5-2018

Divergent evolution of di-lysine ER retention vs. farnesylation motif-mediated anchoring of the ankb virulence effector.

John David Perpich
University of Louisville

DIVERGENT EVOLUTION OF DI-LYSINE ER RETENTION VS. FARNESYLATION MOTIF-MEDIATED ANCHORING OF THE ANKB VIRULENCE EFFECTOR

By

John David Perpich
B.S., Michigan State University, 1999
M.S., Michigan State University, 2001
Pharm.D., Medical University of South Carolina, 2007

A Dissertation
Submitted to the Faculty of the
School of Medicine of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in Microbiology and Immunology

Department of Microbiology and Immunology
University of Louisville
Louisville, KY

May 2018

DIVERGENT EVOLUTION OF DI-LYSINE ER RETENTION VS. FARNESYLATION MOTIF-MEDIATED ANCHORING OF THE ANKB VIRULENCE EFFECTOR

By

John David Perpich
B.S., Michigan State University, 1999
M.S., Michigan State University, 2001
Pharm.D., Medical University of South Carolina, 2007

A Dissertation Approved on

April 12, 2018

by the following Dissertation Committee:

_____
Dissertation Director
Dr. Yousef Abu Kwaik

_____
Dr. Donald Demuth

_____
Dr. Matthew Lawrenz

_____
Dr. Jonathan Warawa

_____
Dr. Michael Merchant

DEDICATION

To my wife, Amy, for her willingness to support me in this endevour and in the next phase of my career. For staying by my side during our most difficult times. To my son, Ethan, for always putting a smile on my face and teaching me to always be positive. I love them both more than words can express.

ACKNOWLEDGMENTS

I thank my mentor, Yousef Abu Kwaik, for always being available and willing to guide me and for helping me find my faults and grow beyond them. Also for his many pieces of valuable advice. I also thank my committee for their guidance. In particular, Dr. Michael Merchant for many hours of troubleshooting as well as career guidance. I am grateful to Dr. Michele Kosiewicz for all her time and invaluable advice on my career.

ABSTRACT

DIVERGENT EVOLUTION OF DI-LYSINE ER RETENTION VS. FARNESYLATION MOTIF-MEDIATED ANCHORING OF THE ANKB VIRULENCE EFFECTOR

John Perpich

April 12, 2018

Legionella pneumophila is an aquatic organism capable of intracellular replication within a wide range of protozoan hosts and within human macrophages where it causes Legionnaires' Disease. *L. pneumophila* manipulates a variety of host cell processes by translocating, into the host cell cytosol, more than 300 effector proteins via the type IVB secretion system (T4SS). The AnkB effector from the AA100/130b strain (AnkB-AA100) contains a C-terminal CaaX motif that rapidly becomes farnesylated upon translocation into the host cell and anchored to the cytosolic face of the *Legionella*-containing vacuole (LCV) membrane, which is essential for intracellular replication. We show that a homolog of AnkB from the Paris strain (AnkB-Paris) has a frameshift mutation that truncates the C-terminus eliminating the CaaX motif, but creating a unique KNKYAP sequence that resembles a eukaryotic di-lysine ER retention motif (KxKxx). AnkB-Paris localizes to the cytosolic face of the LCV membrane most likely through the ER retention motif. Phylogenetic analyses indicate that the AnkB-Paris allele is common among environmental isolates and is positively

selected specifically for the di-lysine motif. Ectopic expression of AnkB-Paris results in a perinuclear distribution and *trans*-rescues the *ankB* mutant of the AA100/130b strain for intravacuolar replication. The *trans*-rescue is dependent on an intact di-lysine ER-retention motif, which most likely enables anchoring of AnkB to the ER-derived LCV membrane. AnkB contains a N-terminal eukaryotic F-box domain that recruits the cellular ubiquitylation machinery by interacting with the host Skp1 protein resulting in decoration of the LCV with polyubiquitylated proteins. Proteins targeted for ubiquitylation during infection are likely bound by ankyrin domains in the C-terminus of AnkB. Based on the crystal structure, we identified four residues within the ankyrin domains likely involved in binding specific substrate proteins. Mutation of these residues did not affect protein folding but resulted in loss of accumulation of ubiquitylated proteins around the LCV and a severe defect in intracellular replication, similar to the *ankB* null mutant. Overall, we identified a di-lysine ER retention motif in the C-terminus of AnkB-Paris and 4 substrate binding residues within the 3 ankyrin domains of AnkB-AA100 both of which are essential for biological function.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

INTRODUCTION

1. History and Etiology of Legionnaires' Disease

Legionnaires' disease was first described in the aftermath of an outbreak of a severe pneumonia in 221 people at a convention of the American Legion in Philadelphia that ultimately claimed 34 lives [1, 2]. Many of the people who became ill had visited the Bellevue Stratford Hotel in Philadelphia, however, some affected individuals had merely walked passed the hotel [3]. Nevertheless, there was a strong association between staying at the hotel and development of the pneumonia. It is apropos that the epidemic occurred on the centennial of Koch's discovery of the bacterium responsible for anthrax, since the identification of the causative agent of Legionnaires' disease closely followed Koch's postulates. The postulates state that to establish the cause of an infectious disease, the organism must be present in diseased but not healthy hosts [4]. It must be isolated in pure culture from diseased individuals and reproduce the disease when inoculated into a healthy host [4]. Finally, it must be re-isolated in pure culture from the diseased experimental host [4]. After an extensive epidemiological investigation, a small filamentous gram-negative rod was isolated from infected patients. The causative agent was later identified and classified into the new family *Legionellaceae* as *Legionella pneumophila* [5].

L. pneumophila is a gram-negative, motile, non-spore-forming bacillus. Although the bacterium is found free-living and biofilm-associated, it primarily persists in freshwater environments by infecting and replicating within several protozoans including members of *Acanthamoeba*, *Hartmannella*, *Naegleria*, and *Tetrahymena* [6]. *L. pneumophila* has a biphasic lifecycle that alternates between a nonmotile replicative phase and a flagellated transmissive phase [7, 8].

More than 50 species and 70 serogroups in the genus have been described with almost 50% of the species causing disease in humans [9]. Among the various isolates of *L. pneumophila*, serogroup 1 accounts for 85% of cases in the USA. In Australia and New Zealand, 30% of cases of Legionnaires' disease are caused by *L. longbeachae*, which is found mostly in soil [5].

2. Epidemiology and Clinical Manifestations

The worldwide incidence of Legionnaires' disease is difficult to determine due to differences in awareness levels, diagnosis, and reporting throughout the world [10]. The data are better in countries where it is a reportable disease such as the USA, New Zealand, Australia, Japan, Canada, and Europe where the incidence ranges from 2-14 per million people [11]. *L. pneumophila* is thought to account for 2-9% of community-acquired pneumonia with most cases occurring during the warmer and wetter months of the year [10]. In the USA, the reported incidence has increased from 3.9 to 11.5 cases per million people during the period from 2000-2009 [12]. Still, the disease is thought to be largely underreported even in countries with surveillance programs. For example, a

German study of community-acquired pneumonia that utilized a standard protocol to rigorously diagnose *Legionella* pneumonia found a prevalence of 180-360 cases per million people [10].

Risk factors for Legionnaires' disease include age >50 years, male gender, smoking, chronic lung disease, diabetes, and end-stage renal disease [10, 13]. Additional risk factors include immunosuppression, hematologic malignancies, lung cancer, and treatment with chemotherapy. Persons infected with human immunodeficiency virus (HIV) may be at greater risk according to CDC surveillance data, however, clinical studies in HIV cohorts have failed to verify this [14-16].

Clinically, *L. pneumophila* causes two distinct illnesses. Pontiac fever is a self-limiting flu-like illness with symptoms of fever, headache, malaise, and myalgia [17]. This manifestation of *L. pneumophila* infection was first described in 1968 as an epidemic that occurred in a county health department building in Pontiac, Michigan [17]. At the time, the causative agent was not identified. However, after the description of Legionnaires' disease in 1976, the cause of Pontiac fever was retrospectively identified as *L. pneumophila*. The other clinical syndrome caused by *L. pneumophila* is the multisystem disease involving pneumonia known as Legionnaires' disease, which mimics other types of pneumonia with fever, cough, dyspnea, headache, myalgia, and delirium [13, 18]. The similarity to other types of pneumonia and the empiric nature of treatment for community-acquired pneumonia likely contribute to under-reporting of cases of Legionnaires' disease. However, gastrointestinal or neurological symptoms

suggest *Legionella* as the cause of the pneumonia [10]. Nevertheless, Legionnaires' disease is difficult to distinguish clinically from other pneumonias.

Diagnosis of Legionnaires' disease can be achieved using several modalities, including culture of the organism, immunofluorescence microscopy, urine antigen testing, and nucleic acid based assays. Diagnosis by culture remains the gold standard for diagnosis, but can suffer from low sensitivity due to inadequate sample collection and, potentially, poor survival of the bacteria within respiratory secretions [13]. Urine antigen testing is a commonly used test due to its ease of sample collection and fast turnaround time. This test employs monoclonal antibodies to detect lipopolysaccharide (LPS) antigens shed by the organism [9]. The diagnostic utility of this test is limited by its ability to detect only serogroup 1 and by the fact that about 8% of patients do not shed LPS in their urine [10]. In addition, the results cannot be used in epidemiologic investigations that endeavor to trace the source of the contagion. The community acquired pneumonia guidelines published by the Infectious Disease Society of America (IDSA) recommend blood and sputum cultures as well as urinary antigen testing for *L. pneumophila* and *S. pneumoniae* in patients with severe community acquired pneumonia (CAP) [19]. Ultimately, this would seem a prudent approach in most cases.

The mortality rate of Legionnaires' disease is estimated to be 10-40% with the number of fatalities likely decreasing. One study found that mortality rates dropped from 34% to 12% between 1980-1998 [20]. This may be due to newer antibiotic therapies, which are the mainstay of treatment for Legionnaires'

disease, or more aggressive treatment of community acquired pneumonia. Prompt treatment is important since delayed initiation of antibiotics may increase mortality associated with *Legionella* pneumonia [21]. The empiric antibiotic therapy for CAP recommended by the IDSA will be effective in patients infected with *Legionella*. The bacteria are susceptible to macrolides including erythromycin and azithromycin as well as fluoroquinolones such as levofloxacin [13]. *Legionella* are naturally resistant to beta-lactam antibiotics, which is an important consideration, since clinicians do not always follow IDSA guidelines particularly in an outpatient setting.

3. Microbial Ecology and Transmission

Freshwater is the major habitat for *L. pneumophila* and the bacteria have been isolated from these environments throughout the world [22]. The organism is relatively sensitive to drying, pH extremes, and temperatures much above or below the preferred 25-42°C [23]. Within these aquatic environments, *Legionella* bacteria invade into, and persist within, free living protozoan hosts, including at least 14 species of amoebae, two species of ciliated protozoa, and a slime mold [13]. The intracellular environment of these hosts provides a nutrient-rich and sheltered environment for replication [24]. In addition to their persistence in natural bodies of water, *Legionella* bacteria can also survive and replicate in man-made bodies of water.

L. pneumophila has been isolated from air conditioning systems, cooling towers, hot tubs, showers, and dental water lines [25-28]. Legionnaires' disease outbreaks are strongly associated with these man-made aquatic environments.

The increased presence of man-made water systems in the latter half of the 20th century is what led to the initial outbreak and description of Legionnaires' disease [29]. These man-made aquatic environments may provide a habitat more favorable to *Legionella* proliferation, as bacterial counts within these environments can exceed those found in rivers, lakes, and ponds. Yamamoto and colleagues found that bacterial counts within cooling tower water samples correlated positively with increases in water temperature, pH, and levels of protozoans [30]. Several methods for disinfection of these reservoirs have been utilized including physical, thermal, and chemical treatments [31]. It is generally recommended that treatment and maintenance plans be in place for these water systems, however, the are no definitive recommendations [31]. When water from these man-made reservoirs is aerosolized, it can be inhaled by people in close association and cause the disease.

4. Metabolism

Early work using chemically defined media revealed *L. pneumophila* could only grow on amino acids and did not utilize glucose or starch [32, 33]. Pyruvate and α-ketoglutarate both stimulated growth, however, addition of glucose or starch did not impact growth rate. George *et al.*, found that *L. pneumophila* is auxotrophic for arginine, cysteine, isoleucine, leucine, threonine, valine, methionine, and phenylalanine or tyrosine [32, 34-36]. The sole sources of carbon and energy were serine and, to a lesser degree, threonine. Later, it was determined that *L. pneumophila* can utilize glucose as well as glycerol for synthesis of histidine and mannose at certain times [37]. The bacteria have two

growth phases within cells, a replicative and a transmissive phase. From the perspective of gene expression, these phases correspond closely with *in vitro* grown bacteria in exponential (E) and post-exponential (PE) states, respectively [38]. It is now appreciated that *L. pneumophila* alters its basic metabolism in each of these phases [13].

