOCK [200]. To evaluate the accuracy of docking geometry between MavA protein and each of the predicted ligands, the Root Mean Square Deviation (RMSD) of the reference position of the ligand and MavA was reported by the docking tools after ensuring optimal superimposition of the two structures. A RMSD value < 4 Å is considered a good index of potential interaction [201]. The analysis outcome showed a high possibility of interaction with Ras protein with high docking confidence and a low RMSD of 1.78 Å, the shortest length recorded between the MavA protein, and the ligands analyzed (Fig. 42a and

b). The best-ranked model shows a GOAP score and rank of -84721.80 and 51, respectively. It has a minimum RSMD value between atoms of 1.78Å between GLU63 of Ras protein and TYR328 of MavA protein. Other interactions include a bond between TYR64 of Ras protein and ARG323 of MavA protein with an RMSD of 2.627Å. The binding of GTP with MavA protein has a RMSD value of 1.931Å at PHE 17 residue. Interaction of the MavA protein with carboxylase is seen on the same pocket in the tertiary structure of the MavA protein. The MavA protein and actin filament were predicted to interact with each other by simulated interaction with a minimum RMSD value of 1.562 Å (Fig. 42c).

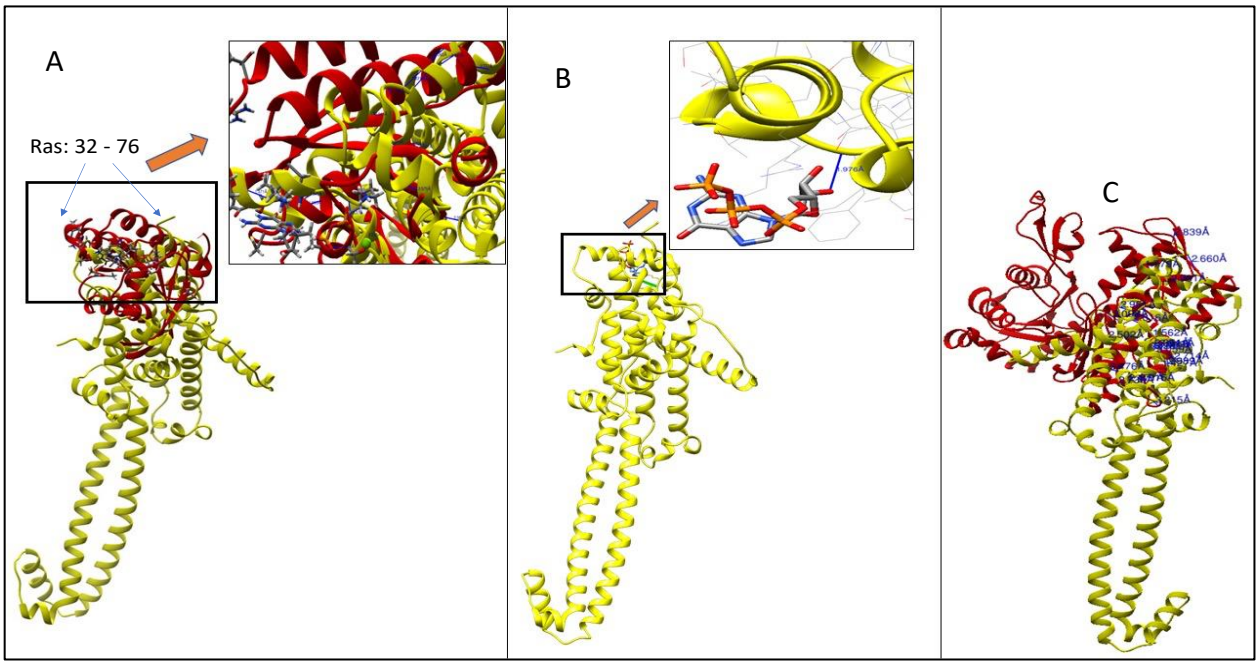


Figure 42: Predicted molecular interaction between MavA (yellow) and Ras Protein (red). A: MavA and Ras/GTP interaction simulation using lzerd docking server. B: MavA Interaction with GTP. MavA docked with GTP shows that the GTP is attached to a pocket in the MavA structure for a successful transfer to the Ras protein. C: Docking showing the interaction between the actin cytoskeleton and MavA protein.

DISCUSSION

With several open-access server, the tertiary structure of MavA protein was predicted to be made up of nineteen alpha helices, four beta strands and coil-coil domain. With the WaggaWagga server, the prediction of coil-coil domain in MavA protein was emphasized at the protein C-terminal. There is a speculation that the coil-coil domains of MavA (Ras—guanine exchange factor) may play a role in the biological role of the MavA protein because the coil-coil domain in protein have been identified to confer a high likelihood of nucleotide binding capacity irrespective of species or length and have been a crucial domain that dictates the functions of resident proteins [202]. For instance, the active site of LegC3, another effector protein in *L. pneumophila*, possesses the coil-coil domain. The LegC3 protein do prevent the homotypic fusion of yeast vacuoles, which heavily depends on the SNARE and Rab-GTPase proteins [203]. The coiled-coil domain is also mentioned as a crucial conserved area of proteins involved in cytoskeletal remodeling and cell adhesion during *L. pneumophila* infection is the coil-coil domain [204].[203][204] This hypothesis was supported by a robust contact between the MavA protein and the actin filament on SwissDock, with an RMSD value of 1.562 Å.

Seven similar proteins to MavA protein in other species of Legionella were identified however, none of these have been functionally described (Fig. 2a). This suggest that MavA protein may be a crucial protein that is conserved by natural selection during the speciation of Legionella, as shown in the conservation of the protein within the genera. However, the uncharacterized status of these neighbors made it impossible to define the function of the MavA protein based solely on the amino acid sequence. However, a hypothesis of the RasGEF activity of the MavA protein was made by combining analytical servers that require the FASTA amino acid residues with other structure-dependent servers. Such an approach involved the alignment of MavA protein's sequence

with two proteins from Legionella species that have been identified as RasGEF proteins and able to see that the N terminal of the proteins has been significantly conserved. There has been a conserved N-terminal RasGEF in most characterized RasGEF proteins [205].

The molecular docking of MavA with the predicted ligands shows that the MavA binding sites with GTP is possibly located at the C terminal. Since Ras protein is mainly found in eukaryotic cells the identification of possible RasGEF containing effector protein in *L. pneumophila* provides evidence for horizontal transfer of genes to the bacteria as a result of co-evolution [206]. MavA protein tertiary structure was docked with Ras protein using Lzerd, which showed that potential interactions with human Ras protein through hydrogen bonding at crucial residues on the Ras protein, including positions 32–40 and 60–76 [207]. The purine nucleotide (GTP) is predicted to binds to the MavA protein at position F17 at a close bond length of 1.976 Å RSMD. The GTP was the dominant ligand of the protein according to our data for analyzing the exchange of GTP for GDP by CB-Dock. MavA protein, Ras protein, and GTP are predicted to interact with an RSMD value of less than 2.0 Å, which suggests a high affinity in their interactions.

Ectopically expressed MavA revealed differential regulation of seventeen genes, including sixteen upregulated genes and one downregulated gene. The relationship between these genes helped better inform the molecular function of the MavA protein. For instance, the prediction of the interaction between MavA protein and actin is also seen in the upregulation of cilia and flagella associated protein (CFAP20) in the cell [208].

Evidence for the activation of the adrenergic receptor in the host by the MavA protein and mRNA splicing, respectively, was provided by the differential expression of the ARRDC3 and BUD31 genes. To control the sympathetic nervous system and prevent hyperinflammation and lower tract infections, the adrenergic receptors, which are known to be linked to G-protein coupled

receptors, bind to, and activate the neurotransmitters norepinephrine and epinephrine [211]. This suggest that it could be of importance for the pathogen by making increased ARRDC3 receptor in the macrophage lead to hypersensitivity to inflammatory signal.

Sortilin receptor protein (SORL1), a retrograde membrane cargo protein, is the sole gene in the MavA-expressing HEK293 cell that has been differently downregulated. In humans, the SORL1 is known to play important role in protein ubiquitination, endocytosis, and protein degradation [215]. According to a study by Canuel et al., cathepsin D is retained in the cells independent of the mannose 6-phosphate receptor (M6PR) pathway when the sortilin receptor protein is missing, which disrupts the creation of lysosomes [216]. This suggests that the sortilin protein contributes significantly to the development of lysosomes [215]. Sortilin is also known to localize on the LCV in the absence of the RidL effector protein in *L. pneumophila*, according to a study by Bärlocher and colleagues. This localization help encourage bacterial replication inside the host cells, through avoidance of retrograde membrane trafficking by the RidL effector protein [217]. Based on this, sortilin downregulation could cause MavA protein to impair host retrograde membrane trafficking during infection and hinder lysosomal formation by making the essential sortilin receptor needed for the process unavailable [218], leading to increased *L. pneumophila* macrophage proliferation. Therefore, the modulation of the sortilin receptor revealed that in addition to the postulated endosomal remodeling and cytoskeletal reorganization by MavA, another complementing method used by the MavA protein is to shield the LCV of *L. pneumophila* from the lysosome.

In summary, the suggested molecular function of the MavA protein is guanine exchange factor and it is predicted to have coiled-coil domain, based on the bioinformatic analysis and transcriptome study of the protein. Ras protein is predicted to be activated by MavA protein's

RasGEF activity, which starts the cell signaling process required for cytoskeletal reorganization and endocytosis which are essential for LCV creation and bacterial internalization into the LCV. Additionally, previous study by Barlocher showed that during the intracellular replication of *Legionella pneumophila*, the sortilin receptor protein which point to area of future investigation as whether the downregulation of sortilin receptor gene by MavA protein is also seen at protein level (translation level), preventing the usual sorting of retrograde payloads from the early endosome to the lysosome in the host cell [217].

REFERENCES

[1] B. A. Wee *et al.*, "Population analysis of Legionella pneumophila reveals a basis for resistance to complement-mediated killing," *Nat Commun*, vol. 12, no. 1, p. 7165, Dec. 2021, doi: 10.1038/s41467-021-27478-z.

[2] F. W. Chandler, M. D. Hicklin, and J. A. Blackmon, "Demonstration of the agent of Legionnaire disease in tissue," *N Engl J Med*, vol. 297, no. 22, pp. 1218–1220, Dec. 1977, doi: 10.1056/NEJM197712012972206.

[3] C. Chidiac *et al.*, "Factors associated with hospital mortality in community-acquired legionellosis in France," *European Respiratory Journal*, vol. 39, no. 4, pp. 963–970, Apr. 2012, doi: 10.1183/09031936.00076911.

[4] W. C. Winn, "Legionnaire disease: historical perspective," *Clin Microbiol Rev*, vol. 1, no. 1, pp. 60–81, Jan. 1988, doi: 10.1128/CMR.1.1.60.

[5] J. Spiegelman, T. Pedutem, and M. J. Francisco, "Legionnaire disease Cases at a Large Community Hospital—Common and Underdiagnosed," *Int J Environ Res Public Health*, vol. 17, no. 1, p. 332, Jan. 2020, doi: 10.3390/ijerph17010332.

[6] M. J. Donohue, J. H. Mistry, N. Tucker, and S. J. Vesper, "Hot water plumbing in residences and office buildings have distinctive risk of Legionella pneumophila contamination," *Int J Hyg Environ Health*, vol. 245, p. 114023, Aug. 2022, doi: 10.1016/j.ijheh.2022.114023.

[7] J. O. Falkinham, E. D. Hilborn, M. J. Arduino, A. Pruden, and M. A. Edwards, "Epidemiology and Ecology of Opportunistic Premise Plumbing Pathogens: *Legionella pneumophila*, *Mycobacterium avium*, and *Pseudomonas aeruginosa*," *Environ Health Perspect*, vol. 123, no. 8, pp. 749–758, Aug. 2015, doi: 10.1289/ehp.1408692.