During its replicative phase, *L. pneumophila* uses serine as its major carbon source for amino acid and protein synthesis and also as its major energy source [37]. Serine is converted to pyruvate, which is then converted to acetyl-CoA. The acetyl-CoA can drive energy production via the tricarboxylic acid (TCA) cycle and also serves as a substrate for synthesis of the energy storage compound poly-3-hydroxybutyrate (PHB) [38, 39]. The replicative form is nonmotile, contains little to no PHB, and is minimally infectious [13]. As nutrients within the host cell are depleted, the alarmone, 3',5'-bispyrophosphate (ppGpp), accumulates within the bacterial cell and causes an increase in the stationary-phase σ factor RpoS [40]. During stationary phase, the PHB is broken down and used as the major source of carbon and energy [41, 42]. RpoS expression causes the bacteria to transition into the transmissive phase, which is characterized by osmotic resistance, sodium sensitivity, motility, accumulation of PHB granules, and high infectivity [13, 41]. These infectivity traits include egress from the spent host cell, extracellular survival, invasion of a new host, and lysosome evasion [38]. Thus, *L. pneumophila* is an asaccharolytic organism that subsists on amino acids via the TCA cycle and oxidative phosphorylation.

5. Intracellular Life Cycle

The life cycle of *L. pneumophila* has been described in multiple host cell types ranging from protozoans, including *Acanthameoba castellanii*, *Hartmannella vermiformis*, and *Dictyostelium discoideum,* to various mammalian cells, including mouse bone marrow-derived macrophages, human mononuclear-derived macrophages (hMDM), HeLa, A549, and CHO-K1 cells [29]. While some host cell specific differences exist, most of the fundamental life cycle is the same in all of these eukaryotic cells [29]. Preliminary studies on entry indicated that the bacteria enter the host cell by coiling phagocytosis, however, they can also enter by traditional phagocytosis [43]. Unlike most bacteria whose phagosomes fuse with lysosomes, phagosomes containing *Legionella* do not become acidic or fuse with lysosomes [44]. Instead, they remodel the phagosome into a replicative niche called the *Legionella* containing vesicle (LCV) [18]. During the first hour, the LCV becomes surrounded by rough ER-derived vesicles and mitochondria [45-48]. The vacuoles acquire ER markers such as BiP, but do not accumulate endocytic or lysosomal markers such as LAMP-1, cathepsin D, and Rab5 [46, 47, 49]. Since amoebae feed on bacteria, it is likely that *L. pneumophila* evolved to become an intracellular pathogen due to selective pressure to avoid the lysosomal pathway.

6. T4SS and Manipulation of Host Cell Processes

Control over host cell processes is achieved by virtue of the type IVB secretion system (T4SS), which is a protein translocation machine designed to inject bacterial effector proteins into the host cell that modulate host cell functions

and promote formation of a replicative niche [29]. This virulence system was co-discovered by two groups and is now known as the *dot/icm* system for defect in organelle trafficking or intracellular multiplication, respectively [50, 51]. This system is essential for *L. pneumophila* survival and proliferation within phagocytic cells as mutations that impair translocation of all T4SS effector proteins render the bacteria unable to replicate [52]. In addition, the system is also involved in bacterial entry into host cells, establishing the LCV, manipulation of various cellular processes, and exit from the host cell [29, 53, 54]. Host cell processes that are affected during an infection include vesicle trafficking, protein translation, apoptosis, and ubiquitination pathways [29].

L. pneumophila encodes at least 330 effector proteins, which is nearly 5 times more than other intracellular pathogens including *Salmonella enterica*. In fact, the bacterium dedicates nearly 10% of its proteome to effectors [52]. Knocking out any single effector protein rarely results in a bacterial strain defective for intracellular replication. Functional redundancy is the widely accepted explanation for the large number of effectors and the absence of a growth defect in strains lacking any one effector [55]. Evidence for this, in part, comes from work in the Isberg laboratory [56]. A strain lacking 5 large regions of the genome including 31% of the known effectors of the T4SS was not defective for *in vitro* growth and only minimally defective for growth in mouse macrophages. However, deletion of just one of the regions caused reduced growth in amoebae. In addition, the importance of each of the 5 regions differed depending on the species of amoebae that was infected. Interestingly, deletion

of a region five, which contained *ankB,* resulted in a growth defect *in vitro*. The strain variations observed in *L. pneumophila* reflect its highly plastic genome that has evolved to adapt to multiple host cells. Many effectors contain eukaryotic protein domains and *Legionella* has evolved by acquiring genetic material through horizontal gene transfer from its host cells [57]. Therefore, the functional redundancy exists to expand *Legionella's* host range to be a generalist pathogen to ensure survival in diverse hosts in the aquatic environment.

Many effector proteins contain eukaryotic domains to facilitate manipulation of host cell processes [58]. These domains are mixed and matched in different combinations to make new effectors. One comprehensive study of effectors in 38 *Legionella* species found 608 Legionella effector ortholog groups, but only seven effectors were conserved as "core effectors" and found in all species [59]. The authors identified numerous new effector domains as well as novel domain combinations.

7. Genome Plasticity

Members of the *Legionella* genus are capable of infecting protozoans in many freshwater environments throughout the world and they have a very broad host range [29]. Protozoans that serve as hosts are diverse and include amoeba, slime molds, and ciliates. In response to the need to parasitize many different protozoa, the genome of *Legionella*, and *L. pneumophila* in particular, is highly plastic [60].

An analysis of the genomes of 6 *L. pneumophila* strains (Paris, Lens, Philadelphia, Corby, Alcoy, and AA100/130b) revealed that, although the GC

content was consistent at 38% for all strains, there was significant heterogeneity among them [60]. The total size of the genomes ranged from 3.345 mega bases (Mb) in Lens to 3.576 Mb in Corby, while the number of genes ranged from 2980 in Lens to 3288 in AA100/130b [60]. The number of genes specific to a given strain varied from 144 in Corby to 386 in AA100/130b, which translates to between 6 and 11% of the genes as strain specific [60]. Intriguingly, a large percentage of genes are dispensable for growth of the Philadelphia strain in bacterial culture. Deleting 27% of all protein encoding genes from various places in the chromosome did not result in a growth defect *in vitro* [56]. Together, these regions represent 49% of the known Dot/Icm translocated effectors [56]. In addition, they had a lower GC content than the rest of the genome suggesting acquisition by horizontal gene transfer [56]. These data indicate that the genome of *L. pneumophila* is highly adaptable and that genome plasticity, particularly of the effectors, has played a major role in shaping the different strains [60].

In support of the idea of genome plasticity, the effector repertoires also differ between strains within the same species. Cazalet *et al*., compared the genome sequences of *L. pneumophila* from strains Paris and Lens [61]. They identified 10% of Lens genes and 14% of Paris genes as strain-specific. The Paris strain contains 10 eukaryotic-like genes, including several with ankyrin domains, that are not found in the Lens strain [61]. Likewise, the Lens strain contains 3 eukaryotic-like genes not found in the Paris strain [61]. This diversity is remarkable especially when one considers that these differences occur between members of the same serotype and species [61]. It is possible that this

plasticity exists to allow *L. pneumophila* to parasitize the multitude of protozoan hosts it may encounter and, thereby, to adapt to virtually any aquatic environment.

Burstein *et al*., analyzed the genome sequence of 38 *Legionella* species, including *L. pneumophila,* and compared the repertoire of effectors contained within each [59]. The genome size varied significantly from 2.37 Mb in *L. adelaidensis* up to 4.82 Mb in *L. santicrucis* [59]. The percentage of guanine and cytosine nucleotides in the genome (GC content) was also highly variable and ranged from 36.7% to 51.1% [59]. The number of putative effector genes varied greatly from 52 in *L. adelaidensis* to 247 in *L. waltersii* [59]. Interestingly, the genome size is not necessarily predictive of the number of effectors, since species with very different size genomes can have similar numbers of effectors [59]. Their analysis identified a total of 5,885 predicted effectors among all the species with very little overlap and only 7 core effectors common to all [59]. These core effectors seemed to evolve over a long period of time suggesting they are fundamental to *Legionella* pathogenesis. The effectors unique to a given species also had low GC content suggestive of recent acquisition by horizontal gene transfer from ameobal or protozoal hosts which have a lower GC content than *L. pneumophila* [59]. *L. pneumophila* had the most species-specific effectors and the most dynamic complement of effectors [59]. This suggests that *L. pneumophila* is particularly prone to horizontal gene transfer and may explain why most infections in humans are due to *L. pneumophila*. Overall, the genomes

of *Legionella* spp. are highly variable in relation to size and GC content, as well as the number of effectors and their predicted function.

Identification of the function of individual *L. pneumophila* effectors using a genetic approach has been challenging because most single gene knock-outs do not result in a growth defect due to the high degree of functional redundancy among the effectors [62, 63]. Ghosh and colleagues describe five types of redundancy [62]. In molecular redundancy, two or more effectors target the same host protein using the same mechanism. Target redundancy occurs when multiple effectors target the same host protein but by distinct molecular mechanisms. In pathway redundancy, effectors alter a single host pathway but each target different components. Cellular process redundancy occurs when multiple effectors target analogous or complementary host pathways that control the same host cell process. Finally, system redundancy is where effectors target multiple host cell processes to accomplish a single goal. Modulating cell death is an example of this type of redundancy, since *L. pneumophila* can alter apoptosis, necrosis, pyroptosis, or autophagy. The presence of so many types of redundancies suggests that *L. pneumophila* has evolved over a long period of time and become highly sophisticated as an intracellular pathogen. *L. pneumophila* has the highest rate of gain and loss of effectors among *Legionella spp*., making it particularly malleable [59]. By frequently shifting its effector repertoire, *L. pneumophila* may have had the best chance at making the jump from infecting protozoans to infecting human macrophages.

8. Ankyrin Domain-Containing Proteins in *L. pneumophila*

The majority of ankyrin domain-containing proteins are found in eukaryotic organisms where they are some of the most common protein domains. The domain was first described in yeast and Drosophila but is named after the human protein responsible for attachment of the cytoskeleton to the plasma membrane [64]. Ankyrin domain proteins are involved in numerous cellular processes including cell-cell signaling, development, cytoskeleton anchoring, transcription, and cell cycle regulation [65].

The ankyrin repeat (ANK) is a 33-residue motif that forms a helix-loop-helix conformation and is involved in protein-protein interactions [66]. The helices are arranged anti-parallel with the loop projecting out to make interactions with adjacent loops forming a β-sheet structure [65]. This basic structure can be repeated from 1 to 33 domains, although most proteins contain six or fewer repeats [65]. Key residues that define the motif and stabilize the first α-helix include the TPLH sequence at positions 4 through 7. The V/I-V-X-L/V-L-L motif (X is any hydrophilic amino acid) contributes to the second α-helix and stabilizes adjacent ANK repeats [65, 67]. The substrate binding surface of the ankyrin domain is composed of the β-hairpin/loop region as well as the inner short helices [67]. These interaction surfaces were determined from co-crystal structures of six ankyrin repeat proteins and determine the specificity of the protein-protein interaction.

Proteins containing ANK domains are also found among bacteria, archaea, and many viral genomes [65]. Fifteen *ank* genes have been identified

in four sequenced *L. pneumophila* genomes with 11 shared among all four (*ankB, ankC, ankD, ankF, ankG, ankh, ankl, ankJ, ankK, ankN, ankQ*) [64]. Ankyrin repeat genes in *L. pneumophila* were likely acquired through horizontal gene transfer from amoebal or protozoal hosts and appear to be novel without significant homology to known *ank* genes [57, 68]. *L. pneumophila* effectors containing ANK domains are diverse in that they can be combined with F-box domains, leucine-rich repeats, cysteine peptidase domains, and glycosyltransferase domains to name a few [59]. Eight of the 11 ANK effectors are translocated through the T4SS based on adenylate cyclase fusion assays [69]. Three of the 11 effectors, *ankB*, *ankH*, and *ankJ*, are essential for intracellular growth in human macrophages and amoebae [70, 71]. While the functions of AnkH and AnkJ are unknown, AnkB is responsible for recruiting polyubiquitylated proteins to the LCV during infection in macrophages and amoebae.