[8] S. A. Collier *et al.*, “Estimate of Burden and Direct Healthcare Cost of Infectious Waterborne Disease in the United States,” *Emerg Infect Dis*, vol. 27, no. 1, pp. 140–149, Jan. 2021, doi: 10.3201/eid2701.190676.

[9] N. Saini and R. S. Gupta, “A robust phylogenetic framework for members of the order Legionellales and its main genera (Legionella, Aquicella, Coxiella and Rickettsiella) based on phylogenomic analyses and identification of molecular markers demarcating different clades,” *Antonie Van Leeuwenhoek*, vol. 114, no. 7, pp. 957–982, Jul. 2021, doi: 10.1007/s10482-021-01569-9.

[10] R. Lesnik, I. Brettar, and M. G. Höfle, “Legionella species diversity and dynamics from surface reservoir to tap water: from cold adaptation to thermophily,” *ISME J*, vol. 10, no. 5, pp. 1064–1080, May 2016, doi: 10.1038/ismej.2015.199.

[11] F. David W *et al*. “Legionnaires’ Disease.” *New England Journal of Medicine*, vol. 297, no. 22, 1977, pp. 1189–1197, https://doi.org/10.1056/nejm197712012972201.

[12] I. Hasni *et al.*, “Intracellular Behaviour of Three Legionella pneumophila Strains within Three Amoeba Strains, Including Willaertia magna C2c Maky,” *Pathogens*, vol. 9, no. 2, p. 105, Feb. 2020, doi: 10.3390/pathogens9020105.

[13] J. W. Conlan and L. A. E. Ashworth, “The relationship between the serogroup antigen and lipopolysaccharide of Legionella pneumophila,” *J. Hyg., Camb*, vol. 96, pp. 39–48, 1986, doi: 10.1017/S0022172400062513.

[14] D. M. Pierre, J. Baron, V. L. Yu, and J. E. Stout, “Diagnostic testing for Legionnaire disease,” *Ann Clin Microbiol Antimicrob*, vol. 16, no. 1, p. 59, Dec. 2017, doi: 10.1186/s12941-017-0229-6.

[15] G. D'Auria, N. Jiménez-Hernández, F. Peris-Bondia, A. Moya, and A. Latorre, "Legionella pneumophila pangenome reveals strain-specific virulence factors," *BMC Genomics*, vol. 11, no. 1, p. 181, 2010, doi: 10.1186/1471-2164-11-181.

[16] WHO, "LEGIONELLA and the prevention of Legionellosis," Geneva, 2007. Accessed: Mar. 24, 2023. [Online]. Available: http://apps.who.int/iris/bitstream/handle/10665/43233/9241562978_eng.pdf;jsessionid=C3F60BD14B7380EFB22159CB43766A44?sequence=1

[17] "Regulations and Guidelines on Legionella Control in Water Systems - Management of Legionella in Water Systems - NCBI Bookshelf." https://www.ncbi.nlm.nih.gov/books/NBK555112/ (accessed Jun. 25, 2023).

[18] R. R. Isberg, T. J. O'Connor, and M. Heidtman, "The Legionella pneumophila replication vacuole: making a cosy niche inside host cells," *Nat Rev Microbiol*, vol. 7, no. 1, pp. 13–24, Jan. 2009, doi: 10.1038/nrmicro1967.

[19] S. S. Weber, C. Ragaz, K. Reus, Y. Nyfeler, and H. Hilbi, "Legionella pneumophila Exploits PI(4)P to Anchor Secreted Effector Proteins to the Replicative Vacuole," *PLoS Pathog*, vol. 2, no. 5, p. e46, May 2006, doi: 10.1371/JOURNAL.PPAT.0020046.

[20] A. Chong, C. A. Lima, D. S. Allan, G. K. Nasrallah, and R. A. Garduño, "The Purified and Recombinant *Legionella pneumophila* Chaperonin Alters Mitochondrial Trafficking and Microfilament Organization," *Infect Immun*, vol. 77, no. 11, pp. 4724–4739, Nov. 2009, doi: 10.1128/IAI.00150-09.

[21] A. Best, C. Price, M. Ozanic, M. Santic, S. Jones, and Y. Abu Kwaik, "A Legionella pneumophila amylase is essential for intracellular replication in human macrophages and amoebae," *Sci Rep*, vol. 8, no. 1, p. 6340, Apr. 2018, doi: 10.1038/s41598-018-24724-1.

[22] S. Ninio, J. Celli, and C. R. Roy, "A Legionella pneumophila Effector Protein Encoded in a Region of Genomic Plasticity Binds to Dot/Icm-Modified Vacuoles," *PLoS Pathog*, vol. 5, no. 1, p. e1000278, Jan. 2009, doi: 10.1371/journal.ppat.1000278.

[23] D. Chauhan and S. R. Shames, "Pathogenicity and Virulence of Legionella: Intracellular replication and host response," *Virulence*, vol. 12, no. 1, p. 1122, 2021, doi: 10.1080/21505594.2021.1903199.

[24] C. T. Price *et al.*, "Molecular Mimicry by an F-Box Effector of Legionella pneumophila Hijacks a Conserved Polyubiquitination Machinery within Macrophages and Protozoa," *PLoS Pathog*, vol. 5, no. 12, p. e1000704, Dec. 2009, doi: 10.1371/JOURNAL.PPAT.1000704.

[25] L. Gomez-Valero, C. Rusniok, C. Cazalet, and C. Buchrieser, "Comparative and Functional Genomics of Legionella Identified Eukaryotic Like Proteins as Key Players in Host? Pathogen Interactions," *Front Microbiol*, vol. 2, 2011, doi: 10.3389/fmicb.2011.00208.

[26] S. R. Shames, L. Liu, J. C. Havey, W. B. Schofield, A. L. Goodman, and C. R. Roy, "Multiple *Legionella pneumophila* effector virulence phenotypes revealed through high-throughput analysis of targeted mutant libraries," *Proceedings of the National Academy of Sciences*, vol. 114, no. 48, Nov. 2017, doi: 10.1073/pnas.1708553114.

[27] D. C. Lockwood, H. Amin, T. R. D. Costa, and G. N. Schroeder, "The Legionella pneumophila Dot/Icm type IV secretion system and its effectors," *Microbiology (N Y)*, vol. 168, no. 5, May 2022, doi: 10.1099/mic.0.001187.

[28] M. G. Rittig *et al.*, "Coiling phagocytosis is the preferential phagocytic mechanism for Borrelia burgdorferi," *Infect Immun*, vol. 60, no. 10, pp. 4205–4212, Oct. 1992, doi: 10.1128/iai.60.10.4205-4212.1992.

[29] A. Lührmann, C. V Nogueira, K. L. Carey, C. R. Roy, and R. R. Isberg, “Inhibition of pathogen-induced apoptosis by a Coxiella burnetii type IV effector protein,” *Source*, vol. 107, no. 44, pp. 18997–19001, 2010, doi: 10.1073/pnas.

[30] A. J. Allombert *et al.*, “Title Page Deciphering Legionella effector delivery by Icm/Dot secretion system reveals a new role for c-diGMP signaling”, doi: 10.1101/754762.

[31] W. Eisenreich and K. Heuner, “The life stage-specific pathometabolism of *Legionella pneumophila*,” *FEBS Lett*, vol. 590, no. 21, pp. 3868–3886, Nov. 2016, doi: 10.1002/1873-3468.12326.

[32] M. H. Saier, “A Functional-Phylogenetic Classification System for Transmembrane Solute Transporters,” *Microbiology and Molecular Biology Reviews*, vol. 64, no. 2, pp. 354–411, Jun. 2000, doi: 10.1128/MMBR.64.2.354-411.2000.

[33] Q. Ren, K. Chen, and I. T. Paulsen, “TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels,” *Nucleic Acids Res*, vol. 35, no. Database, pp. D274–D279, Jan. 2007, doi: 10.1093/nar/gkl925.

[34] S. Depluverez, S. Devos, and B. Devreese, “The Role of Bacterial Secretion Systems in the Virulence of Gram-Negative Airway Pathogens Associated with Cystic Fibrosis,” *Front Microbiol*, vol. 7, Aug. 2016, doi: 10.3389/fmicb.2016.01336.

[35] V. Waters, “New Treatments for Emerging Cystic Fibrosis Pathogens other than Pseudomonas,” *Curr Pharm Des*, vol. 18, no. 5, pp. 696–725, Feb. 2012, doi: 10.2174/138161212799315939.

[36] S. M. Hinsa, M. Espinosa-Urgel, J. L. Ramos, and G. A. O’Toole, “Transition from reversible to irreversible attachment during biofilm formation by Pseudomonas fluorescens WCS365 requires an ABC transporter and a large secreted protein,” *Mol Microbiol*, vol. 49, no. 4, pp. 905–918, Jul. 2003, doi: 10.1046/j.1365-2958.2003.03615.x.

[37] O. Spitz *et al.*, "Identity Determinants of the Translocation Signal for a Type 1 Secretion System," *Front Physiol*, vol. 12, p. 804646, Feb. 2022, doi: 10.3389/FPHYS.2021.804646/BIBTEX.

[38] F. Fuche, A. Vianney, C. Andrea, P. Doublet, and C. Gilbert, "Functional Type 1 Secretion System Involved in Legionella pneumophila Virulence," *J Bacteriol*, vol. 197, no. 3, p. 563, 2015, doi: 10.1128/JB.02164-14.

[39] T. Qin, H. Zhou, H. Ren, and W. Liu, "Distribution of secretion systems in the genus Legionella and its correlation with pathogenicity," *Front Microbiol*, vol. 8, no. MAR, Mar. 2017, doi: 10.3389/FMICB.2017.00388/FULL.

[40] S. Letoffe, V. Redeker, and C. Wandersman, "Isolation and characterization of an extracellular haem-binding protein from Pseudomonas aeruginosa that shares function and sequence similarities with the Serratia marcescens HasA haemophore," *Mol Microbiol*, vol. 28, no. 6, pp. 1223–1234, Jun. 1998, doi: 10.1046/j.1365-2958.1998.00885.x.

[41] F. Duong, E. Bonnet, V. Géli, A. Lazdunski, M. Murgier, and A. Filloux, "The AprX protein of *Pseudomonas aeruginosa:* a new substrate for the Apr type I secretion system," *Gene*, vol. 262, no. 1–2, pp. 147–153, Jan. 2001, doi: 10.1016/S0378-1119(00)00541-2.

[42] D. N. Harland, E. Dassa, R. W. Titball, K. A. Brown, and H. S. Atkins, "ATP-binding cassette systems in *Burkholderia pseudomallei* and *Burkholderia mallei*," *BMC Genomics*, vol. 8, no. 1, p. 83, Dec. 2007, doi: 10.1186/1471-2164-8-83.

[43] M. Ferrer-Navarro *et al.*, "Abundance of the Quorum-Sensing Factor Ax21 in Four Strains of *Stenotrophomonas maltophilia* Correlates with Mortality Rate in a New Zebrafish Model of Infection," *PLoS One*, vol. 8, no. 6, p. e67207, Jun. 2013, doi: 10.1371/journal.pone.0067207.

[44] N. H. Carbonetti, "Pertussis toxin and adenylate cyclase toxin: key virulence factors of *Bordetella pertussis* and cell biology tools," *Future Microbiol*, vol. 5, no. 3, pp. 455–469, Mar. 2010, doi: 10.2217/fmb.09.133.

[45] A. Cameron, R. Zaheer, E. H. Adator, R. Barbieri, T. Reuter, and T. A. McAllister, "Bacteriocin Occurrence and Activity in Escherichia coli Isolated from Bovines and Wastewater," *Toxins (Basel)*, vol. 11, no. 8, p. 475, Aug. 2019, doi: 10.3390/toxins11080475.

[46] D. N. Atapattu and C. J. Czuprynski, "*Mannheimia haemolytica* Leukotoxin Binds to Lipid Rafts in Bovine Lymphoblastoid Cells and Is Internalized in a Dynamin-2- and Clathrin-Dependent Manner," *Infect Immun*, vol. 75, no. 10, pp. 4719–4727, Oct. 2007, doi: 10.1128/IAI.00534-07.

[47] A. K. May, T. G. Gleason, R. G. Sawyer, and T. L. Pruett, "Contribution of *Escherichia coli* Alpha-Hemolysin to Bacterial Virulence and to Intraperitoneal Alterations in Peritonitis," *Infect Immun*, vol. 68, no. 1, pp. 176–183, Jan. 2000, doi: 10.1128/IAI.68.1.176-183.2000.

[48] F. Fuche, A. Vianney, C. Andrea, P. Doublet, and C. Gilbert, "Functional Type 1 Secretion System Involved in Legionella pneumophila Virulence," *J Bacteriol*, vol. 197, no. 3, pp. 563–571, Feb. 2015, doi: 10.1128/JB.02164-14.

[49] K. V. Korotkov, M. Sandkvist, and W. G. J. Hol, "The type II secretion system: biogenesis, molecular architecture and mechanism," *Nat Rev Microbiol*, vol. 10, no. 5, p. 336, May 2012, doi: 10.1038/NRMICRO2762.

[50] R. C. White and N. P. Cianciotto, "Assessing the impact, genomics and evolution of type II secretion across a large, medically important genus: the Legionella type II secretion paradigm," *Microb Genom*, vol. 5, no. 6, pp. 1–30, Jul. 2019, doi: 10.1099/MGEN.0.000273.

[51] L. G. De Masi, C. D. Sturey, J. A. Lieberman, and M. S. Donnenberg, "The type 2 secretion and type 4 pilus systems of Escherichia coli," *Escherichia coli: Pathotypes and Principles of*

Pathogenesis: Second Edition, pp. 387–416, Aug. 2013, doi: 10.1016/B978-0-12-397048-0.00013-9.

[52] K. V. Korotkov, M. Sandkvist, and W. G. J. Hol, "The type II secretion system: biogenesis, molecular architecture and mechanism," *Nat Rev Microbiol*, vol. 10, no. 5, p. 336, May 2012, doi: 10.1038/NRMICRO2762.

[53] S. P. Howard *et al.*, "Structure and assembly of pilotin-dependent and -independent secretins of the type II secretion system," *PLoS Pathog*, vol. 15, no. 5, p. e1007731, May 2019, doi: 10.1371/JOURNAL.PPAT.1007731.

[54] A. Fulara, I. Vandenberghe, R. J. Read, B. Devreese, and S. N. Savvides, "Structure and oligomerization of the periplasmic domain of GspL from the type II secretion system of Pseudomonas aeruginosa," *Sci Rep*, vol. 8, no. 1, Dec. 2018, doi: 10.1038/S41598-018-34956-W.

[55] T. R. D. Costa *et al.*, "Secretion systems in Gram-negative bacteria: structural and mechanistic insights," *Nature Publishing Group*, vol. 13, p. 343, 2015, doi: 10.1038/nrmicro3456.

[56] M. Sandkvist *et al.*, "General secretion pathway (eps) genes required for toxin secretion and outer membrane biogenesis in Vibrio cholerae," *J Bacteriol*, vol. 179, no. 22, pp. 6994–7003, 1997, doi: 10.1128/JB.179.22.6994-7003.1997.

[57] B. Wretlind and O. R. Pavlovskis, "Genetic mapping and characterization of Pseudomonas aeruginosa mutants defective in the formation of extracellular proteins," *J Bacteriol*, vol. 158, no. 3, pp. 801–808, 1984, doi: 10.1128/JB.158.3.801-808.1984.

[58] K. R. Hardie, A. Schulze, M. W. Parker, and J. T. Buckley, "Vibrio spp. secrete proaerolysin as a folded dimer without the need for disulphide bond formation," *Mol Microbiol*, vol. 17, no. 6, pp. 1035–1044, 1995, doi: 10.1111/J.1365-2958.1995.MMI_17061035.X.

[59] C. M. Harding, R. L. Kinsella, L. D. Palmer, E. P. Skaar, and M. F. Feldman, “Medically Relevant Acinetobacter Species Require a Type II Secretion System and Specific Membrane-Associated Chaperones for the Export of Multiple Substrates and Full Virulence,” *PLoS Pathog*, vol. 12, no. 1, 2016, doi: 10.1371/JOURNAL.PPAT.1005391.

[60] T. L. Johnson, U. Waack, S. Smith, H. Mobley, and M. Sandkvist, “Acinetobacter baumannii Is Dependent on the Type II Secretion System and Its Substrate LipA for Lipid Utilization and In Vivo Fitness,” *J Bacteriol*, vol. 198, no. 4, pp. 711–719, 2015, doi: 10.1128/JB.00622-15.

[61] M. Lindeberg and A. Collmer, “Analysis of eight out genes in a cluster required for pectic enzyme secretion by Erwinia chrysanthemi: sequence comparison with secretion genes from other gram-negative bacteria,” *J Bacteriol*, vol. 174, no. 22, pp. 7385–7397, 1992, doi: 10.1128/JB.174.22.7385-7397.1992.

[62] V. Aragon, S. Kurtz, and N. P. Cianciotto, “Legionella pneumophila major acid phosphatase and its role in intracellular infection,” *Infect Immun*, vol. 69, no. 1, pp. 177–185, 2001, doi: 10.1128/IAI.69.1.177-185.2001.

[63] M. Wilton, T. W. R. Halverson, L. Charron-Mazenod, M. D. Parkins, and S. Lewenza, “Secreted Phosphatase and Deoxyribonuclease Are Required by Pseudomonas aeruginosa To Defend against Neutrophil Extracellular Traps,” *Infect Immun*, vol. 86, no. 9, Sep. 2018, doi: 10.1128/IAI.00403-18.

[64] T. J. DiChristina, C. M. Moore, and C. A. Haller, “Dissimilatory Fe (III) and Mn (IV) reduction by Shewanella putrefaciens requires ferE, a homolog of the pulE (gspE) type II protein secretion gene,” *J Bacteriol*, vol. 184, no. 1, pp. 142–151, 2002, doi: 10.1128/JB.184.1.142-151.2002.

[65] C. Duncan *et al.*, “Lcl of Legionella pneumophila is an immunogenic GAG binding adhesin that promotes interactions with lung epithelial cells and plays a crucial role in biofilm formation,” *Infect Immun*, vol. 79, no. 6, pp. 2168–2181, Jun. 2011, doi: 10.1128/IAI.01304-10.

[66] P. J. Christie, N. Whitaker, and C. González-Rivera, “Mechanism and structure of the bacterial type IV secretion systems,” *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1843, no. 8, pp. 1578–1591, Aug. 2014, doi: 10.1016/j.bbamcr.2013.12.019.

[67] M. J. Sheedlo, M. D. Ohi, D. B. Lacy, and T. L. Cover, “Molecular architecture of bacterial type IV secretion systems,” *PLoS Pathog*, vol. 18, no. 8, p. e1010720, Aug. 2022, doi: 10.1371/journal.ppat.1010720.

[68] K. Wallden, A. Rivera-Calzada, and G. Waksman, “Microreview: Type IV secretion systems: versatility and diversity in function,” *Cell Microbiol*, vol. 12, no. 9, pp. 1203–1212, Sep. 2010, doi: 10.1111/j.1462-5822.2010.01499.x.

[69] R. Zhang, J. J. LiPuma, and C. F. Gonzalez, “Two type IV secretion systems with different functions in Burkholderia cenocepacia K56-2,” *Microbiology (N Y)*, vol. 155, no. 12, pp. 4005–4013, Dec. 2009, doi: 10.1099/mic.0.033043-0.

[70] D. P. Souza *et al.*, “Bacterial killing via a type IV secretion system,” *Nat Commun*, vol. 6, no. 1, p. 6453, Mar. 2015, doi: 10.1038/ncomms7453.

[71] M. Juhas *et al.*, “Novel Type IV Secretion System Involved in Propagation of Genomic Islands,” *J Bacteriol*, vol. 189, no. 3, pp. 761–771, Feb. 2007, doi: 10.1128/JB.01327-06.

[72] M. Juhas, D. W. Crook, and D. W. Hood, “Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence,” *Cell Microbiol*, vol. 10, no. 12, pp. 2377–2386, Dec. 2008, doi: 10.1111/j.1462-5822.2008.01187.x.

[73] H. Mikkelsen, K. Hui, N. Barraud, and A. Filloux, “The pathogenicity island encoded <scp>PvrSR</scp> / <scp>RcsCB</scp> regulatory network controls biofilm formation and dispersal in <*scp*>*P*</*scp*> *seudomonas aeruginosa* <scp>PA</scp> 14,” *Mol Microbiol*, vol. 89, no. 3, pp. 450–463, Aug. 2013, doi: 10.1111/mmi.12287.

[74] D. E. Voth, L. J. Broederdorf, and J. G. Graham, “Bacterial Type IV Secretion Systems: Versatile Virulence Machines,” *Future Microbiol*, vol. 7, no. 2, p. 241, Feb. 2012, doi: 10.2217/FMB.11.150.

[75] C. D. Vincent, J. R. Friedman, K. C. Jeong, E. C. Buford, J. L. Miller, and J. P. Vogel, “Identification of the core transmembrane complex of the Legionella Dot/Icm type IV secretion system,” *Mol Microbiol*, vol. 62, no. 5, pp. 1278–1291, Dec. 2006, doi: 10.1111/j.1365-2958.2006.05446.x.

[76] H. Kim *et al.*, “Structural basis for effector protein recognition by the Dot/Icm Type IVB coupling protein complex,” *Nat Commun*, vol. 11, no. 1, Dec. 2020, doi: 10.1038/S41467-020-16397-0.

[77] D. Chauhan and S. R. Shames, “Pathogenicity and Virulence of Legionella: Intracellular replication and host response,” *Virulence*, vol. 12, no. 1, p. 1122, 2021, doi: 10.1080/21505594.2021.1903199.

[78] A. Meir, D. Chetrit, L. Liu, C. R. Roy, and G. Waksman, “Legionella DotM structure reveals a role in effector recruiting to the Type 4B secretion system,” *Nat Commun*, vol. 9, no. 1, Dec. 2018, doi: 10.1038/S41467-017-02578-X.

[79] P. J. Christie, “Structural biology: Loading T4SS substrates,” *Nat Microbiol*, vol. 2, no. 9, p. 17125, Aug. 2017, doi: 10.1038/nmicrobiol.2017.125.

[80] E. Fernandez-Moreira, J. H. Helbig, and M. S. Swanson, “Membrane vesicles shed by Legionella pneumophila inhibit fusion of phagosomes with lysosomes,” *Infect Immun*, vol. 74, no. 6, pp.

3285–3295, Jun. 2006, doi: 10.1128/IAI.01382-05/ASSET/334B20E5-E80C-49C7-93EA-91878CE319E1/ASSETS/GRAPHIC/ZII0060659350009.JPEG.

[81] K. Wennerberg, K. L. Rossman, and C. J. Der, "The Ras superfamily at a glance," *J Cell Sci*, vol. 118, no. 5, pp. 843–846, Mar. 2005, doi: 10.1242/jcs.01660.

[82] L. Goitre, E. Trapani, L. Trabalzini, and S. F. Retta, "The Ras Superfamily of Small GTPases: The Unlocked Secrets," 2014, pp. 1–18. doi: 10.1007/978-1-62703-791-4_1.