9. Protein Ubiquitylation

Ubiquitylation is a post-translational modification of proteins that involves the covalent attachment of the 76-amino acid ubiquitin protein. The process occurs only in eukaryotic cells and is highly conserved from amoebae to mammals [72, 73]. Lysine residues on the target protein are the most common site of attachment, however, ubiquitin can also be conjugated on other residues including cysteine or the free α-amino group of an N-terminal residue [72]. Additional ubiquitin molecules can be added to other residues in the substrate protein resulting in multi-monoubiquitylation [74-77]. In polyubiquitylation,

additional ubiquitin molecules are added onto the first ubiquitin molecule via one of seven lysines (K^6, K^{11}, K^{27}, K^{29}, K^{33}, K^{48}, and K^{63}) creating a polymeric ubiquitin chain [74, 77]. Homogenous ubiquitin chains are named based on which K residue within ubiquitin contains the additional ubiquitin molecules [74, 77]. The nature of the polyubiquitin linkage determines the fate of the ubiquitylated protein. Typically, linkage of a single ubiquitin moiety alters protein localization or activity, whereas addition of multiple moieties, particularly on K^{48}, targets the protein for degradation by the 26S proteasome [72, 78, 79]. In contrast to targeting K^{48}-linked proteins for degradation via the proteasome, K^{63}-linked proteins can promote lysosomal degradation of plasma membrane proteins [74]. In addition, K^{63}-linkages can promote protein-protein interactions. In canonical NF-κB activation, K^{63}-linked ubiquitin chains are generated on the IL-1β receptor and recruit the TAK1-TAB1-TAB2/3 complex and the IKK complex, which allows TAK1 to phosphorylate IKK2 [80]. Ubiquitination of ribosomal proteins, for example, can stabilize interactions and promote translation [81]. In addition, K^{63}-linked proteins can serve as a scaffold to recruit DNA repair enzymes and cell cycle checkpoint proteins to sites of DNA damage [82].

Ubiquitylation is a three-step process that begins with an E1 activating enzyme that covalently attaches ubiquitin to itself via a high energy thioester linkage [72]. The activated ubiquitin is then transferred to the active site of an E2 ubiquitin conjugating enzyme. Finally, the E2-ubiquitin interacts with an E3 ubiquitin ligase complex that transfers ubiquitin to a specific substrate. E3 ubiquitin ligases can be either HECT, RING, or RING-IBR-RING (RBR) type [83,

84]. RING-type E3 ligases consist of either a single U-box domain protein or as part of a complex, such as the SKP1-CUL1-F-box (SCF) complex [78, 83]. The SCF complex is composed of a RING-box 1 protein (RBX1), cullin 1 (CUL1), S-phase kinase associated protein 1 (SKP1), and an F-box domain containing protein [85, 86]. The F-box domain protein also contains a protein-protein interaction domain such as WD40 or leucine-rich repeat (LRR) domain, which directs the E3 ubiquitin ligase complex to specific substrates [87]. This makes F-box proteins modular in nature, and they contain a domain for interacting with the cellular ubiquitylation machinery and a domain for interacting with substrate proteins.

E3 ligases of the SCF family recognize substrates by multiple mechanisms including by binding specific degron motifs within the targeted proteins, by altered protein localization, and by domain-based recognition [88, 89]. Canonical degrons are short defined motifs in target proteins that are recognized by the F-box protein only after they have been phosphorylated by specific kinases. For example, the FBXW1 protein binds the degron Asp-Ser-Gly-X-X-Ser (where X is any amino acid and both Ser are phosphorylated) [88, 90]. This provides specificity to the E3 and restricts degradation to cellular conditions under which the protein kinase is active. Recognition of substrates by F-box proteins can be restricted to certain subcellular compartments [91]. The FBXL20 protein contains a CaaX motif, which undergoes isoprenylation directing it to membranes [92]. FBXL20 can only target its substrate protein for degradation when it is localized to membranes, which provides control of

recognition by restricting access of the E3 to its substrate [92]. Finally, F-box proteins can recognize substrates by binding specific domains [93]. The FBXO4 protein binds TRF1 via an intermolecular beta sheet as opposed to a short degron [94]. *L. pneumophila* encodes multiple F-box proteins which have the potential to interface with host cell processes and drastically alter the ubiquitylation state of the cell.

10. Manipulation of Host Ubiquitylation Pathways

During infection of macrophages and amoebae, the LCV acquires an aggregate of polyubiquitinated proteins. *L. pneumophila* strain Philadelphia-1 encodes five F-box proteins including LegU1, LicA, AnkB, PpgA, and Lpg2525. Ensminger and Isberg characterized the E3 ubiquitin ligase activity of these proteins [95]. All 5 of the effectors are delivered by the T4SS during infection. Three of these F-box proteins (LegU1, AnkB, and LicA) associate with components of the host ubiquitylation machinery [95]. LegU1 and AnkB were shown to interact with host SKP1 and CUL1 in a manner dependent upon an intact F-box domain [95]. LegU1 can bind to the host protein BAT3 when ectopically expressed in HEK-293T cells and can direct its ubiquitylation *in vitro*. The fact that *Legionella* encodes multiple F-box proteins and that these proteins interface with the host ubiquitylation machinery, and target host proteins for ubiquitylation, suggests that this process should be critical for intracellular growth. However, *legU1, ankB*, and *lpg2160* were all shown to be dispensable for intracellular growth of the Philadelphia strain in both murine macrophages and amoebae [95]. In addition, a triple F-box mutant (*legU1, ankB*, and *licA*) in the

Philadelphia-1 strain background still accumulated polyubiquitinated proteins around its LCV [96]. Although this may seem to argue against the importance of manipulating the ubiquitylation machinery, it may simply be explained by functional redundancy and highlight the importance of this process for intracellular growth in various hosts. If manipulating the ubiquitylation machinery is crucial for infection, one might expect *L. pneumophila* to have acquired multiple F-box proteins from the many hosts it has encountered over time.

The importance of each of the F-box proteins in *Legionella* depends on the strain. Different strains encode different compliments of F-box proteins. For example, the Paris and Lens strains have three genes that encode F-box proteins, whereas the Philadelphia strain encodes five [97]. The F-box effector, AnkB, is essential for accumulation of polyubiquitinated proteins around the LCV during infection in some strains. In the Paris strain, the percentage of LCVs associated with ubiquitin is reduced by about half in *ankB* mutant infected cells compared with WT infected cells in the macrophage-like THP-1 cell line [98]. In the AA100/130b strain, the percentage of LCVs surrounded by ubiquitin dropped from 88% in WT infected hMDMs to 30% in *ankB* mutant infected cells [99]. In both the AA100/130b and Paris strains, AnkB is essential for intracellular growth in macrophages (hMDMs and U937 for AA100/130b and THP-1 for Paris) and amoebae [71, 98]. However, AnkB is not essential in the Philadelphia strain for robust intracellular growth within mouse macrophages [95]. These results indicate that among F-box proteins, there is a high degree of functional

redundancy and that AnkB may make the largest contribution to manipulation of host ubiquitin pathways in certain strains.

In addition to F-box proteins, *L. pneumophila* also utilizes several other means to promote ubiquitylation of proteins. First, the LubX effector contains an eukaryotic U-box domain and has E3 ubiquitin ligase activity [100]. Interestingly, it ubiquitylates another *L. pneumophila* effector, SidH, and is designated as a metaeffector or an effector that regulates other effectors [100]. Second, the SidC and SdcA effectors also possess E3 ligase activity and use a mechanism that involves the CHD catalytic triad [101]. The targets of these proteins are unknown, however, they are involved in recruiting ER vesicles and polyubiquitylated proteins to the LCV [73]. Finally, *L. pneumophila* also employs a novel ubiquitylation strategy that uses NAD as an energy source and is independent of E1 and E2 enzymes [102]. An ADP-ribosyl group from NAD is added to a ubiquitin molecule resulting in an activated intermediate. This intermediate is attacked by a Ser residue in a substrate protein resulting in the covalent addition of ubiquitin to the substrate through a phosophor-ribosyl linkage. The SidE family of effectors contains a mono-ADP-ribosyltransferase (mART) motif that catalyzes the ubiquitylation of host Rab33b and Rab1 by this mechanism [103]. Thus, *L. pneumophila* utilizes diverse strategies to manipulate the host ubiquitylation machinery to promote its replication within eukaryotic cells.

11. Manipulation of Host Prenylation

Protein prenylation is a eukaryotic post-translational modification of proteins that facilitates anchoring to membranes [104-106]. The process

involves covalent attachment of either a 15 carbon farnesyl or 20 carbon geranylgeranyl moiety onto a conserved cysteine residue within the CaaX motif, where “a” is an aliphatic amino acid and “X” is any amino acid [104, 106-108]. The first step is attachment of the farnesyl group by farnesyltransferase (FTase) or geranylgeranyl group by geranylgeranyltransferase (GGTase) to the cysteine within the CaaX motif [109]. The prenylated protein is further processed by the activity of the Ras-converting enzyme I (Rce1) which removes the -aaX tripeptide from the C-terminus [109]. Finally, the isoprenylcysteine carboxyl methyltransferase (Icmt) enzyme methylates the prenylated cysteine residue [109]. The modified protein can be anchored to the plasma membrane or other subcellular membranes including the LCV membrane [104, 109, 110].

In two studies, eleven genes encoding proteins containing a eukaryotic CaaX motif were identified among four *L. pneumophila* strains (*Philadelphia, Paris, Lens,* and *Corby*) with five being conserved in all 4 strains [111, 112]. When ectopically expressed in mammalian cells, six of these displayed a plasma membrane localization indicative of prenylation. When the cysteine residue within the CaaX motif was mutated to alanine, all the proteins redistributed to the cytoplasm. In addition, the LCV accumulated prenylated proteins during infection in a *dot/icm*-dependent manner. It is likely that these prenylated proteins are either *dot/icm* translocated effectors or host proteins recruited to the LCV in a *dot/icm*-dependent manner. Therefore, *L. pneumophila* hijacks the protein prenylation machinery to lipidate and anchor specific CaaX motif-containing effectors to the surface of the LCV.

There is evidence that prenylation of *L. pneumophila* effectors is critical for intracellular replication. The AnkB effector from strain AA100/130b is a CaaX motif containing effector that has been characterized in more detail [110]. Ectopically expressed AnkB localizes to the plasma membrane in a CaaX motif-dependent manner [110]. In addition to the CaaX motif, host cell FTase activity is required for the membrane localization of ectopically expressed AnkB [110]. The CaaX motif and host cell FTase activity are also required for localization of AnkB to the cytosolic face of the LCV membrane during infection in hMDMs and amoebae [110]. Importantly, farnesylation of AnkB is essential for intracellular replication within macrophages and amoebae as well as intrapulmonary proliferation in a mouse model of Legionnaires' disease [110]. Even though farnesylation of AnkB is essential for its biological function, a naturally occurring variant of AnkB in the Paris strain (AnkB Paris) lacks this motif altogether. This *ankB* allele contains a mutation at nucleotide 450 that truncates the C-terminus eliminating the CaaX motif and creating a new KNKYAP sequence. This sequence matches the eukaryotic di-lysine motif consensus, KXKXX, found in the C-terminus of proteins, which is responsible for trafficking proteins from the Golgi to the ER [113-115]. This will be discussed further in chapter 2.

12. AnkB

AnkB is a 172-amino acid protein that contains a eukaryotic F-box domain, which interacts with the host E3 SCF ubiquitin ligase complex, thus mimicking host F-box proteins [71, 99]. AnkB also contains three eukaryotic-like ANK domains, that likely bind specific substrates, and a C-terminal eukaryotic CaaX

motif [110]. AnkB is the only effector known to be required for intracellular replication in both amoeba and human cells, and in a mouse model of intrapulmonary proliferation [71, 98, 99, 116]. It's importance during infection is strain specific, as knockouts made in the AA100/130b or Paris strain result in a replication defect, however, a knockout made in the Philadelphia strain does not [71, 95, 98]. The Philadelphia strain encodes five F-box proteins and the AA100/130b strain encodes two, while the Paris strain encodes three [95, 98, 117]. Thus, it is possible that the Philadelphia strain does not depend upon AnkB because it contains F-box proteins that can functionally compensate for the loss of AnkB, while the Paris and AA100/130b strains do not.

A model for AnkB function is shown in Figure 1. Upon contact with a suitable host cell, AnkB is translocated through the T4SS into the host cell cytoplasm where it rapidly becomes farnesylated on its C-terminal CaaX motif [110, 118]. Farnesylation of AnkB confers hydrophobicity allowing it to insert into the cytosolic face of the LCV membrane [110]. AnkB recruits the host ubiquitylation machinery by binding to the Skp1 component of the host SCF E3 ubiquitin ligase complex via its F-box domain [99]. As a result, polyubiquitinated proteins accumulate around the LCV during infection [119]. These proteins are polyubiquitylated on K^{48} which targets them for degradation by the 26S proteasome generating a high level of cellular amino acids that drives bacterial replication [120].