[83] J. H. Lin, P. Walter, and T. S. B. Yen, "Endoplasmic Reticulum Stress in Disease Pathogenesis," *Annu Rev Pathol*, vol. 3, p. 399, 2008, doi: 10.1146/ANNUREV.PATHMECHDIS.3.121806.151434.

[84] R. G. Hodge, A. Schaefer, S. V. Howard, and C. J. Der, "RAS and RHO family GTPase mutations in cancer: twin sons of different mothers?," *Crit Rev Biochem Mol Biol*, vol. 55, no. 4, pp. 386–407, Jul. 2020, doi: 10.1080/10409238.2020.1810622.

[85] H. R. Bourne, D. A. Sanders, and F. McCormick, "The GTPase superfamily: conserved structure and molecular mechanism," *Nature*, vol. 349, no. 6305, pp. 117–127, Jan. 1991, doi: 10.1038/349117a0.

[86] J. L. Bos, H. Rehmann, and A. Wittinghofer, "GEFs and GAPs: critical elements in the control of small G proteins.," *Cell*, vol. 129, no. 5, pp. 865–77, Jun. 2007, doi: 10.1016/j.cell.2007.05.018.

[87] S. Paone and A. Olivieri, "Role of Host Small GTPases in Apicomplexan Parasite Infection," *Microorganisms*, vol. 10, no. 7, Jul. 2022, doi: 10.3390/MICROORGANISMS10071370.

[88] J. E. Jun, I. Rubio, and J. P. Roose, "Regulation of Ras Exchange Factors and Cellular Localization of Ras Activation by Lipid Messengers in T Cells," *Front Immunol*, vol. 4, 2013, doi: 10.3389/fimmu.2013.00239.

[89] J. Paez Valencia, K. Goodman, and M. S. Otegui, “Endocytosis and Endosomal Trafficking in Plants,” *Annu Rev Plant Biol*, vol. 67, pp. 309–335, Apr. 2016, doi: 10.1146/ANNUREV-ARPLANT-043015-112242.

[90] S. Schoebel, L. K. Oesterlin, W. Blankenfeldt, R. S. Goody, and A. Itzen, “RabGDI Displacement by DrrA from Legionella Is a Consequence of Its Guanine Nucleotide Exchange Activity,” *Mol Cell*, vol. 36, no. 6, pp. 1060–1072, Dec. 2009, doi: 10.1016/j.molcel.2009.11.014.

[91] Y. Tan, R. J. Arnold, and Z.-Q. Luo, “*Legionella pneumophila* regulates the small GTPase Rab1 activity by reversible phosphorylcholination,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 52, pp. 21212–21217, Dec. 2011, doi: 10.1073/pnas.1114023109.

[92] A. Mousnier *et al.*, “A New Method to Determine *In Vivo* Interactomes Reveals Binding of the Legionella pneumophila Effector PieE to Multiple Rab GTPases,” *mBio*, vol. 5, no. 4, Aug. 2014, doi: 10.1128/mBio.01148-14.

[93] A. H. Gaspar and M. P. Machner, “VipD is a Rab5-activated phospholipase a1 that protects Legionella pneumophila from endosomal fusion,” *Proc Natl Acad Sci U S A*, vol. 111, no. 12, pp. 4560–4565, Mar. 2014, doi: 10.1073/PNAS.1316376111/-/DCSUPPLEMENTAL.

[94] R. Bulgin *et al.*, “Bacterial Guanine Nucleotide Exchange Factors SopE-Like and WxxxE Effectors,” *Infect Immun*, vol. 78, no. 4, pp. 1417–1425, Apr. 2010, doi: 10.1128/IAI.01250-09.

[95] C. Lim *et al.*, “Crystal structure of a guanine nucleotide exchange factor encoded by the scrub typhus pathogen *Orientia tsutsugamushi*,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30380–30390, Dec. 2020, doi: 10.1073/pnas.2018163117.

[96] S. Bagatella *et al.*, “Bovine neutrophil chemotaxis to Listeria monocytogenes in neurolisteriosis depends on microglia-released rather than bacterial factors,” *J Neuroinflammation*, vol. 19, no. 1, pp. 1–18, Dec. 2022, doi: 10.1186/S12974-022-02653-1/FIGURES/5.

[97] P.; K. H. ; S. R. ; K. A. ; G. M. ; S. K. ; T. A. K. ; S. Y. Chopra, "Nucleoside diphosphate kinase as GTPase-activating FEBS 28622 of Mycobacterium tuberculosis acts protein for Rho-GTPases," in *Nucleoside diphosphate kinase as GTPase-activating FEBS 28622 of Mycobacterium tuberculosis acts protein for Rho-GTPases*, FEBS, 2004, pp. 212–216.

[98] M.-D. Chilton *et al.*, "Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis," *Cell*, vol. 11, no. 2, pp. 263–271, Jun. 1977, doi: 10.1016/0092-8674(77)90043-5.

[99] H. Bierne and R. Pourpre, "Bacterial Factors Targeting the Nucleus: The Growing Family of Nucleomodulins," *Toxins (Basel)*, vol. 12, no. 4, p. 220, Mar. 2020, doi: 10.3390/toxins12040220.

[100] L. H. M. Le, L. Ying, and R. L. Ferrero, "Nuclear trafficking of bacterial effector proteins," *Cell Microbiol*, vol. 23, no. 6, Jun. 2021, doi: 10.1111/cmi.13320.

[101] H. E. Hanford, J. Von Dwingelo, and Y. Abu Kwaik, "Bacterial nucleomodulins: A coevolutionary adaptation to the eukaryotic command center," *PLoS Pathog*, vol. 17, no. 1, p. e1009184, Jan. 2021, doi: 10.1371/journal.ppat.1009184.

[102] D. Sviridov and M. Bukrinsky, "Interaction of pathogens with host cholesterol metabolism," *Curr Opin Lipidol*, vol. 25, no. 5, pp. 333–338, Oct. 2014, doi: 10.1097/MOL.0000000000000106.

[103] A. Chaumet, G. D. Wright, S. H. Seet, K. M. Tham, N. V. Gounko, and F. Bard, "Nuclear envelope-associated endosomes deliver surface proteins to the nucleus," *Nat Commun*, vol. 6, no. 1, p. 8218, Sep. 2015, doi: 10.1038/ncomms9218.

[104] A.-S. Stolle *et al.*, "T3SS-Independent Uptake of the Short-Trip Toxin-Related Recombinant NleC Effector of Enteropathogenic Escherichia coli Leads to NF-κB p65 Cleavage," *Front Cell Infect Microbiol*, vol. 7, Apr. 2017, doi: 10.3389/fcimb.2017.00119.

[105] A. Lange, R. E. Mills, C. J. Lange, M. Stewart, S. E. Devine, and A. H. Corbett, "Classical Nuclear Localization Signals: Definition, Function, and Interaction with Importin α," *Journal of Biological Chemistry*, vol. 282, no. 8, pp. 5101–5105, Feb. 2007, doi: 10.1074/jbc.R600026200.

[106] H. Zhao, Y. Zhang, P. Wu, J. Wang, and H. Li, "The Shigella type three secretion system effector OspF invades host nucleus by binding host importin α1," *World J Microbiol Biotechnol*, vol. 35, no. 5, p. 71, May 2019, doi: 10.1007/s11274-019-2635-8.

[107] L. Galluzzi *et al.*, "Molecular definitions of autophagy and related processes," *EMBO J*, vol. 36, no. 13, p. 1811, Jul. 2017, doi: 10.15252/EMBJ.201796697.

[108] H. Khalil *et al.*, "Interruption of Autophagosome Formation in Cardiovascular Disease, an Evidence for Protective Response of Autophagy," *Immunol Invest*, vol. 49, no. 3, pp. 249–263, Apr. 2020, doi: 10.1080/08820139.2019.1635619.

[109] D. R. Thomas, P. Newton, N. Lau, and H. J. Newton, "Interfering with Autophagy: The Opposing Strategies Deployed by Legionella pneumophila and Coxiella burnetii Effector Proteins," *Front Cell Infect Microbiol*, vol. 10, p. 1, Nov. 2020, doi: 10.3389/FCIMB.2020.599762.

[110] H. Behrouj *et al.*, "Epigenetic regulation of autophagy in coronavirus disease 2019 (COVID-19)," *Biochem Biophys Rep*, vol. 30, Jul. 2022, doi: 10.1016/J.BBREP.2022.101264.

[111] A. I. Abd El Maksoud *et al.*, "Methylomic Changes of Autophagy-Related Genes by Legionella Effector Lpg2936 in Infected Macrophages," *Front Cell Dev Biol*, vol. 7, Jan. 2020, doi: 10.3389/fcell.2019.00390.

[112] L. Mengue, M. Régnacq, W. Aucher, E. Portier, Y. Héchard, and A. Samba-Louaka, "Legionella pneumophila prevents proliferation of its natural host Acanthamoeba castellanii," *Sci Rep*, vol. 6, no. 1, p. 36448, Nov. 2016, doi: 10.1038/srep36448.

[113] Z. Ge *et al.*, "New Global Insights on the Regulation of the Biphasic Life Cycle and Virulence Via ClpP-Dependent Proteolysis in Legionella pneumophila," *Molecular & Cellular Proteomics*, vol. 21, no. 5, p. 100233, May 2022, doi: 10.1016/j.mcpro.2022.100233.

[114] R. Schuelein *et al.*, "Targeting of RNA Polymerase II by a nuclear *Legionella pneumophila* Dot/Icm effector SnpL," *Cell Microbiol*, vol. 20, no. 9, p. e12852, Sep. 2018, doi: 10.1111/cmi.12852.

[115] J. Von Dwingelo *et al.*, "Interaction of the Ankyrin H Core Effector of *Legionella* with the Host LARP7 Component of the 7SK snRNP Complex," *mBio*, vol. 10, no. 4, Aug. 2019, doi: 10.1128/mBio.01942-19.

[116] S. Allgood *et al.,* "Legionella Effector AnkX Disrupts Host Cell Endocytic Recycling in a Phosphocholination-Dependent Manner". *Front Cell Infect Microbiol*, vol. 8, no. 7, Sep 2017, doi: 10.3389/fcimb.2017.00397.

[117] X. Yu *et al.*, "Legionella effector AnkX interacts with host nuclear protein PLEKHN1," *BMC Microbiol*, vol. 18, no. 1, p. 5, Dec. 2018, doi: 10.1186/s12866-017-1147-7.

[118] M. K. Shanmugam *et al.*, "Role of novel histone modifications in cancer," *Oncotarget*, vol. 9, no. 13, pp. 11414–11426, Dec. 2017, doi: 10.18632/ONCOTARGET.23356.

[119] F. Cozzolino, I. Iacobucci, V. Monaco, T. Angrisano, and M. Monti, "Lysines Acetylome and Methylome Profiling of H3 and H4 Histones in Trichostatin A—Treated Stem Cells," *Int J Mol Sci*, vol. 22, no. 4, p. 2063, Feb. 2021, doi: 10.3390/ijms22042063.

[120] L. Gomez-Valero *et al.*, "Comparative analyses of Legionella species identifies genetic features of strains causing Legionnaire disease," *Genome Biol*, vol. 15, no. 11, p. 505, Nov. 2014, doi: 10.1186/s13059-014-0505-0.

[121] M. Rolando *et al.*, "Legionella pneumophila Effector RomA Uniquely Modifies Host Chromatin to Repress Gene Expression and Promote Intracellular Bacterial Replication," *Cell Host Microbe*, vol. 13, no. 4, pp. 395–405, Apr. 2013, doi: 10.1016/j.chom.2013.03.004.