Evidence for the role of K^{48}-linked polyubiquitylated proteins in generating amino acids for intracellular growth comes from data showing that inhibition of

the proteasome with MG132 prevents replication of WT bacteria in hMDMs [120]. Importantly, this defect could be completely rescued by supplementation with a 5-fold excess of amino acids, which suggests that the sole purpose of manipulating the ubiquitylation machinery is to generate a surplus of free amino acids [120]. It was shown that AnkB has a prominent role in this process. Replication of the *ankB* mutant can be rescued by addition of a 5-fold excess of amino acids or by specific single amino acids such as serine and cysteine, which are converted by the bacteria to pyruvate that feeds the TCA cycle [120]. Importantly, some non-essential amino acids also rescue the *ankB* mutant. Interestingly, addition of pyruvate or citrate can also rescue the *ankB* mutant for intracellular replication [120]. The fact that the *ankB* mutant can be rescued by addition of pyruvate or citrate indicates that sufficient amino acid levels are present within the host cell, and that higher levels of amino acids are required specifically as sources of energy to feed the TCA cycle [120]. RelA, an indicator of amino acid starvation, is upregulated in the *ankB* mutant during an infection [121]. Importantly, induction of RelA was reversed, and intracellular replication of the *ankB* mutant was rescued, by addition of a 5-fold excess of amino acids [120]. Overall, the *ankB* mutant undergoes a starvation response within amoebas and hMDMs, which is consistent with a nutritional role for AnkB. These data indicate that the K^{48}-linked polyubiquitylated proteins surrounding the LCV are assembled by AnkB and are degraded by the proteasome, which results in a surplus of free amino acids that are taken up by the bacteria and shunted through the TCA cycle to generate ATP to drive replication.

While AnkB seems to be a nutritional virulence factor that recruits polyubiquitylated proteins to the LCV in order to generate amino acids for replication, the specific substrates of AnkB are currently unknown and their identification could reveal novel roles. It seems unlikely that AnkB would indiscriminately target proteins for degradation, since ankyrin domains typically mediate specific protein-protein interactions [66]. Identification of AnkB substrates could reveal that it also has a role in modulating certain host pathways by targeting specific proteins for degradation. Alternatively, AnkB may target specific host proteins rich in amino acids for which *L. pneumophila* is auxotrophic or those rich in serine or cysteine to drive energy production via the TCA cycle.

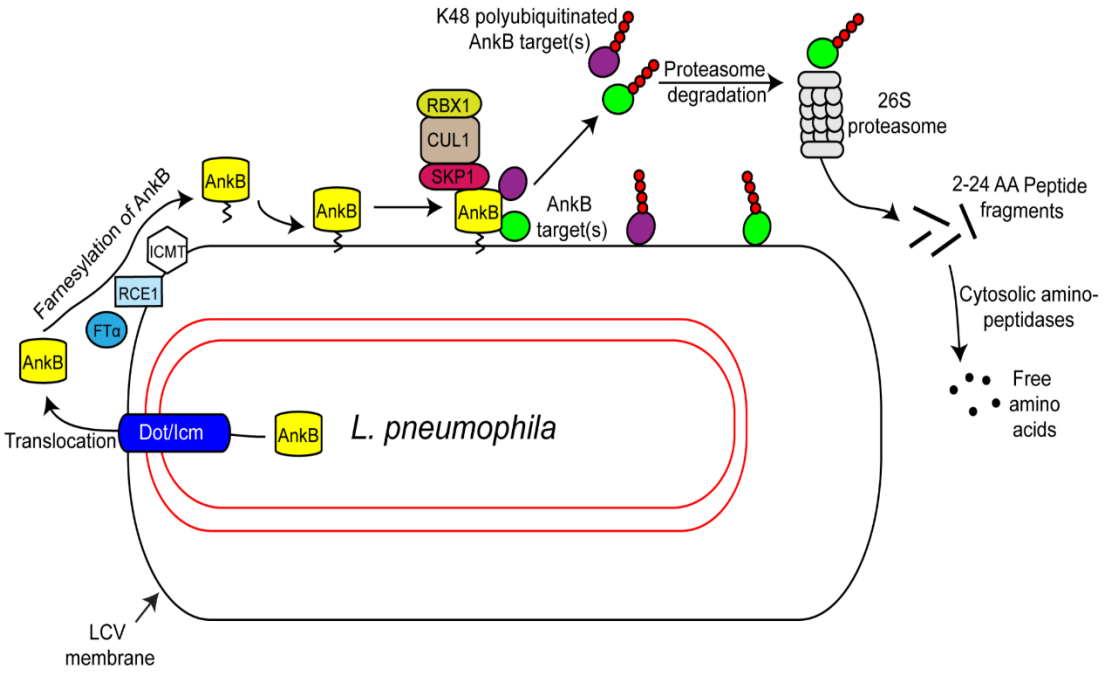

Fig. 1. Model of AnkB Function. AnkB is translocated via the Dot/Icm system where it rapidly becomes farnesylated by the host enzymes FTα, RCE1, and ICMT. This allows it to insert into the cytosolic face of the LCV membrane where it interacts with the host ubiquitylation machinery and directs the K^{48}-linked polyubiquitylation of specific host proteins. These proteins are degraded by the proteasome into peptide fragments which are further processed into free amino acids. Adapted and modified from reference [78].

SPECIFIC AIMS

AnkB of the AA100/130b strain (AnkB AA100) is farnesylated and inserts into the cytoplasmic face of the LCV membrane, a process which is essential for recruiting polyubiquitylated proteins to the vacuole and for full virulence of *L. pneumophila* [110]. Substitution of the farnesylated cysteine to alanine blocks farnesylation of AnkB and causes a loss of LCV polyubiquitin accumulation and results in a replication defect identical to the *ankB* null mutant. Ectopic expression of AnkB-AA100 in HEK293T cells results in farnesylation of the effector which directs its localization to the plasma membrane and *trans*-rescues the *ankB* mutant for intracellular replication [110].

Unlike AnkB AA100, a naturally occurring *ankB* allele found in the Paris strain (AnkB Paris) localizes to the leading edge of lamellipodium formation along with α-actinin [98]. The *ankB Paris* allele contains a mutation at nucleotide 450 that truncates the C-terminal 18 amino acids, deletes half of the last ankyrin repeat, and eliminates the CaaX motif. This frame shift mutation creates a new C-terminus, which resembles a eukaryotic di-lysine motif. In eukaryotic cells, this motif is responsible for trafficking proteins from the Golgi to the ER [115]. Despite lacking the farnesylation motif essential for function of AnkB AA100, AnkB Paris still mediates recruitment of polyubiquitylated proteins to the LCV and is essential for intracellular replication and virulence in a mouse model, although

to a lesser degree compared to strain AA100/130b [71, 98]. Therefore, AnkB Paris likely contains an alternative anchoring mechanism that allows it to localize to the ER-derived LCV membrane during infection.

Except for a single target of AnkB Paris, the proteins ubiquitylated by AnkB have not been identified [98]. The primary role of decorating the LCV with polyubiquitylated proteins appears to be in the generation of free amino acids to power replication, since an *ankB* mutant can be rescued by a 5-fold excess of amino acids [120]. However, this does not exclude the possibility that AnkB may target specific host proteins to manipulate specific host pathways. Identification of the specific proteins targeted by AnkB could provide new insights into the virulence mechanisms employed by *L. pneumophila*.

I hypothesize, that AnkB from the Paris strain contains a eukaryotic-like di-lysine motif that is critical for directing it to the cytoplasmic face of the ER-derived LCV during infection and that, once there, it binds and ubiquitylates specific host proteins.

Specific Aim 1: Determine the mechanism by which AnkB Paris is anchored to the cytosolic face of the LCV membrane.

Specific Aim 2: Determine the residues within the ankyrin domains of AnkB that are responsible for binding substrates and promoting their ubiquitylation.

Specific Aim 3: Identify host and/or *L. pneumophila* substrate proteins that are ubiquitylated by AnkB during infection.

MATERIALS AND METHODS

Bacterial strains, cell cultures, and infections

L. pneumophila strain AA100/130b (ATCC BAA-74), its isogenic *dotA* and *ankB* mutants, and complemented mutants were grown on BCYE agar plates for 3-4 days at 37^0C prior to infection as previously described [71]. When required, antibiotics were used at a concentration of 50 µg/mL for kanamycin and 5 µg/mL for chloramphenicol. The *E. coli* strain DH5α was used for cloning. *E. coli* was grown in Luria-Bertani (LB) and antibiotics were used at a concentration of 100 µg/mL for ampicillin and 40 µg/mL for chloramphenicol. HEK293T cell line was maintained in DMEM (Gibco, Grand Island, NY) supplemented with 10% FBS.

Purification and preparation of human monocyte-derived macrophages (hMDMs) was performed as previously described. Monocytes were isolated from whole blood of healthy donors and then allowed to adhere to 6 well low adherence cell culture plates for 3 days at 37°C and 5% CO_2 in RPMI 1640 supplemented with 20% FBS. Monocytes were then counted and re-suspended RPMI 1640 supplemented with 10% FBS and plated on coverslips at a density of 2 X 10^5 cells per well of a 24 well cell culture plate and incubated for a further 2 days. The cell culture media was then replaced with RPMI 1640 supplemented with 5% FBS for one day, and then with RPMI 1640 supplemented with 1% FBS for one day. The resulting hMDMs were then used for infection.

All methods were carried out in accordance with relevant guidelines and regulations. We confirm that all experimental protocols were approved by the Institutional Review Board (IRB) committee (IRB# 04.0358). We confirm that informed consent was obtained from all subjects, as required per our approved IRB protocol.

Infection of hMDMs was performed as previously described. Bacteria were resuspended in RPMI 1640 with 10% FBS and macrophages were infected in triplicate for 1 hour at a multiplicity of infection (MOI) of 10. Plates were centrifuged at 200 g for 5 minutes to synchronize the infection. Infected cells were treated with 50 µg/mL gentamicin for 1 hour to kill extracellular bacteria. Following gentamicin treatment, cells were washed three times with Hank's buffered saline solution (HBSS) and then RPMI containing 10% FBS was added. At 10 hours post infection, cells were fixed and processed for confocal microscopy. Infection of HEK-293 cells was performed at an MOI of 50 for 1 hour followed by treatment with gentamicin 50 µg/mL for 1 hour. At 10 hours post infection, cells were fixed and processed for confocal microscopy.

HEK293T cell transfection and infection

To create *ankB*-Paris and *ankB*-ParisK$^{149, 151}$R, *ankB* from strain AA100/130b cloned into the mammalian vector p3XFlag-CMV-10 (Sigma) was used as a template for site directed mutagenesis by PCR. HEK293T cells (85% confluent) were re-plated onto poly-L lysine treated coverslips in 24 well plates at a density of 5x10^4 cells/well. After overnight incubation, cells were transfected with 0.625 µg plasmid DNA per well using 1.5 µg polyethylenimine (PEI) per well.

After 24 hours, cells were infected with bacteria suspended in DMEM at an MOI of 100 for 1 hour at 37^0C and 5% CO_2. Plates were centrifuged at 200 g for 10 minutes to synchronize the infection. Extracellular bacteria were eliminated by treatment with gentamicin 50 µg/mL for 1 hour. At 10 hours post infection, cells were fixed and processed for confocal microscopy.

LCV AnkB localization

To the determine localization of AnkB on the LCV surface during infection, post-nuclear supernatants of infected hMDMs were prepared and then differentially labeled as described previously [110]. Briefly, a total of 1 X 10^6 hMDMs were infected with *L. pneumophila* at an MOI of 10 for 2h. Post nuclear supernatants were prepared as described previously and LCVs were allowed to adhere to poly-L-lysine coated glass coverslips and fixed using 4% paraformaldehyde [110]. To differentiate between intact LCVs and extracellular bacteria, the LCVs were labeled prior to permeabilization with mouse anti-*Legionella* antisera (1/1000 dilution) and rabbit anti-AnkB antisera (1/200 dilution) for 1h. LCVs were then permeabilized with -20°C methanol and counter-labeled with goat anti-*Legionella* antisera (1/1000 dilution) for 1h to detect intact LCVs. The LCVs were then labeled with Alexa-Fluor conjugated secondary antibodies (anti-mouse 488, anti-rabbit 555 and anti-goat 647) following the manufacturers recommendations (Invitrogen).

Confocal microscopy

Processing of infected cells for confocal microscopy was performed as we described previously [110]. Purification of the LCVs and their labeling prior to

permeabilization to localize AnkB on the cytosolic side of the LCV membrane was performed as we described previously [110]. For antibody labeling, goat polyclonal anti-*L. pneumophila* was used at a dilution of 1:500 and detected by Alexa-Fluor 488-conjugated donkey anti-goat IgG (Invitrogen, Carlsbad, CA). Poly-ubiquitinated proteins were detected using mouse anti-polyubiquitin FK1 antibody at a dilution of 1:50 (BIOMOL International/Affiniti, Exeter, United Kingdom), followed by Alexa-Fluor 647-conjugated goat anti-mouse IgM (Invitrogen, Carlsbad, CA). For detection of 3X-Flag tagged proteins during transfection experiments, mouse monoclonal anti-Flag (Sigma) antibodies were used followed by detection with Alexa-Fluor 488-conjugated donkey anti-mouse (Invitrogen, Carlsbad, CA). An Olympus FV1000 laser scanning confocal microscope was used to examine cells as we described previously. On average, 8-15 0.2 µm serial Z sections of each image were captured and stored for further analyses, using Adobe Photoshop CS3.

Adenylate cyclase and Western blot analysis

L. pneumophila strain AA100/130b (ATCC BAA-74) or its isogenic *dotA* mutant harboring pCya-AnkB-Paris, pCya empty vector, or pCya-AnkB-Paris with $K^{149}R$ or $K^{151}R$ were grown on BCYE agar plates for 3-4 days at 37^0C prior to infection as previously described [71]. U937 macrophages differentiated with PMA at 50 ng/mL were infected at MOI 50 in triplicate and plates were centrifuged to synchronize the infection. After 2 hours at 37°C and 5% CO_2, cells were washed three times with PBS and lysed by adding 250 µl of 0.1 N HCl containing 0.5% Triton X-100 and incubating at room temperature for 20 minutes.

Lysates were assayed for cAMP using the Direct Cyclic AMP Enzyme Immunoassay kit (Enzo Life Sciences, Inc.). Aliquots of bacteria used for infection (1 x 10^8 bacteria) were lysed by adding SDS-PAGE loading buffer and boiling for 5 minutes. Fusion protein expression was assessed by Western blot using anti-M45 (1:50 dilution) according to standard procedures. Blots were re-probed with anti-CAT (1:2000).

PCR and Sequencing of *ankB* alleles

The *ankB* allele was amplified with the following primers: ankB1F: 5'-GGATCCCAAGAGATTTTTAG-3' and ankB1R: 5' – CATTTAACAAACAAGGCACT-3' using standard PCR conditions. PCR primers were located in genes flanking *ankB*. Briefly, 25 ng of genomic DNA was used as template in a 25 µL PCR reaction containing 1U of Taq polymerase (Midsci, St. Louis MO), 150 µM dNTPs, 20 pmol/µl of each primer with the following cycling parameters: 94^{o}C-5' – 1 cycle followed by 30 cycles of 94^{o}C-1', 55^{o}C-1, 72^{o}C 1' and a final 5 min extension at 72^{o}C. DNA sequencing was performed on both strands at the University of Washington Sequencing Core, and the sequence data was assembled and edited using the DNASTAR suite (DNASTAR Inc., Madison, WI).

Phylogenetic Analysis

Maximum likelihood (ML) tree was constructed using MEGA version 6 assuming the TN93 + G substitution model [122]. The percentage of trees in which the associated taxa clustered together was determined by a bootstrap analysis of 1000 trees. Initial tree for the heuristic search was obtained

automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.2388)). Nucleotide sequence data has been submitted to GenBank® and assigned the following accession numbers: KM276667-KM276681.

Analysis of Selection Pressures

(a) Site Models: To identify the different selective forces, i.e., negative, neutral or positive selection, that acted upon *ankB* codons during its evolutionary history we tested the fit of the sequence data to several codon-based models implemented in CODEML package of PAML ver 4.7 accessed via its GUI interface PAMLX1.2 essentially as described before [123-125]. In brief, we used *site models* to determine selective pressures on each *ankB* codon by comparing the differences in the likelihood score of each model's fit to the sequence data via a series of likelihood ratio tests [LRTs) [123]. To verify or supplement CODEML outcomes, we conducted several other alternate tests including GARD (genetic algorithms for recombination detection) to detect recombination among *ankB* sequences, and SLAC (single-likelihood ancestor counting), FEL (fixed effects likelihood), IFEL (internal fixed effects likelihood), REL (random effects likelihood), and MEME (mixed effects models of evolution), which can each detect positive and negatively selected codon in protein coding genes and can explicitly account and correct for recombination within sequences. All these

methods were accessed and their outcomes analyzed via the www.datamonkey.org server [126].

(b) Branch Site Models. To determine whether the *ankB1* allele branch experienced positive selection in its evolutionary history we used two versions of the branch-site models A (M2N2) implemented in CODEML: 1) M2N2A1, which specifically tested for evidence of positive selection in the clade leading up to *ankB1* and *ankB8; and* 2) M2N2A2, which specifically sought evidence for positive selection in the *ankB1* branch itself. The fit of each model to the data was tested via LRTs with 1 degree of freedom and that measured the difference in the likelihood score of each model (eg., M2N2A1) with a constrained version whereby ω for the branch suspected to be under positive selection was fixed at 1 (e.g. M2N2A1ωf). It has been suggested that selection of branches of interest to test for selection, or testing one branch at a time can sometimes lead to statistical instability or acceptance of poorly supported models [126]. Thus to confirm the outcomes of our CODEML branch site results, we performed supplemental analysis for detecting all branches that may have significantly experienced positive selection in their evolutionary history with the GA (genetic algorithm) branch method implemented at www.datamonkey.org.

Infection of A549 GFP-BirA Cells and Effector Complex Isolation

To create *pBio-ankB* and *pK/A-ankB*, *ankB* from strain AA100/130b cloned into the pICC1544 vector described previously and transformed into WT *L. pneumophila* strain AA100/130b [127]. Bacteria were grown overnight to OD550 of 1.5-1.7, induced for 1 hour with 1 mM IPTG, and used to infect A549 GFP-BirA

cells. A549 GFP-BirA cells were maintained in G418 at a final concentration of 800 µg/mL and plated the day before infection at 5.5-5.75 x 10^6 cells per plate. Cells were infected at MOI 50 and centrifuged for 10 minutes at 400 x g at room temperature before being placed into the incubator at 37°C and 5% CO_2 for 2 hours. Plates were washed three times with 1x PBS and returned to the incubator for 7 hours. At the end of the infection, plates were washed twice with 1x PBS and cross-linked using 10 mL of 1% formaldehyde (in 1x PBS) per plate for 30 minutes at room temperature. Cross-linking was quenched by addition of 1 mL of 1.25 M glycine and 50 mM cysteine (in 1x PBS) for 15 minutes at room temperature. Cells were washed three times in 1x PBS. Plates were scraped at 4°C in 1x PBS, centrifuged for 10 minutes at 500 x g, and stored at -80°C. Cells were lysed for 30 minutes in 50 mM Na_2HPO_4, 150 mM NaCl, and 1% CHAPS, pH 7.3 using 1 mL per plate. Cellular debris was pelleted at 20,000 x g for 20 minutes at 4°C. Sixty µl of equilibrated nickel resin (ThermoFisher 88221) was added to each sample and incubated for 2 hours at 4°C. Resin was washed 5 times in lysis buffer plus 20 mM imidazole and protease inhibitors. Bound complexes were eluted in lysis buffer plus 250 mM imidazole three times for 10 minutes each at 4°C with shaking. Elutions were pooled and 50 µL of equilibrated streptavidin agarose (ThermoFisher 20357) was added and samples were incubated for 2 hours at 4°C. Resin was washed 4 times with lysis buffer and 4 times with 50 mM ammonium bicarbonate (AMBIC). Fifty µL AMBIC was left over beads with the final wash.

Cloning, Protein Expression and Purification of AnkB and Skp1

The human gene Skp1 (residues 1 to 163) was first cloned into pRSFDuet-1 between NdeI and AvrII restriction sites. The gene AnkB (lpg2144, residues 1 to 168) from *Legionella pneumophila* strain Philadelphia was then cloned into the same vector between BamHI and NotI as a N-terminal His-tagged fusion protein. The C-terminal ankyrin domain (residues 54 to 168) was cloned into pET15b as a N-terminal His-tagged fusion and pET29a as a C-terminal His-tagged fusion. Mutagenesis was performed using the QuikChange Multi Site-Directed Mutagenesis Kit (Agilent Technologies). All constructs were verified by DNA sequencing and transformed into a BL21 *E. coli* strain. The cells were grown at 37 ºC in Luria Broth (LB) to an optical density of 0.8, and expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside at 30 ºC for 4 hours or 16 ºC overnight. After centrifuging the cells, the pellets were resuspended in buffer A (50 mM HEPES, 500 mM NaCl, 5% (w/v) glycerol, pH 7.6), containing 1 mM phenylmethylsulfonyl fluoride and 0.1 mg/ml lysozyme, and lysed by sonication. Cell debris was removed by centrifugation, and the fusion protein was bound to Ni-NTA Agarose (Qiagen) beads, washed with buffer A containing 30 mM imidazole and eluted with buffer A containing 250 mM imidazole. The protein was further purified by size-exclusion chromatography on a Superdex75 column (GE Healthcare) in buffer B (10 mM HEPES, 100 mM NaCl, pH 7.0) before crystallization trials. The His-tag in the pET15b constructs was cleaved with thrombin before injecting the protein into a size-exclusion column.

For selenomethionine labeling, the plasmid was transformed into a methionine auxotroph, *E.coli* DL41 (DE3), and the cells were grown in LeMaster medium supplemented with selenomethionine. For ^{15}N-labeling, the cells were grown in M9 minimal medium supplemented with ^{15}N-ammonium chloride as the sole source of nitrogen. The expression and purification protocols were the same as for the native protein.

Crystallization and Structure Determination

Crystallization was performed by the hanging drop vapor diffusion method at 293K using the Classics II commercial screen (Qiagen). Native AnkB (54-168) concentrated to 7.6 mg/mL crystallized in a 1:1 mixture with the reservoir buffer (0.2 M lithium sulfate, 0.1 M HEPES pH 7.5, 25% (w/v) PEG 3350). Crystals of the SeMet-labeled C-terminal domain were obtained at 10 mg/mL with the mother liquor (0.2 M lithium sulfate, 0.1 M Bis-Tris pH 6.5, 25% (w/v) PEG 3350). AnkB (1-168) in complex with Skp1 was concentrated to ~ 4.5 mg/mL and crystals were obtained from a condition containing 0.2 M trimethylamine N-oxide, 0.1 M Tris pH 8.5 and 20% (w/v) PEG 2000 MME.

The ankyrin domain and complex crystals were cryoprotected with 20% glycerol and 20% sucrose, respectively, and flash-cooled in a N_2 cold stream. X-ray diffraction data were collected at beamlines A1 and F1 of Cornell High-Energy Synchrotron Source (CHESS) using an ADSC Quantum 210 CCD detector. Data processing and scaling were performed with HKL-2000 [128]. The diffraction data of the ankyrin domain were phased using anomalous signal from selenium atoms by the single-wavelength anomalous dispersion method,

with the program SHELX [129]. The initial model was built with ARP/wARP and refined with Refmac5 [130, 131]. Full length AnkB in complex with Skp1 was determined by molecular replacement using Skp1 and F-box from a deposited SCF complex structure (PDB code 1LDK) and AnkB (54-168) as the search model [132]. The model was built by ARP/wARP, completed with Coot and improved by several cycles of refinement using Refmac5 [130, 131, 133]. Water molecules were added in the last stage of refinement.

The refinement statistics are shown in Table 3. The final ankyrin domain and complex structures respectively have 0 and 1 outlier in the Ramachandran plot computed using MolProbity [134].

***In vitro* Pull-Down**

AnkB (residues 54-168) was used as bait for the Pierce Pull-Down PolyHis Protein:Protein Interaction Kit (ThermoFisher 21277). Both U937 macrophages and HEK293T cells were used as a source of cellular lysate in separate experiments. The protocol was performed per the manufacturer's instructions. Elutions were run on an SDS-PAGE gel and stained with Sypro Ruby.