[122] T. Li *et al.*, "SET-domain bacterial effectors target heterochromatin protein 1 to activate host rDNA transcription," *EMBO Rep*, vol. 14, no. 8, pp. 733–740, Aug. 2013, doi: 10.1038/embor.2013.86.

[123] M. K. Schuhmacher *et al.*, "The Legionella pneumophila Methyltransferase RomA Methylates Also Non-Histone Proteins during Infection," *J Mol Biol*, vol. 430, no. 13, pp. 1912–1925, Jun. 2018, doi: 10.1016/j.jmb.2018.04.032.

[124] P. Gujral, V. Mahajan, A. C. Lissaman, and A. P. Ponnampalam, "Histone acetylation and the role of histone deacetylases in normal cyclic endometrium," *Reproductive Biology and Endocrinology*, vol. 18, no. 1, p. 84, Dec. 2020, doi: 10.1186/s12958-020-00637-5.

[125] K. Biegeleisen, "A New Histone Structure Which Binds DNA at Its Eight Subunit N-Termini," *OAlib*, vol. 03, no. 02, pp. 1–20, 2016, doi: 10.4236/oalib.1102386.

[126] B. A. Tamburini and J. K. Tyler, "Localized Histone Acetylation and Deacetylation Triggered by the Homologous Recombination Pathway of Double-Strand DNA Repair," *Mol Cell Biol*, vol. 25, no. 12, pp. 4903–4913, Jun. 2005, doi: 10.1128/MCB.25.12.4903-4913.2005.

[127] L. Galdieri, T. Zhang, D. Rogerson, R. Lleshi, and A. Vancura, "Protein Acetylation and Acetyl Coenzyme A Metabolism in Budding Yeast," *Eukaryot Cell*, vol. 13, no. 12, pp. 1472–1483, Dec. 2014, doi: 10.1128/EC.00189-14.

[128] E. Kim *et al.*, "Histone and Non-Histone Targets of Dietary Deacetylase Inhibitors," *Curr Top Med Chem*, vol. 16, no. 7, pp. 714–731, Nov. 2015, doi: 10.2174/1568026615666150825125857.

[129] M. B. Eslaminejad, N. Fani, and M. Shahhoseini, “Epigenetic regulation of osteogenic and chondrogenic differentiation of mesenchymal stem cells in culture.,” *Cell J*, vol. 15, no. 1, pp. 1–10, 2013.

[130] A. M. Grabiec and J. Potempa, “Epigenetic regulation in bacterial infections: targeting histone deacetylases,” *Crit Rev Microbiol*, vol. 44, no. 3, pp. 336–350, May 2018, doi: 10.1080/1040841X.2017.1373063.

[131] X. Han *et al.*, “Epigenetic Regulation of Tumor Necrosis Factor α (TNFα) Release in Human Macrophages by HIV-1 Single-stranded RNA (ssRNA) Is Dependent on TLR8 Signaling,” *Journal of Biological Chemistry*, vol. 287, no. 17, pp. 13778–13786, Apr. 2012, doi: 10.1074/jbc.M112.342683.

[132] L.-P. Chan *et al.*, “IL-8 promotes inflammatory mediators and stimulates activation of p38 MAPK/ERK-NF-κB pathway and reduction of JNK in HNSCC,” *Oncotarget*, vol. 8, no. 34, pp. 56375–56388, Aug. 2017, doi: 10.18632/oncotarget.16914.

[133] B. Schmeck *et al.*, “Histone acetylation and flagellin are essential for Legionella pneumophila-induced cytokine expression,” *J Immunol*, vol. 181, no. 2, pp. 940–947, Jul. 2008, doi: 10.4049/JIMMUNOL.181.2.940.

[134] M. D. Fulton, T. Brown, and Y. G. Zheng, “Mechanisms and Inhibitors of Histone Arginine Methylation,” *Chem Rec*, vol. 18, no. 12, p. 1792, Dec. 2018, doi: 10.1002/TCR.201800082.

[135] M. Rolando and C. Buchrieser, “Legionella pneumophila type IV effectors hijack the transcription and translation machinery of the host cell,” *Trends Cell Biol*, vol. 24, no. 12, pp. 771–778, Dec. 2014, doi: 10.1016/j.tcb.2014.06.002.

[136] B. Schmeck *et al.*, “Histone Acetylation and Flagellin Are Essential for *Legionella pneumophila* -Induced Cytokine Expression,” *The Journal of Immunology*, vol. 181, no. 2, pp. 940–947, Jul. 2008, doi: 10.4049/jimmunol.181.2.940.

[137] M. Chien *et al.*, “The Genomic Sequence of the Accidental Pathogen *Legionella pneumophila*,” *Science (1979)*, vol. 305, no. 5692, pp. 1966–1968, Sep. 2004, doi: 10.1126/science.1099776.

[138] “Ras-GEF domain-containing protein - Legionella pneumophila subsp. pneumophila (strain Philadelphia 1 / ATCC 33152 / DSM 7513) | UniProtKB | UniProt.” https://www.uniprot.org/uniprotkb/Q5ZUV5/entry (accessed Jun. 12, 2023).

[139] “Type IV secretion protein Dot - Legionella pneumophila subsp. pneumophila (strain Philadelphia 1 / ATCC 33152 / DSM 7513) | UniProtKB | UniProt.” https://www.uniprot.org/uniprotkb/Q5ZW15/entry (accessed Jun. 12, 2023).

[140] O. Nevo, T. Zusman, M. Rasis, Z. Lifshitz, and G. Segal, “Identification of Legionella pneumophila Effectors Regulated by the LetAS-RsmYZ-CsrA Regulatory Cascade, Many of Which Modulate Vesicular Trafficking,” *J Bacteriol*, vol. 196, no. 3, pp. 681–692, Feb. 2014, doi: 10.1128/JB.01175-13.

[141] L. Huang *et al.*, “The E Block motif is associated with *Legionella pneumophila* translocated substrates,” *Cell Microbiol*, vol. 13, no. 2, pp. 227–245, Feb. 2011, doi: 10.1111/j.1462-5822.2010.01531.x.

[142] W. R. Pearson, “An Introduction to Sequence Similarity (‘Homology’) Searching,” *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]*, vol. 0 3, no. SUPPL.42, 2013, doi: 10.1002/0471250953.BI0301S42.

[143] I. Kapoor, R. Varada, S. Aroli, and U. Varshney, "Nudix hydrolases with Coenzyme A (CoA) and acyl-CoA pyrophosphatase activities confer growth advantage to Mycobacterium smegmatis," *Microbiology (N Y)*, vol. 165, no. 11, pp. 1219–1232, Nov. 2019, doi: 10.1099/mic.0.000850.

[144] M. Shimizu, S. Masuo, T. Fujita, Y. Doi, Y. Kamimura, and N. Takaya, "Hydrolase Controls Cellular NAD, Sirtuin, and Secondary Metabolites," *Mol Cell Biol*, vol. 32, no. 18, pp. 3743–3755, Sep. 2012, doi: 10.1128/MCB.00032-12.

[145] P. H. Edelstein, B. Hu, T. Shinzato, M. A. C. Edelstein, W. Xu, and M. J. Bessman, "*Legionella pneumophila* NudA Is a Nudix Hydrolase and Virulence Factor," *Infect Immun*, vol. 73, no. 10, pp. 6567–6576, Oct. 2005, doi: 10.1128/IAI.73.10.6567-6576.2005.

[146] E. Dimitrova, A. H. Turberfield, and R. J. Klose, "Histone demethylases in chromatin biology and beyond," *EMBO Rep*, vol. 16, no. 12, pp. 1620–1639, Dec. 2015, doi: 10.15252/embr.201541113.

[147] P. K. Rajan *et al.*, "The Role of Histone Acetylation-/Methylation-Mediated Apoptotic Gene Regulation in Hepatocellular Carcinoma," *Int J Mol Sci*, vol. 21, no. 23, p. 8894, Nov. 2020, doi: 10.3390/ijms21238894.

[148] G. Yu, L.-G. Wang, Y. Han, and Q.-Y. He, "clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters," *OMICS*, vol. 16, no. 5, pp. 284–287, May 2012, doi: 10.1089/omi.2011.0118.

[149] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger, "ColabFold: making protein folding accessible to all," *Nat Methods*, vol. 19, no. 6, pp. 679–682, Jun. 2022, doi: 10.1038/s41592-022-01488-1.

[150] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger, "ColabFold: making protein folding accessible to all," *Nature Methods 2022 19:6*, vol. 19, no. 6, pp. 679–682, May 2022, doi: 10.1038/s41592-022-01488-1.

[151] A. Waterhouse *et al.*, "SWISS-MODEL: homology modelling of protein structures and complexes," *Nucleic Acids Res*, vol. 46, no. W1, pp. W296–W303, Jul. 2018, doi: 10.1093/nar/gky427.

[152] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg, "The Phyre2 web portal for protein modeling, prediction and analysis," *Nat Protoc*, vol. 10, no. 6, pp. 845–858, Jun. 2015, doi: 10.1038/nprot.2015.053.

[153] C. Zhang, P. L. Freddolino, and Y. Zhang, "COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information," *Nucleic Acids Res*, vol. 45, no. W1, pp. W291–W299, Jul. 2017, doi: 10.1093/nar/gkx366.

[154] J. Konc and D. Janežič, "Protein−Protein Binding-Sites Prediction by Protein Surface Structure Conservation," *J Chem Inf Model*, vol. 47, no. 3, pp. 940–944, May 2007, doi: 10.1021/ci6005257.

[155] "ENSEMBL ID to Gene Symbol Converter - Genomics Biotools." https://www.biotools.fr/human/ensembl_symbol_converter (accessed May 17, 2023).

[156] B. Wang *et al.*, "Phosphorylation and acetylation of histone H3 and autoregulation by early growth response 1 mediate interleukin 1β induction of early growth response 1 transcription," *Arterioscler Thromb Vasc Biol*, vol. 30, no. 3, pp. 536–545, Mar. 2010, doi: 10.1161/ATVBAHA.109.193821.

[157] M. Bernhofer *et al.*, "PredictProtein - Predicting Protein Structure and Function for 29 Years," *Nucleic Acids Res*, vol. 49, no. W1, pp. W535–W540, Jul. 2021, doi: 10.1093/NAR/GKAB354.

[158] S. Montgomerie, J. A. Cruz, S. Shrivastava, D. Arndt, M. Berjanskii, and D. S. Wishart, "PROTEUS2: a web server for comprehensive protein structure prediction and structure-based annotation," *Nucleic Acids Res*, vol. 36, no. suppl_2, pp. W202–W209, Jul. 2008, doi: 10.1093/NAR/GKN255.

[159] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger, "ColabFold: making protein folding accessible to all," *Nat Methods*, vol. 19, no. 6, pp. 679–682, Jun. 2022, doi: 10.1038/s41592-022-01488-1.

[160] A. Bateman *et al.*, "UniProt: the universal protein knowledgebase," *Nucleic Acids Res*, vol. 45, no. D1, pp. D158–D169, Jan. 2017, doi: 10.1093/NAR/GKW1099.

[161] J. D. Thompson, D. G. Higgins, and T. J. Gibson, "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice," *Nucleic Acids Res*, vol. 22, no. 22, pp. 4673–4680, 1994, doi: 10.1093/nar/22.22.4673.

[162] V. Swaminathan, A. H. Kishore, K. K. Febitha, and T. K. Kundu, "Human Histone Chaperone Nucleophosmin Enhances Acetylation-Dependent Chromatin Transcription," *Mol Cell Biol*, vol. 25, no. 17, p. 7534, Sep. 2005, doi: 10.1128/MCB.25.17.7534-7545.2005.