HEK293T Transfections and Co-IP

Cells were grown to 70% confluence and plated into T-175 flasks at 4 x 10^6 cells per flask and incubated overnight. Cells were transfected using 11 µg of plasmid DNA per flask with polyethylenimine as the transfection reagent (Polysciences) for 24 hours. Cells were lysed in 50 mM Tris HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, and 1% Triton X-100 with protease inhibitors. Flag-tagged or

HA-tagged proteins were captured using anti-Flag magnetic beads (Sigma) or anti-HA magnetic beads (ThermoFisher) per the manufacturer's instructions.

Antibodies and Western Blotting

Protein samples were heated to 95°C for 5 minutes in sample buffer, separated by SDS-PAGE and transferred to a PVDF membrane. Anti-Flag (Sigma) was used at 1:1000, anti-HA (Bethyl) was used at 1:1000, and anti-AnkB was used at 1:30,000 in 6% non-fat milk in PBS+0.05% Tween-20 overnight at 4°C. Secondary anti-mouse-HRP was used at 1:50,000 and anti-rabbit-HRP was used at 1:100,000. Blots were developed using chemiluminescent substrate (ThermoFisher) and imaged using a CCD camera.

CHAPTER 1

IDENTIFICATION OF HOST PROTEIN TARGETS OF ANKB

Results

In vitro Pull Down

Identification of biologically-relevant protein interacting partners can be challenging, particularly in the context of proteins involved in bacterial pathogenesis. Approaches include yeast 2-hybrid assays, co-immunoprecipitations, label transfer interaction analyses, far-Western blots, and *in vitro* pull downs [135, 136]. *In vitro* pull downs have been successfully used to identify novel host interacting partners of bacterial proteins in *L. pneumophila* and *Coxiella burnetii* [137-139]. We chose an *in vitro* pull down to identify host proteins targeted by AnkB. Since AnkB is an E3 ubiquitin ligase responsible for recruiting K^{48}-linked polyubiquitylated proteins to the LCV, and K^{48}-linked proteins are typically targeted for degradation; it is likely that proteins ubiquitylated by AnkB will be unstable during a natural infection. Due to solubility issues, we purified a His6 tagged version of the three ankyrin domains of AnkB (residues 54-168) from *E. coli* by nickel affinity chromatography followed by size exclusion chromatography. This purified version of AnkB was used as bait in an *in vitro* pull down with lysates from the U937 macrophage-like cell line or HEK293T cells serving as prey. Columns were washed and interacting partners

were eluted. A column without AnkB was used as a negative control. Elutions were run on an SDS-PAGE gel, stained, and unique bands were excised and identified by mass spectrometry (MS).

The results for both U937 and HEK293T cell lines are shown (Fig. 2). Each band indicated was excised from the gel along with the corresponding area of the control lane and proteins were identified by MS. There were five bands common to both cell lines of the same apparent molecular weight (MW). A total of 176 proteins were identified among these five bands (Appendices 1 and 2). The two to four most abundant proteins in each of the five bands found only in the AnkB condition were chosen for confirmatory studies using co-IP (Table 1).

To confirm putative AnkB binding partners, HEK293T cells co-expressing Flag-tagged AnkB or bacterial alkaline phosphatase (BAP) and hemagglutinin (HA)-tagged versions of each candidate substrate were subjected to co-IP using anti-Flag magnetic resin followed by Western blot using anti-HA antibodies. Overall, none of the 13 candidate proteins was found to interact with AnkB. Each potential interacting partner either non-specifically bound to the anti-Flag resin, showed no binding to AnkB or the resin, or was poorly expressed.

An example of a protein that expressed well but failed to bind either AnkB or BAP is also shown in Figure 3. Expression of PYCR was apparent in both AnkB and BAP co-expressing cells, however to a lesser degree in the BAP condition. PYCR did not co-IP with either AnkB or BAP indicating that it is not a true binding partner of AnkB under our experimental conditions. Similar results were obtained for ANP32A, ANP32B, ANP32E, CSNK2A1, and SET (data not

shown). Candidate substrates that showed low expression include DBR1, TACO1, and UGT8 (Fig. 4 and data not shown).

A representative example of non-specific binding to the beads is shown in Figure 3. HA-PYCRL co-precipitated with AnkB to a greater degree than with BAP. However, expression of PYCRL in PYCRL/BAP co-expressing cells was also lower, which likely accounts for the difference in the intensity of the PYCRL band in the AnkB and BAP lanes. Similar non-specific binding was also observed for TBL1XR1, FAM98A, and RACK1 (data not shown). PIP4k2b appeared to specifically bind AnkB as shown in Figure 5a. PIP4k2b was present in the AnkB but not the BAP immunoprecipitate. However, further investigation revealed that PIP4k2b bound non-specifically to the anti-Flag resin to a similar degree when co-transfected with either *p3x flag* empty vector or *p3x-ankB* (Fig. 5b compare AnkB/PIP4k2b to p3x/PIP4k2b). Competition of BAP for PIP4k2b binding to anti-Flag resin is one possible explanation for the conflicting results observed when empty vector or BAP is used as a negative control. A positive control co-IP using the known AnkB binding partner, Skp1, showed that Skp1 bound specifically to AnkB under these experimental conditions (Fig. 5b). Overall, this indicates that PIP4k2b is not a binding partner for AnkB under our experimental conditions.

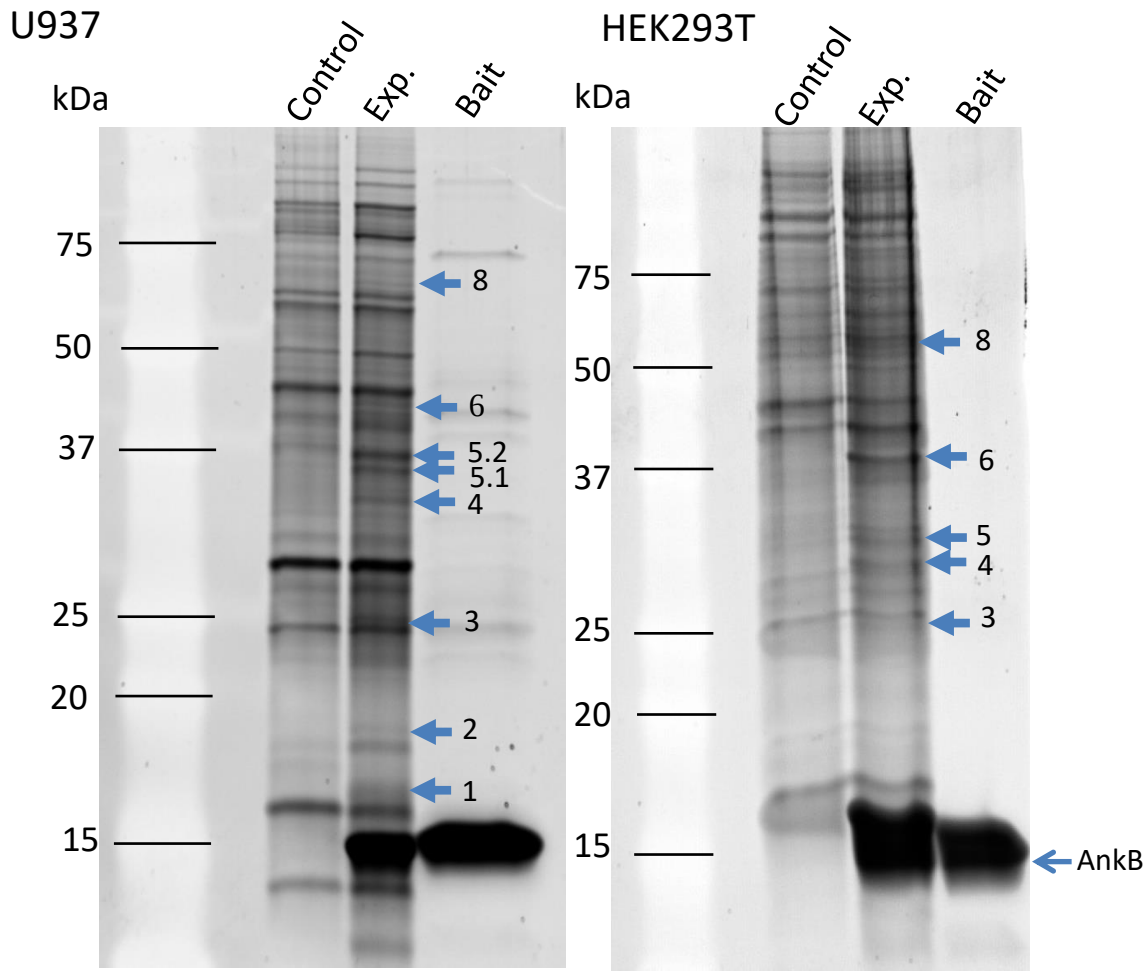

Fig. 2. ***In vitro*** **Pull Down** His-6 tagged AnkB (54-168) was bound to a nickel column and lysates from either the U937 or HEK293T cell line were passed over the column. Columns were washed and eluted. Elutions were run on an SDS-PAGE gel and stained with SyproRuby. Bands that appear to be common to both cell lines are labeled with the same number. Control- no AnkB bound to the column, Exp- AnkB bound, Bait- purified AnkB.

Table 1. Candidate AnkB-Interacting Host Proteins

Name	Description	Band
*ANP32B	Acidic leucine-rich nuclear phosphoprotein B	3
RACK1	Receptor of activated protein kinase C	3
ANP32A	Acidic leucine-rich nuclear phosphoprotein A	4
*SET	Protein SET	4
ANP32E	Acidic leucine-rich nuclear phosphoprotein E	5
TACO1	Translational activator of cytochrome c oxidase	5
PYCRL	Pyrroline 5-carboxylate reductase	5
*CSNK2A1	Casein kinase II subunit alpha	5
FAM98A	Protein FAM98A	6
*PIP4K2B	Phosphatidylinositol 5-phosphate 4-kinase type 2-beta	6
DBR1	Lariat debranching enzyme	8
UGT8	2-hydroxyacylsphingosine 1-beta-galactosyltransferase	8
TBL1XR1	F-box-like/WD repeat-containing protein	8

* Found in both U937 and HEK293T

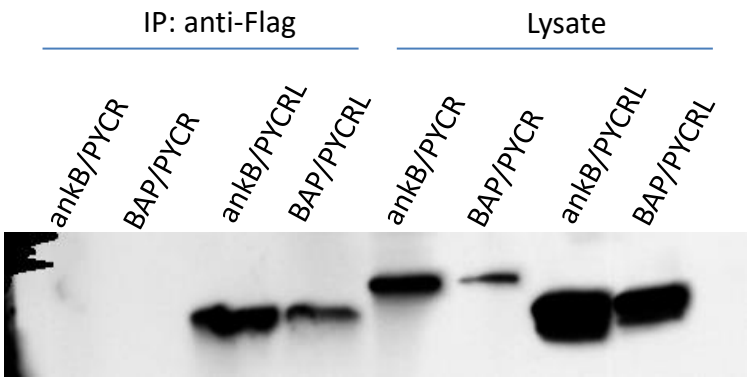

Fig. 3. Co-IP of AnkB and either PYCRL or PYCR. HEK293T cells were transfected with *Flag-ankB* or *Flag-BAP* and either *HA-PYCR* or *HA-PYCRL* as indicated. Cell lysates were incubated with anti-Flag magnetic resin, washed, and eluted. Elutions were run on SDS-PAGE gels and subjected to Western blotting with anti-HA antibodies.

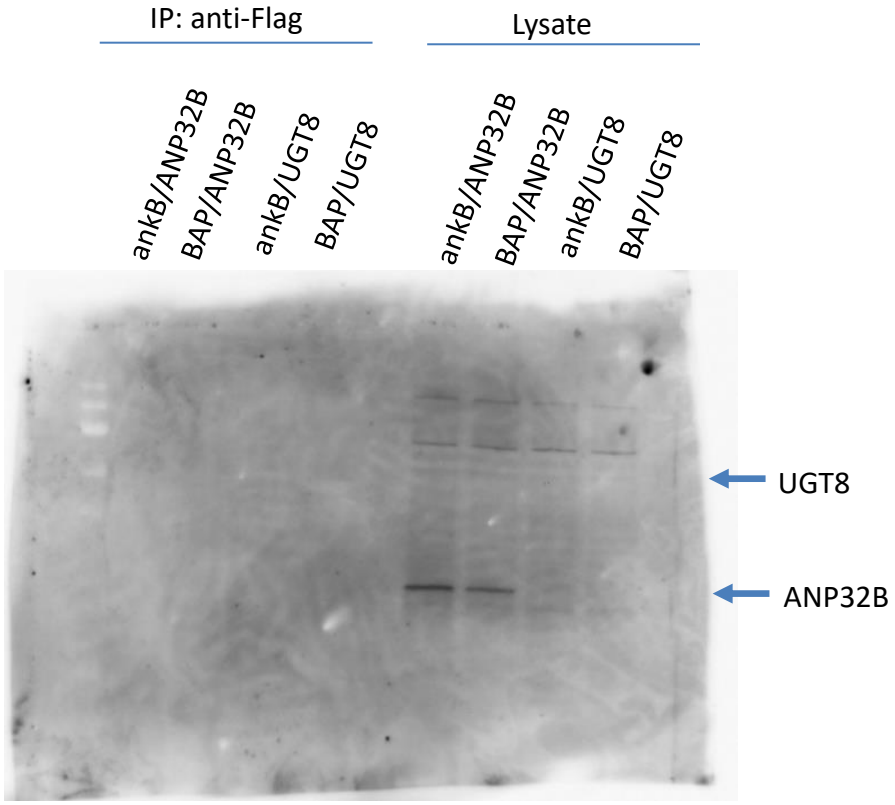

Fig. 4. Co-IP of AnkB and either UGT8 or ANP32B. HEK293T cells were transfected with *flag-ankB* or *flag-BAP* (control) and either *ha-ugt8* or *ha-anp32b* as indicated. Cell lysates were incubated with anti-Flag magnetic resin and elutions were run on SDS-PAGE gels and subjected to Western blotting with anti-HA antibodies. The expected positions of HA-UGT8 and HA-ANP32B are indicated.

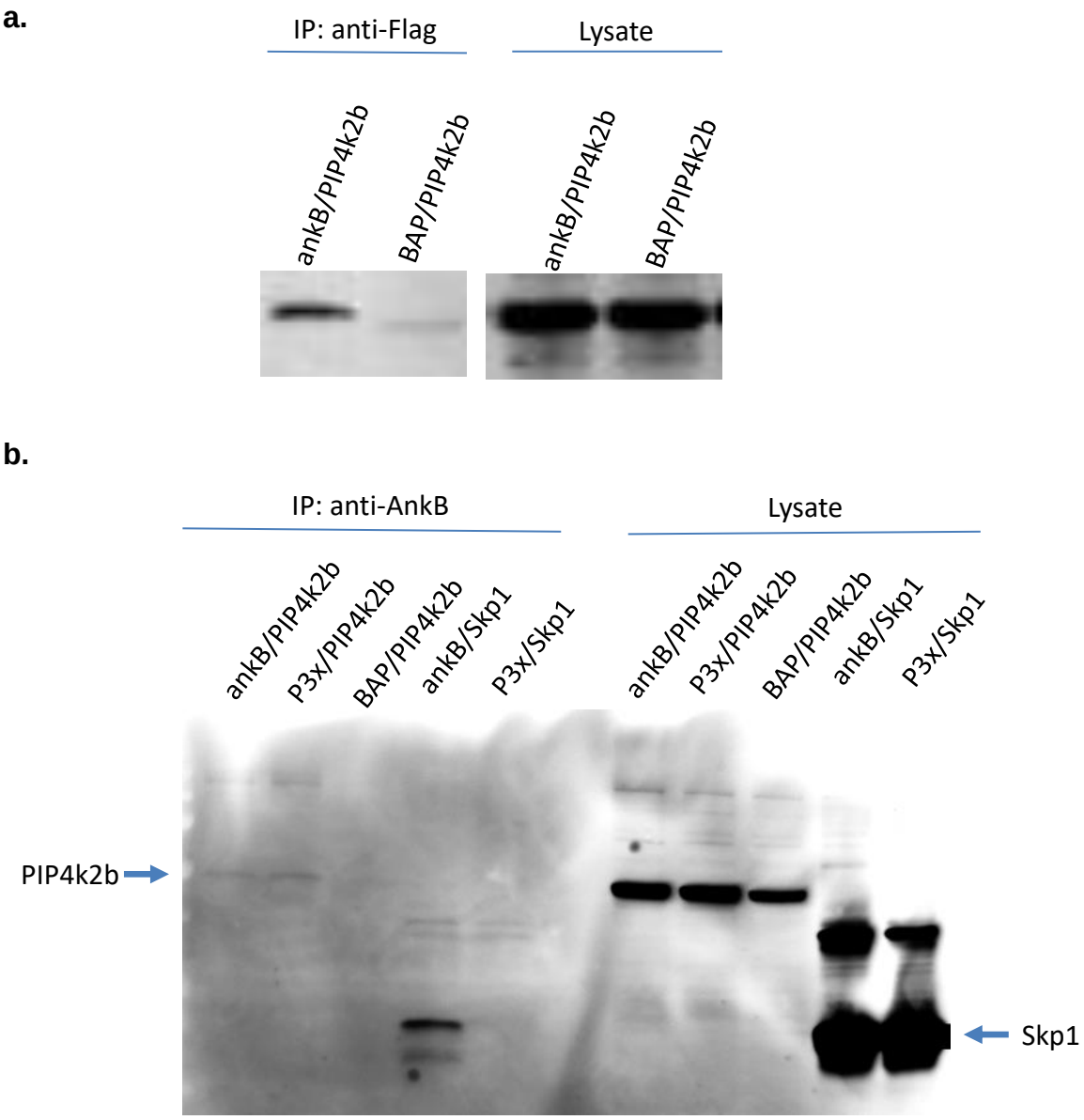

Fig. 5. Co-IP of AnkB and PIP4k2b. (a) HEK293T cells were transfected with *Flag-ankB* or *Flag-BAP* and *HA-PIP4k2b* as indicated. Lysates were incubated with anti-Flag magnetic beads and elutions were run on SDS-PAGE gel and subjected to Western blotting with anti-HA antibodies. **(b)** Similar to (a) except that an additional negative control using empty *p3X flag* vector and a positive control using *HA-Skp1* were also included. Expected positions of HA-Skp1 and HA-PIP4k2b are indicated.

Identification of AnkB Substrates by Two-Step Immunoprecipitation

The failure of the *in vitro* pull down to identify AnkB substrates may be due to various potential caveats. E3 ubiquitin ligase substrate identification is challenging due to the transient and low affinity nature of E3 ligase-substrate interactions and the instability of the resulting ubiquitylated substrates [135, 140]. Ensminger and colleagues successfully identified a substrate of the *L. pneumophila* F-box protein LegU1 using a two-step enrichment approach [95]. They ectopically expressed *HA-ubiquitin* and *3x-flag legU1* in HEK293T cells and purified in tandem using anti-Flag resin followed by anti-HA resin. The resulting two-step immunoprecipitate was resolved on an SDS-PAGE gel and proteins were identified by MS. We employed a similar approach whereby HEK293T cells were transfected with *3x-flag ankB* (or *3x-flag* vector as a control) and *HA-ubiquitin* and treated with MG132 to inhibit the proteasome. Treating cells with MG132 can result in the stabilization of substrate-F-box protein complexes within the cell, which helps facilitate isolation of ubiquitylated substrates [140]. Ubiquitylated AnkB substrates were purified first using anti-Flag resin followed by anti-HA resin. The resulting two-step immunoprecipiate was resolved on a SDS-PAGE gel and proteins were identified by MS.

Three distinct bands on the gel unique to the AnkB condition were identified as potential substrates (Fig. 6). The entire list of proteins identified in all the bands is shown in Appendix 3, while a list of proteins identified in at least 2 of 3 replicates is shown in Table 2 (numbers indicate total spectrum counts). Bands 1 and 2 contained the E3 ubiquitin ligase HUWE1 as well as ubiquitin

itself (Table 2 and Appendix 3). Region 3 contained no proteins identified from at least 2 replicates. Region 4 contained peptides derived from AnkB and ubiquitin which likely represents ubiquitylated AnkB based on the apparent molecular weight and the fact that ubiquitylation of AnkB has been described previously [141]. To confirm HUWE1 as an AnkB substrate, *ha-huwe1* and *3x-flag ankB* (or *3x-flag* vector as a control) were co-transfected into HEK293T cells and immunoprecipitated with anti-Flag magnetic beads. The resulting immunoprecipitants were subjected to Western blotting using anti-HA antibodies. As shown in Figure 7, HA-HUWE1 expressed well but failed to bind AnkB. Since HUWE1 was never detected in the control sample among any of the three replicates (Table 2), it is unlikely to non-specifically bind the anti-Flag resin. As the confirmatory co-IP was done using fewer cells than the initial two-step enrichment, it is possible that the binding between AnkB and HUWE1 is transient enough to not be detectable with fewer cells.

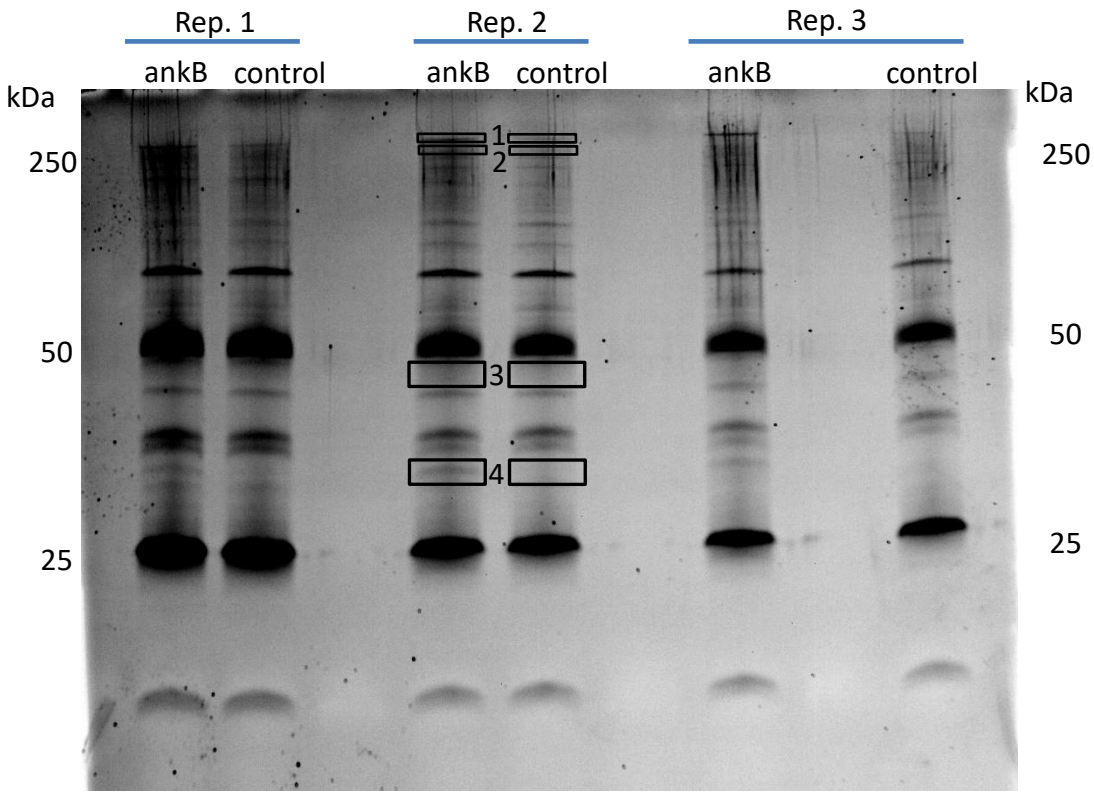

Fig. 6. SDS-PAGE Gel of Two-Step Immunoprecipitation. *p3x-flag ankB* or *p3x* vector was co-transfected with *ha-ubiquitin* into HEK293T cells. AnkB was purified using anti-Flag resin and elutions were subjected to a second IP with anti-HA resin. Elutions from anti-HA resin were ran on an SDS-PAGE gel and stained with Sypro Ruby. Numbers indicate regions that were excised from the gel and identified by MS. Three replicates were performed.

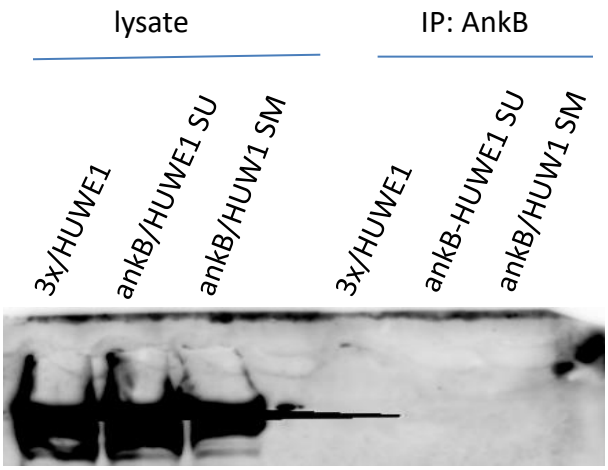

Fig. 7. Co-IP of HUWE1 and AnkB. *p3x-flag ankB* or *p3x* vector was co-transfected with *ha-huwe1* into HEK293T cells. AnkB was immunoprecipitated with anti-Flag resin. Elutions were subjected to Western blotting with anti-HA antibodies. SU indicates same µg of DNA used for co-transfection and SM indicates same number of moles.