[163] H. M. Berman, "The Protein Data Bank," *Nucleic Acids Res*, vol. 28, no. 1, pp. 235–242, Jan. 2000, doi: 10.1093/nar/28.1.235.

[164] C. Christoffer, V. Bharadwaj, R. Luu, and D. Kihara, "LZerD Protein-Protein Docking Webserver Enhanced With de novo Structure Prediction," *Front Mol Biosci*, vol. 8, Aug. 2021, doi: 10.3389/fmolb.2021.724947.

[165] H. Zhou and J. Skolnick, "GOAP: a generalized orientation-dependent, all-atom statistical potential for protein structure prediction," *Biophys J*, vol. 101, no. 8, pp. 2043–2052, Oct. 2011, doi: 10.1016/J.BPJ.2011.09.012.

[166] H. Zhou and Y. Zhou, "Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction," *Protein Sci*, vol. 11, no. 11, pp. 2714–2726, Apr. 2002, doi: 10.1110/PS.0217002.

[167] Y. Liu, M. Grimm, W. Dai, M. Hou, Z.-X. Xiao, and Y. Cao, “CB-Dock: a web server for cavity detection-guided protein–ligand blind docking,” *Acta Pharmacol Sin*, vol. 41, no. 1, pp. 138–144, Jan. 2020, doi: 10.1038/s41401-019-0228-6.

[168] M. Leal *et al.*, “Environmental cultures and hospital-acquired Legionnaire disease: a 5-year prospective study in 20 hospitals in Catalonia, Spain.,” *Infect Control Hosp Epidemiol*, vol. 25, no. 12, pp. 1072–1076, Dec. 2004, doi: 10.1017/S019594170007750X.

[169] S. Banga *et al.*, “Legionella pneumophila inhibits macrophage apoptosis by targeting pro-death members of the Bcl2 protein family,” *Proc Natl Acad Sci U S A*, vol. 104, no. 12, pp. 5121–5126, Mar. 2007, doi: 10.1073/PNAS.0611030104/SUPPL_FILE/11030TABLE3.PDF.

[170] “Legionnaire disease — Description of an Epidemic of Pneumonia | NEJM.” https://www.nejm.org/doi/10.1056/NEJM197712012972201 (accessed May 30, 2023).

[171] J. L. Yang, D. Li, and X. Y. Zhan, “Concept about the Virulence Factor of Legionella,” *Microorganisms 2023, Vol. 11, Page 74*, vol. 11, no. 1, p. 74, Dec. 2022, doi: 10.3390/MICROORGANISMS11010074.

[172] A. M. Joseph and S. R. Shames, “Affecting the Effectors: Regulation of Legionella pneumophila Effector Function by Metaeffectors,” *Pathogens*, vol. 10, no. 2, pp. 1–7, Feb. 2021, doi: 10.3390/PATHOGENS10020108.

[173] R. R. Isberg, T. J. O’Connor, and M. Heidtman, “The Legionella pneumophila replication vacuole: making a cozy niche inside host cells,” *Nat Rev Microbiol*, vol. 7, no. 1, p. 13, 2009, doi: 10.1038/NRMICRO1967.

[174] L. Liu and C. R. Roy, “The Legionella pneumophila Effector RavY Contributes to a Replication-Permissive Vacuolar Environment during Infection,” *Infect Immun*, vol. 89, no. 12, Nov. 2021, doi: 10.1128/IAI.00261-21.

[175] C. E. McCullough and R. Marmorstein, “Molecular basis for histone acetyltransferase regulation by binding partners, associated domains, and autoacetylation,” *ACS Chem Biol*, vol. 11, no. 3, p. 632, Mar. 2016, doi: 10.1021/ACSCHEMBIO.5B00841.

[176] H. Wapenaar and F. J. Dekker, “Histone acetyltransferases: challenges in targeting bi-substrate enzymes,” *Clin Epigenetics*, vol. 8, no. 1, pp. 1–11, May 2016, doi: 10.1186/S13148-016-0225-2/TABLES/3.

[177] M. E. Lalonde, X. Cheng, and J. Côté, “Histone target selection within chromatin: an exemplary case of teamwork,” *Genes Dev*, vol. 28, no. 10, pp. 1029–1041, May 2014, doi: 10.1101/GAD.236331.113.

[178] P. Zheng, *et al.* “Early Growth Response 1 Deficiency Protects the Host against Pseudomonas Aeruginosa Lung Infection.” *Infection and Immunity*, vol. 88, no. 1, 2019, https://doi.org/10.1128/iai.00678-19.

[179] Y. Han *et al.*, “Decreased DHRS2 expression is associated with HDACi resistance and poor prognosis in ovarian cancer.,” *Epigenetics*, vol. 15, no. 1–2, pp. 122–133, Jan. 2020, doi: 10.1080/15592294.2019.1656155.

[180] N. Shafqat *et al.*, “Hep27, a member of the short-chain dehydrogenase/reductase family, is an NADPH-dependent dicarbonyl reductase expressed in vascular endothelial tissue,” *Cell Mol Life Sci*, vol. 63, no. 10, pp. 1205–1213, May 2006, doi: 10.1007/S00018-006-6013-Y.

[181] C. Deisenroth, A. R. Thorner, T. Enomoto, C. M. Perou, and Y. Zhang, “Mitochondrial Hep27 is a c-Myb target gene that inhibits Mdm2 and stabilizes p53,” *Mol Cell Biol*, vol. 30, no. 16, pp. 3981–3993, Aug. 2010, doi: 10.1128/MCB.01284-09.

[182] Z. Sun *et al.*, "EGR1 recruits TET1 to shape the brain methylome during development and upon neuronal activity," *Nat Commun*, vol. 10, no. 1, p. 3892, Aug. 2019, doi: 10.1038/s41467-019-11905-3.

[183] J. C. Tsai, L. Liu, B. C. Cooley, M. R. DiChiara, J. N. Topper, and W. C. Aird, "The egr-1 promoter contains information for constitutive and inducible expression in transgenic mice," *The FASEB Journal*, vol. 14, no. 13, pp. 1870–1872, Oct. 2000, doi: 10.1096/FJ.99-1072FJE.

[184] Á. L. Riffo-Campos *et al.*, "Nucleosome-specific, Time-dependent Changes in Histone Modifications during Activation of the Early Growth Response 1 (Egr1) Gene," *Journal of Biological Chemistry*, vol. 290, no. 1, pp. 197–208, Jan. 2015, doi: 10.1074/JBC.M114.579292.

[185] F. Duclot and M. Kabbaj, "The role of early growth response 1 (EGR1) in brain plasticity and neuropsychiatric disorders," *Front Behav Neurosci*, vol. 11, p. 35, Mar. 2017, doi: 10.3389/FNBEH.2017.00035/BIBTEX.

[186] Z. Zhao and A. Shilatifard, "Epigenetic modifications of histones in cancer," *Genome Biology 2019 20:1*, vol. 20, no. 1, pp. 1–16, Nov. 2019, doi: 10.1186/S13059-019-1870-5.

[187] D. Schator *et al.*, "Legionella para-effectors target chromatin and promote bacterial replication," *Nature Communications 2023 14:1*, vol. 14, no. 1, pp. 1–14, Apr. 2023, doi: 10.1038/s41467-023-37885-z.

[188] G. A. Stepanov, J. A. Filippova, A. B. Komissarov, E. V. Kuligina, V. A. Richter, and D. V. Semenov, "Regulatory Role of Small Nucleolar RNAs in Human Diseases," *Biomed Res Int*, vol. 2015, pp. 1–10, 2015, doi: 10.1155/2015/206849.

[189] Human Gene Database, "BUD31 Gene - BUD31 Homolog," *Gene Cards*, 2022.

[190] I. N. Mohamed, L. Li, S. Ismael, T. Ishrat, and A. B. El-Remessy, "Thioredoxin interacting protein, a key molecular switch between oxidative stress and sterile inflammation in cellular response," *World J Diabetes*, vol. 12, no. 12, pp. 1979–1999, Dec. 2021, doi: 10.4239/wjd.v12.i12.1979.

[191] M. Cassandri *et al.*, "Zinc-finger proteins in health and disease," *Cell Death Discov*, vol. 3, no. 1, p. 17071, Nov. 2017, doi: 10.1038/cddiscovery.2017.71.

[192] M. Miyake, T. Fukui, and Y. Imai, "Differences in protein synthesis between wild type and intracellular growth-deficient strains of Legionella pneumophila in U937 and Acanthamoeba polyphaga," *Microb Pathog*, vol. 40, no. 4, pp. 161–170, Apr. 2006, doi: 10.1016/j.micpath.2005.12.005.

[193] J. W. Barnes, M. Aarnio-Peterson, J. Norris, M. Haskins, H. Flanagan-Steet, and R. Steet, "Upregulation of Sortilin, a Lysosomal Sorting Receptor, Corresponds with Reduced Bioavailability of Latent TGFβ in Mucolipidosis II Cells," *Biomolecules*, vol. 10, no. 5, p. 670, Apr. 2020, doi: 10.3390/biom10050670.

[194] M. J. Gonzales, J. M. Dugan, and R. W. Shafer, "Synonymous–non-synonymous mutation rates between sequences containing ambiguous nucleotides (Syn-SCAN)," *Bioinformatics*, vol. 18, no. 6, pp. 886–887, Jun. 2002, doi: 10.1093/bioinformatics/18.6.886.

[195] J. Yang, Y. Wang, and Y. Zhang, "ResQ: An Approach to Unified Estimation of B -Factor and Residue-Specific Error in Protein Structure Prediction," *J Mol Biol*, vol. 428, no. 4, pp. 693–701, Feb. 2016, doi: 10.1016/j.jmb.2015.09.024.

[196] M. C. Deller, L. Kong, and B. Rupp, "Protein stability: a crystallographer's perspective," *Acta Crystallogr F Struct Biol Commun*, vol. 72, no. 2, pp. 72–95, Feb. 2016, doi: 10.1107/S2053230X15024619.

[197] D. Simm, K. Hatje, and M. Kollmar, “Waggawagga: comparative visualization of coiled-coil predictions and detection of stable single α-helices (SAH domains),” *Bioinformatics*, vol. 31, no. 5, pp. 767–769, Mar. 2015, doi: 10.1093/bioinformatics/btu700.

[198] J. Konc, T. Česnik, J. T. Konc, M. Penca, and D. Janežič, “ProBiS-Database: Precalculated Binding Site Similarities and Local Pairwise Alignments of PDB Structures,” *J Chem Inf Model*, vol. 52, no. 2, p. 604, Feb. 2012, doi: 10.1021/CI2005687.

[199] R. M. Waterhouse, E. M. Zdobnov, F. Tegenfeldt, J. Li, and E. V. Kriventseva, “OrthoDB: the hierarchical catalog of eukaryotic orthologs in 2011,” *Nucleic Acids Res*, vol. 39, no. Database issue, p. D283, Jan. 2011, doi: 10.1093/NAR/GKQ930.

[200] Y. Yan, D. Zhang, P. Zhou, B. Li, and S.-Y. Huang, “HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy,” *Nucleic Acids Res*, vol. 45, no. W1, pp. W365–W373, Jul. 2017, doi: 10.1093/nar/gkx407.

[201] I. Kufareva and R. Abagyan, “Methods of protein structure comparison,” *Methods Mol Biol*, vol. 857, p. 231, 2012, doi: 10.1007/978-1-61779-588-6_10.

[202] K. Baid *et al.*, “Direct binding and internalization of diverse extracellular nucleic acid species through the collagenous domain of class A scavenger receptors,” *Immunol Cell Biol*, vol. 96, no. 9, pp. 922–934, Oct. 2018, doi: 10.1111/IMCB.12052.

[203] D. Yao, M. Cherney, and M. Cygler, “Structure of the N-terminal domain of the effector protein LegC3 from *Legionella pneumophila*,” *Acta Crystallogr D Biol Crystallogr*, vol. 70, no. 2, pp. 436–441, Feb. 2014, doi: 10.1107/S139900471302991X.

[204] K. Miyake *et al.*, “Actin Cytoskeletal Reorganization Function of JRAB/MICAL-L2 Is Fine-tuned by Intramolecular Interaction between First LIM Zinc Finger and C-terminal Coiled-coil Domains,” *Sci Rep*, vol. 9, no. 1, p. 12794, Dec. 2019, doi: 10.1038/s41598-019-49232-8.

[205] J. Schultz, F. Milpetz, P. Bork, and C. P. Ponting, "SMART, a simple modular architecture research tool: Identification of signaling domains," *Proc Natl Acad Sci U S A*, vol. 95, no. 11, pp. 5857–5864, May 1998, doi: 10.1073/PNAS.95.11.5857/ASSET/7CA4C22C-EFFB-4036-BB0D-263ECE5D1AB3/ASSETS/GRAPHIC/PQ0980595003.JPEG.

[206] L. Gomez-Valero *et al.*, "Extensive recombination events and horizontal gene transfer shaped the Legionella pneumophila genomes," *BMC Genomics*, vol. 12, no. 1, p. 536, Dec. 2011, doi: 10.1186/1471-2164-12-536.

[207] L. A. Quilliam *et al.*, "Involvement of the Switch 2 Domain of Ras in Its Interaction with Guanine Nucleotide Exchange Factors," *Journal of Biological Chemistry*, vol. 271, no. 19, pp. 11076–11082, May 1996, doi: 10.1074/jbc.271.19.11076.

[208] P. W. Chrystal *et al.*, "The inner junction protein CFAP20 functions in motile and non-motile cilia and is critical for vision," *Nat Commun*, vol. 13, no. 1, p. 6595, Nov. 2022, doi: 10.1038/s41467-022-33820-w.

[209] Y. Zhou *et al.*, "DHRS2 inhibits cell growth and motility in esophageal squamous cell carcinoma," *Oncogene*, vol. 37, no. 8, pp. 1086–1094, Feb. 2018, doi: 10.1038/onc.2017.383.

[210] K. M. Jannie *et al.*, "Vinculin-dependent actin bundling regulates cell migration and traction forces," *Biochemical Journal*, vol. 465, no. 3, pp. 383–393, Feb. 2015, doi: 10.1042/BJ20140872.

[211] A. Koenecke *et al.*, "Alpha-1 adrenergic receptor antagonists to prevent hyperinflammation and death from lower respiratory tract infection," *Elife*, vol. 10, Jun. 2021, doi: 10.7554/eLife.61700.

[212] K. M. Draheim, H.-B. Chen, Q. Tao, N. Moore, M. Roche, and S. Lyle, "ARRDC3 suppresses breast cancer progression by negatively regulating integrin β4," *Oncogene*, vol. 29, no. 36, pp. 5032–5047, Sep. 2010, doi: 10.1038/onc.2010.250.

[213] A. F. O'Donnell, L. Huang, J. Thorner, and M. S. Cyert, "A Calcineurin-dependent Switch Controls the Trafficking Function of α-Arrestin Aly1/Art6," *Journal of Biological Chemistry*, vol. 288, no. 33, pp. 24063–24080, Aug. 2013, doi: 10.1074/jbc.M113.478511.

[214] Y. Liu *et al.*, "Arrestin domain containing 3 promotes Helicobacter pylori–associated gastritis by regulating protease-activated receptor 1," *JCI Insight*, vol. 5, no. 15, Aug. 2020, doi: 10.1172/jci.insight.135849.

[215] S. Kwon and J. L. Christian, "Sortilin Associates with Transforming Growth Factor-β Family Proteins to Enhance Lysosome-mediated Degradation," *Journal of Biological Chemistry*, vol. 286, no. 24, pp. 21876–21885, Jun. 2011, doi: 10.1074/jbc.M111.228262.

[216] M. Canuel, A. Korkidakis, K. Konnyu, and C. R. Morales, "Sortilin mediates the lysosomal targeting of cathepsins D and H," *Biochem Biophys Res Commun*, vol. 373, no. 2, pp. 292–297, Aug. 2008, doi: 10.1016/j.bbrc.2008.06.021.

[217] K. Bärlocher, A. Welin, and H. Hilbi, "Formation of the Legionella Replicative Compartment at the Crossroads of Retrograde Trafficking," *Front Cell Infect Microbiol*, vol. 7, Nov. 2017, doi: 10.3389/fcimb.2017.00482.

[218] S. Itoh, K. Mizuno, M. Aikawa, and E. Aikawa, "Dimerization of sortilin regulates its trafficking to extracellular vesicles," *Journal of Biological Chemistry*, vol. 293, no. 12, pp. 4532–4544, Mar. 2018, doi: 10.1074/JBC.RA117.000732.

[219] D. J. MacPhee, "Methodological considerations for improving Western blot analysis," *J Pharmacol Toxicol Methods*, vol. 61, no. 2, pp. 171–177, Mar. 2010, doi: 10.1016/j.vascn.2009.12.001.

[220] “Thermo Scientific Pierce BCA Protein Assay Kit: Protein Analysis Reagents:Protein | Fisher Scientific.” https://www.fishersci.com/shop/products/pierce-bca-protein-assay/p-4531640 (accessed Jun. 13, 2023).

APPENDICES

Appendix 01: Expression of His-tagged LneB in *Escherichia coli*, BL21.

a. A colony from fresh transformed *E. coli* on Laura agar plate was inoculated in 30mL Laura broth containing 30µL ampicillin and incubated overnight on a shaker at 37^0C.

b. 10mL of the overnight grown culture was transferred into a flask containing 500 mL L.B. (ration 1:50) containing 500µL of ampicillin.

c. The culture was incubated on a shaker at 37^0C for 3-4 hours until the bacteria reach the exponential phase (OD_{600}=0.5-0.6).

d. 1mL of the bacteria culture was transferred to a 1.5mL tube before the addition of IPTG, spun at 10,000 for 1 min to form a pellet and kept at -20^0C for SDS-PAGE. This served as the uninduced sample.

e. The remaining 499mL of the bacteria culture was induced using IPTG at a final concentration of 200µM.

f. Induced culture was incubated on a shaker at 18^{o}C overnight for slow induction.

g. 1mL tubes containing the induced cells were centrifuged at 10,000 rpm for 1 minute for SDS-PAGE.

h. The induced cells were collected into a 250 ml centrifuge tube and centrifuged at the G rotor at 3000 rpm for 10 min at 4^0C. The pellets were saved and stored at -80^{o}C for the protein purification step.

Appendix 02: Purification of His-LneB Protein Using Ni-ion Column Chromatography

Buffer Preparation

A. Lysis Buffer

2 mM sodium azide

500 mM NaCl

5 mM imidazole

20 mM Tris-Cl, pH 7.9

B. High-Salt Wash Buffer

2 mM sodium azide

2 M NaCl

5 mM imidazole

20 mM Tris-Cl, pH 7.9

C. His Elution Buffer

2 mM sodium azide

500 mM NaCl

500 mM imidazole

2M Tris-Cl, pH 7.9

Steps

a. Approximately 1 gram of pellet was resuspended into 5mL of reconstituted lysis buffer and vortexed intermittently without generating heat.

   - Reconstituted lysis buffer = 10mL lysis buffer + 250µL (50mg/mL) lysozyme (Thermo Scientific™ 9008) + 1 tablet of protease inhibitor (Pierce™ Protease Inhibitor Tablets A32963).

b. Bacteria cells was further lysed 6mm probe using a sonicator on ice with 10s on and 20s rest (pulse-off) at 50% amplitude and repeated six times. 50µL of the lysate was taken after sonication into a 1.5mL tube for SDS-PAGE analysis. The remaining lysate was transferred into an appropriate microcentrifuge tube.

c. The resulting lysate was centrifuged for 40 minutes at 15,000g at 4^0C, and the supernatant was recovered.

d. The resin beads stock containing Ni-ions was mixed by rocking at room temperature for 10 minutes before use, and 0.8mL of resin was transferred into a 1.5mL tube.

e. The beads were spun at 3000g for 1 min and 0.8 mL of lysis buffer was used to wash the resin twice.

f. Resin was transferred to the purification column at 4^0C. Resin was washed with 1mL of the lysis buffer in the column.

g. Syringe filter of 0.45µm pores was used to filter the supernatant into a sterile 50mL falcon tube.

h. Filtrate was passed through the column bed containing Ni-ion resin and left to set to aid in order to aid the interaction of the protein with the resin for 30 minutes.

i. The column was covered with parafilm and set on the rotating device for ten minutes to aid the interaction of resin and the protein. After mixing, beads were allowed to settle and 50µL of the sample was collected into a 1.5mL tube for SDS-PAGE.

j. 10mL of wash buffer was added to the column, which was later sealed with parafilm.

k. The column was transferred to a rotating device for 10 minutes, the parafilm was removed and column was opened to collect 50µL of the filtrate as wash-1 into a labeled 1.5mL tube for SDS-PAGE. This was repeated thrice to generate wash 2 and wash 3.

l. HIS-tagged protein was eluted in triplicate by 1mL of elution buffer, incubated on rotation at 4^0C. Elution was repeated two more times and the eluent containing HIS-tagged protein was collected into a labeled 1.5mL tube.

Appendix 03: SDS-PAGE

A. SDS-Gel Preparation

a. 12% of resolving SDS gel was made using the components in Table 9.

b. The resolving gel was first cast into 0.75mm glass-casting tray, followed by the stacking gel and 10wells were created using the 0.75mm thick comb.

Table 9: SDS-gel Preparation

Component	Volume -Resolving gel (mL)	Volume -Stacking gel (mL)
H_2O	1.6	3.4
30% acrylamide mis	2.0	0.83
1.5M Tris (pH 8.8)	1.3	0.63
10% SDS	0.05	0.05
10% ammonium persulfate	0.05	0.05
TEMED	0.002	0.005

B. Sample preparation

a. The volume of cell pellet was weighed, and equal volume of nano-pure water was added to make a suspension followed by dilution with 2X SDS Loading dye into 1X.

b. Sample was boiled at 100^{o}C for 5 minutes and centrifuged at full speed (11000g) for 3 mins.

c. 5µL of protein molecular weight marker (Thermo Scientific™ 26612) and 10µL of treated samples were loaded into designated wells and electrophoretic separation was carried out at 100V for 15 minutes to drive the protein along the stacking gel and 130V for 60mins for driving the protein along the stacking gel until the dye reached the bottom of the gel.

d. Gel was stained with Coomassie brilliant blue for 4 hours at room temperature on a rocker and distained afterward with a distaining solution (made of glacial acetic acid and methanol).

e. Gel was viewed with white background under white light.

Appendix 04: Western blot

Buffer Preparation

A. Transfer buffer

1. 800mL of nanopure water containing 5.8g Tris Base, 2.9g Glycine, and 0.37g SDS.
2. 200mL MeOH.

B. Blocking solution

    1. 2% powdered non-fat milk solution was made by dissolving 2g of the powdered non-fat milk in 20mL of 1X PBS.
    2. PBST containing 500µL Tween20 in 1X PBS was added to the milk solution to make up to 100mL.

Wash buffer

100mL1X PBS

899mL ddH2O

1mL Tween20

Steps

a. Nitrocellulose membrane of 0.45µm pore size was pre-wetted in transfer buffer for 5 minutes.
b. Transfer sandwich was assembled as shown below by making the gel on the negative side and membrane on the red positive side.

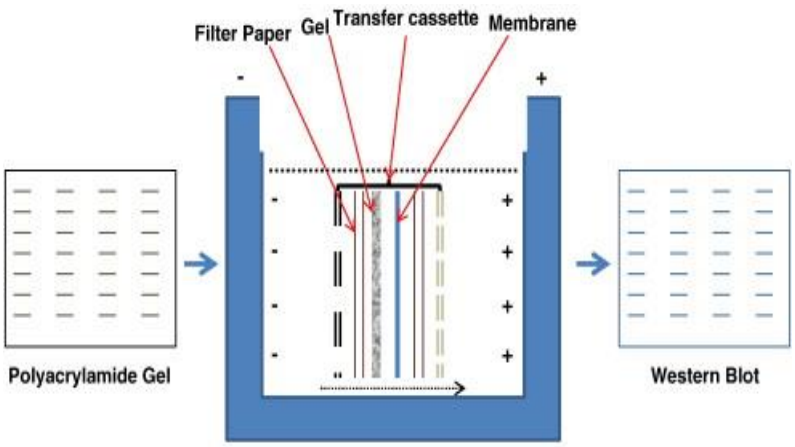


Figure 43 Arrangement of blotting accessories for western blot [219]

c. Protein was transferred to the membrane for 1hr at 100volts on ice to reduce the heat and preserve protein.
d. After transfer, the membrane was incubated in 10 – 20mL blocking buffer for 45mins – 1hr at room temperature with rocking.
e. Anti-HIS-tag primary antibody was diluted with blocking buffer at a ratio of 1: 5000.
f. The membrane was incubated in 10mL of diluted 1^0 Antibody for 1hr at room temperature on a rocker.
g. The blot was washed with 5mL wash buffer five times. a. Brief wash B. 4 X 5min wash.
h. Anti-rabbit horse radish peroxidase 2^0Ab was diluted in blocking buffer at 1:20000.
i. The blot was incubated in diluted 2^0Ab for 1hr at room temperature on rocking.
j. The blot was washed with wash buffer for 5mins on rocking.
k. Components 1 and 2 of Enhanced Chemiluminescence (ECL) substrate was mixed at 1:1 ratio.
l. The blot was incubated in ECL substrate for 2mins at room temperature.
m. Blot was view under translumiscent as high/medium intensity and 1minute exposure.

Appendix 05: Protein Concentration Estimation – BCA assay

As shown in Figure 44 below, spectrophotometric analysis, Bicinchoninic acid, BCA, Thermo Scientific™ 23225 was used for protein concentration estimation.

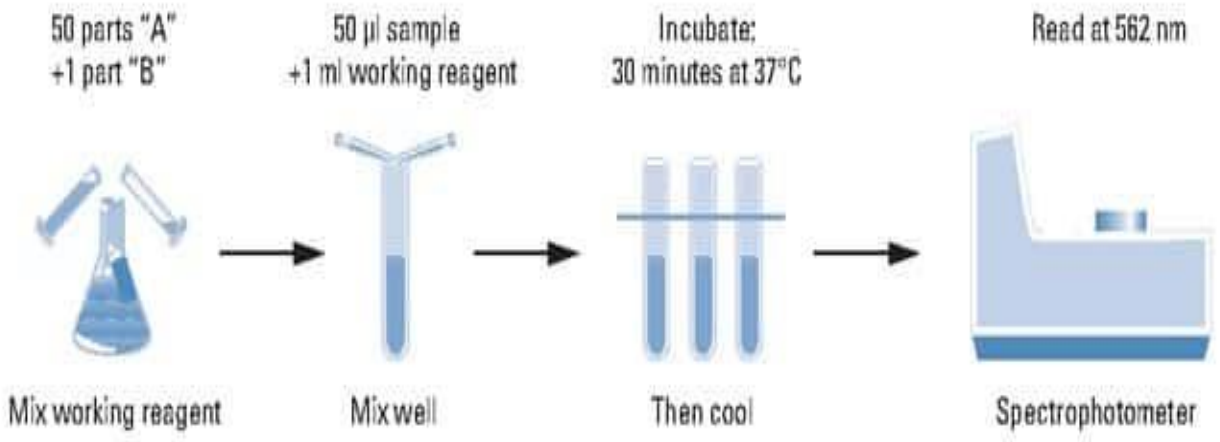


Figure 44: Schematic diagram of BSA assay [220]

Appendix 06: Transfection of 293T Cell

Treating Cells with DNA

a. 500µL of Opti-mem, 20µg of DNA and 40µL of p3000 was mixed into 1.5mL tube, vortexed and incubated at room temperature for 5minutes.

b. 500µL of Opti-mem and 40µL of lipofectamine 3000 (Invitrogen) (ref# L3000-015) was mixed into 1.5mL tube.

c. Components of tubes A and B in step a and b were mixed into tube C without vortex to avoid generating bubbles.

d. The tube was incubated at room temperature for 15 minutes.

e. HEK 293T (Human Embryonic Kidney) cells were grown in tissue culture dishes (100 mm x 20 mm) in Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal bovine serum (batch # 70032512).

f. The cells were incubated with 5% CO2 at 37°C. The cells were transfected using 5 x 106 cells/mL cell density on a 100 x 20 mm culture dish.

g. Cells that were 70–90% confluent were transfected.

h. The DNA mix in tube C was added to cells. Cells were incubated for 5hrs at 37^0C under CO_2. Afterwards, 5mL of media with 10% FBS was added to the cells.

i. The medium was removed after transfection and cells were detached from the flask using cold 4mL PBS and transferred into 15mL tube. The second cell recovery was done with 4mL to remove remnant cells.

j. Cell suspension was spun at 1000g for 5mins, and supernatant was discarded after spinning.

k. The cell pellet was suspended in 1mL PBS, transferred into 1.5mL tube and centrifuged spun at 5000g for 1minutes at 4^0C.

l. To investigate HAT activity in the nuclear extracts of the transfected cells, cell pellets were recovered from step k above.

Appendix 07: Nuclear Extraction Protocol

a. Protease inhibitor was added to CER1 and NER at 1:100 ratio, with volume depending on the packed cell volume of the pellets.

b. Ice-cold CER 1 was added to the cell pellet (Table 10) and vortexed vigorously at the highest setting for 15 minutes.

Packed Cell Volume (µL)	CER I (µL)	CER II (µL)	NER (µL)
10	100	5.5	50
20	200	11	100
50	500	27.5	250
100	1000	55	500

*For HeLa cells, 2×10^6 cells is equivalent to 20µL packed cell volume.

Table 10: Reconstituting nuclear and cytoplasmic extraction reagent based on packed cell volume. Maintain the volume ratio: CER I: CER II: NER at 200:11:100µL, respectively.

c. The tube was incubated on ice for 10 minutes and ice-cold CER II was added to the tube.

d. The tube was vortex for 5 seconds at the highest setting and was further incubated on ice for 1 minute and 5 seconds at the highest setting.

e. The tube was spun for 5 minutes at maximum speed in a microcentrifuge at 4°C (~16,000 × g).

f. Immediately, the supernatant containing the cytoplasmic extract was transferred to a clean pre-chilled tube which was stored at -80°C until use.

g. The insoluble pellet fraction was resuspended in ice-cold NER and vortexed at the highest setting for 15 seconds.

h. The sample was placed on ice and vortexed every 10 minutes for 15 seconds for 40 minutes.

i. The tube was spun for 10 minutes at maximum speed in a microcentrifuge (~16,000 × g) at 4°C.

j. Immediately, the supernatant containing the nuclear extract was transferred to a clean pre-chilled tube which was stored at -80°C until use.

Appendix 08: Histone Acetylation Protocol (EpiQuik™ HAT Activity/Inhibition Assay Kit Base Catalog # P-4003).

a. HT1 (wash buffer) was diluted to a 1:10 ratio at pH 7.2 to 7.5.
b. HT2 (histone substrate) was diluted at 1:50 ratio with diluted HT1.
c. 50 µl of the diluted HT2 was transferred into wells of the samples of a 96-well plate and blanks except for the wells for the standard curve.
d. Standards were made by diluting HT3 with diluted HT1 from 20µg/mL stock to obtain the mass of 0.5, 1.0, 2.0, 3.0, 4.0, 6.0, and 8.0ng.
e. 50µl of diluted HT1 was added into the wells of the standards only, followed by 1 µl of the prepared standard concentration range of HT3 (use 0.1 – 10 ng).
f. Cover the wells with Parafilm and incubate at room temperature for 30-45 minutes.
g. Aspirate and wash each well with 150 µl of diluted HT1 three times.
h. HT4 (Acetyl-CoA) at a 1:20 ratio with HT5. The volume that was prepared depended on the number of samples analyzed.
i. For the sample wells, add 26 µl of HT5, 2 µl of the diluted HT4, and 2 µl of nuclear extract samples (4-20 µg), making a total of 30µL.
j. For the standard curve wells, 28 µl of HT5 and 2µL of diluted HT4 was added to the well, making a total of 30µL and 30µl of HT5 into the blank wells.
k. The strips were covered well, plate was gently shaken, and incubate at 37°C for 30-60 minutes.
l. Each well was aspirated and washed with 150 µl of diluted HT1 three times.
m. HT6 (capture antibody) was diluted 1:100 to 1 µg/ml with diluted HT1.

n. 50 µl of the diluted HT6 was added to each strip well and incubated at room temperature for 60 minutes on an orbital shaker (50-100 rpm).

o. Each well was aspirated and washed with 150 µl of diluted HT1 three times.

p. HT7 (detection antibody) was diluted 1:1000 to 0.2 µg/ml with diluted HT1.

q. 50 µl of the diluted HT7 was added to each strip well and incubate at room temperature for 25-30 minutes.

r. Each well was aspirated and washed with 150 µl of diluted HT1 five times.

s. 100 µl of HT8 (developing solution) was added to each well and incubate at room temperature for 2-10 minutes away from light.

t. 50 µl of HT9 (stop solution) was added to each well to stop enzyme reaction.

u. Absorbance was read on a microplate reader at 450nm within 2-15 minutes and HAT activity or inhibition was calculated using instructions in the protocol.

Copyright Permission

Name: Adeyemi, Kayode Gabriel

Email (to receive future readership statistics): kayode.adeyemi891@topper.wku.edu

Type of document: ['Thesis']

Title: CHARACTERIZATION OF Legionella pneumophila EFFECTOR PROTEINS, LneB and MavA.

Keywords (3-5 keywords not included in the title that uniquely describe content): Histone acetyltransferase, RasGEF, Bioinformatics prediction, Transcriptomics.

Committee Chair: Dr. Simran Banga

Additional Committee Members: Dr. Rodney King.
Dr. Ajay Srivastava.

Select 3-5 TopSCHOLAR® disciplines for indexing your research topic in TopSCHOLAR®: Life Sciences

Copyright Permission for TopSCHOLAR® (digitalcommons.wku.edu) and ProQuest research repositories:

I hereby warrant that I am the sole copyright owner of the original work.
I also represent that I have obtained permission from third party copyright owners of any material incorporated in part or in whole in the above described material, and I have, as such identified and acknowledged such third-part owned materials clearly. I hereby grant Western Kentucky University the permission to copy, display, perform, distribute for preservation or archiving in any form necessary, this work in TopSCHOLAR® and ProQuest digital repository for worldwide unrestricted access in perpetuity.
I hereby affirm that this submission is in compliance with Western Kentucky University policies and the U.S. copyright laws and that the material does not contain any libelous matter, nor does it violate third-party privacy. I also understand that the University retains the right to remove or deny the right to deposit materials in TopSCHOLAR® and/or ProQuest digital repository.

['I grant permission to post my document in TopSCHOLAR and ProQuest for unrestricted access.']

The person whose information is entered above grants their consent to the collection and use of their information consistent with the Privacy Policy. They acknowledge that the use of this service is subject to the Terms and Conditions.

['I consent to the above statement.']