Table 2. Proteins Appearing in Two or More Replicates from Two-Step Immunoprecipitation (spectral counts are shown)

Protein	MW kDa	AnkB			Control		
Band 1		**Rep 1**	**Rep 2**	**Rep 3**	**Rep 1**	**Rep 2**	**Rep 3**
Isoform 2 of E3 ubiquitin-protein ligase HUWE1	480	2	7				
Polyubiquitin-B	26	1	2		2	2	
Desmoglein-1	114	1	1		2	4	3
Serum albumin	69	5		2			
Immunoglobulin heavy constant alpha 1	38	2		1			
Band 2							
Isoform 2 of E3 ubiquitin-protein ligase HUWE1	480	4	4	12			
Ubiquitin-40S ribosomal protein S27a	18	3	8	3			1
Desmoglein-1	114	3	5	1			
Band 3							
AnkB Legionella pneumophila	18	4	6	6			
Ubiquitin-60S ribosomal protein L40	15	3	4	3	1	1	
40S ribosomal protein S3	27	2	4		1	3	
Desmoglein-1	114	1		10	4		

AnkB Substrate Identification by Formaldehyde Cross-linking *in vivo*

As a final attempt to identify substrates of AnkB during infection, we modified a method previously described for identification of host protein substrates of *L. pneumophila* effectors [127, 142]. This method takes advantage of formaldehyde cross-linking to stabilize transient protein-protein interactions and uses a tandem affinity purification of cross-linked effector complexes. The effector of interest is cloned into the pICC1544 plasmid which contains a tag with two His-6 epitopes and an *E. coli* biotinylation site. *L. pneumophila* strains expressing this fusion construct are used to infect A549 GFP-BirA cells that express the *E. coli* BirA biotin ligase enzyme. When the effector is translocated, it is recognized and biotinylated by the BirA ligase expressed in the host cell. Effector-host protein complexes are stabilized via formaldehyde cross-linking and purified first on nickel and then on streptavidin. The negative control for this experiment contains a mutation in the biotinylation site such that it no longer becomes biotinylated during infection. In this way, effector complexes containing bound host protein targets can be highly purified and the associated host proteins identified by MS.

AnkB was expressed well from the pICC1544 plasmid (Fig. 8). In the initial experiment, 1.2×10^8 A549 GFP-BirA cells per condition were infected with *L. pneumophila* expressing *pBio-ankB* or *pK/A-ankB* (control). A total of 115 and 116 host proteins were identified in the Bio and K/A conditions, respectively. AnkB was not identified in either condition indicating that the proteins identified are contaminants rather than true AnkB interactors. Samples from this

experiment were run on an SDS-PAGE gel and subjected to Western blotting with antibodies to AnkB. As shown in Figure 9, AnkB was undetectable in lysates prepared from cells infected with *L. pneumophila* expressing either *pBio-ankB* or *pK/A-ankB*. However, AnkB was detected after nickel purification in both conditions. Samples from the unbound fraction after streptavidin purification did not show the expected depletion of biotinylated AnkB, which indicates either most of the purified AnkB was bacterially derived or failed to become biotinylated upon translocation into the host cell. Since it is possible that more infected cells are required for detection of AnkB by MS, a second experiment was done using 2.3×10^8 A549 GFP-BirA cells. This experiment also failed to detect AnkB.

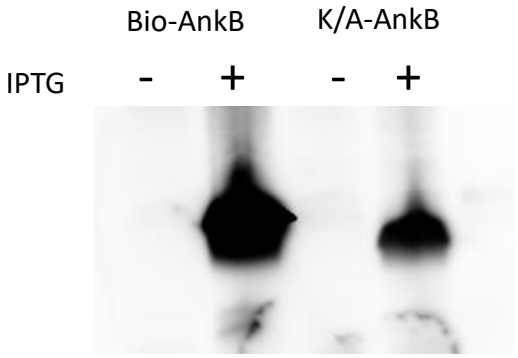

Fig. 8. Expression of Bio-AnkB and K/A-AnkB. *L. pneumophila* harboring *pBio-ankB* or *pK/A-ankB* was grown to an OD550 of 1.5 and induced with 1 mM IPTG for 1 hour. 1 x 10^8 bacteria were lysed in SDS-PAGE loading buffer and ran on an SDS-PAGE gel. AnkB was identified by Western blot using anti-AnkB.

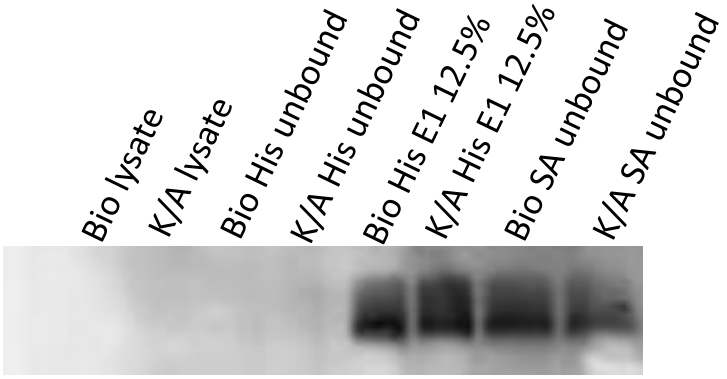

Fig. 9. Purification of Cross-Linked AnkB-Host Protein Complexes. AnkB complexes were purified from infected A549 GFP-BirA cells first using nickel resin and then streptavidin beads. Bio indicates AnkB purified from A549 cells infected with *L. pneumophila* expressing His6-Biotin-tagged AnkB. K/A indicates the negative control which lacks the biotinylation site. His unbound indicates a sample taken from depleted lysate after His purification. His E1 indicates the first elution from the His purification. SA unbound indicates a sample taken from the material left after SA purification. Blots were probed with anti-AnkB.

Discussion

AnkB is one of the few of nearly 350 *L. pneumophila* effectors that is required for robust intracellular proliferation [71]. It is an F-box protein that substitutes for the host F-box component of the SCF E3 ubiquitin ligase complex and mediates the accumulation of K^{48}-linked polyubiquitylated proteins around the LCV [99, 120]. The primary purpose of this process is to generate a surplus of free host amino acids that drive intracellular replication [120]. AnkB may target proteins rich in certain amino acids for which *L. pneumophila* is auxotrophic, however, AnkB may target proteins in specific cellular pathways to modulate specific host processes. The identity of the host proteins targeted by AnkB is currently unknown, and their identification likely provide more insight into the role of AnkB during infection. We utilized three approaches to identify host proteins targeted for ubiquitylation by AnkB.

Our first approach employed an *in vitro* pull-down using AnkB (with the F-box deleted) as bait and cellular lysates from two different cell lines as prey. We identified 13 potential interacting partners and tested each by co-IP. None of these proteins was confirmed as a true specific interacting partner of AnkB. There are several possible explainations for the failure of this approach. First, there was a high background in the in vitro pull down (Fig. 2) possibly owing to less than ideal washing conditions. It is possible that true interacting partners were masked by high abundance contaminating proteins which were not adequately washed away. Second, AnkB was recently shown to be post-translationally modified by hydroxylation within the ankyrin domains [143]. Since

the AnkB protein used for the in vitro pull-down was purified from *E. coli*, it did not have this post-translational modification, which could explain why we were unable to identify true AnkB interacting partners. Finally, it is possible that AnkB may require *L. pneumophila* metaeffectors, which are effectors that control the activity of other effectors, to properly bind and/or ubiquitylate its target proteins. Metaeffectors that target other *L. pneumophila* effectors for ubiquitylation have been described previously [100].

Our second approach utilized a two-step IP in cells expressing AnkB and HA-ubiquitin. Cells were treated with the proteasome inhibitor, MG132. From cellular lysates, AnkB and bound HA-ubiquitin tagged substrates were immunoprecipitated. In step two, the HA-ubiquitin tagged substrates were enriched via anti-HA affinity purification and the results were identified by MS. We identified several peptides for ubiquitin as well as the E3 ubiquitin ligase, HUWE1. Our follow-up co-IP in cells expressing Flag-AnkB and HA-HUWE1 failed to confirm this protein as a true interactor. Reasons may include that AnkB requires specific modifications to its substrates that only occur in the context of an infection. In addition, AnkB-substrate interactions may be too transient and low affinity to be identified in a co-IP experiment where the percentage of transfected cells is only 20-25%. Finally, *L. pneumophila* metaeffectors may be required to observe interaction with or ubiquitylation of AnkB substrates.

Our final approach was to perform the screen for AnkB targets in the context of an infection and stabilize protein interactions using formaldehyde cross-linking. This approach also utilized a two-step enrichment strategy that

included a biotinylated tag to ensure only AnkB that was translocated into the host cell would be purified. Despite using 2.3 x 10^8 infected cells, we failed to purify detectible amounts of AnkB and did not identify any candidate substrates. There are several potential explanations. It is possible that cross-linked AnkB-host protein complexes were insoluble under these conditions. The time of infection may impact if AnkB substrates are available for detection by this method. Thus, successful purification of AnkB substrates may require sampling at different time points during the infection. Since AnkB substrates are K^{48}-linked and targeted for degradation, their isolation and identification may require treating infected cells with a proteasome inhibitor. In addition, although we verified that the tagged version of AnkB is expressed well, translocation efficiency of the tagged protein into the host cell may not be optimal. Finally, the N-terminal tag on AnkB may be masked by the host SCF complex cross-linked to the F-box domain requiring purification under denaturing conditions.

There are two technical considerations that may have impacted two of the three approaches we utilized in our search for AnkB targets. First, it was noted during the crystallization studies that AnkB is insoluble when expressed in *E. coli* without its host binding partner, Skp1. As a result, we used the ankyrin domains of AnkB, which are soluble, as bait for the *in vitro* pull down. It is possible that the entire protein and/or its interacting parter, Skp1, are needed to bind substrate proteins. If this is the case, we would not have been able to detect AnkB targets using the *in vitro* pull down unless we modified our approach by using purified AnkB/Skp1 complex as bait. Second, if binding of AnkB to Skp1 induces a

conformational change that allows AnkB to bind substrates, both the *in vitro* pull down as well as the two-step IP approach may have been affected. Although we overexpressed AnkB in HEK293T cells for the two-step IP, there may not have been enough endogenous Skp1 protein to promote enough stable AnkB-Skp1-target protein complexes for successful purification. Although our first two approaches were subject to these limitations, our third approach, being done in the context of an infection, should have been unaffected by these technical consdierations.

CHAPTER 2

DIVERGENT EVOLUTION OF ER RETENTION VS FARNESYLATION MOTIF-MEDIATED ANCHORING OF THE ANKB VIRULENCE EFFECTOR TO THE *LEGIONELLA*-CONTAINING VACUOLUAR MEMBRANE[1]

Results

Episodic Positive Selection in *ankB* Evolution

Compared to strain AA100/130b, the *ankB* gene of the Paris strain (*ankB*-Paris) has a deletion of an adenine at position 450 (ΔA450), which resulted in a frame shift mutation (Fig. 10). This has led to a truncation of the last 18 amino acids that included the CaaX farnesylation motif, which is essential for anchoring AnkB-AA100/130b into the LCV membrane, which is indispensable for its biologic function in decorating the LCV with polyubiquitinated proteins [110, 144]. Despite this frame shift mutation and deletion of the C-terminal CaaX farnesylation motif, AnkB-Paris is required for decoration of the LCV with polyubiquitinated proteins [98]. Concurrently, a unique NKYAP sequence motif is generated at amino acids 150-154 in AnkB-Paris. To determine whether this frame shift mutation was unique to the Paris strain or more widespread among other *Legionella* isolates,

[1] The contents of this chapter are from "Divergent evolution of Di-lysine ER retention vs. farnesylation motif-mediated anchoring of the AnkB virulence effector to the Legionella-containing vacuolar membrane", by J.D. Perpich, et al, 2017*, Sci Rep*, *7(1)*, p. 5123. Copyright 2017 by Springer Nature. Reprinted with permission.

we examined the abundance of the ΔA450 mutation in fifty-one isolates of clinical (N=25) and environmental (N=26) origin. Analysis of full-length *ankB* sequences revealed 15 distinct *ankB* alleles (Fig. 10b; Fig. 11). Interestingly, among the 15*ankB* alleles, *ankB1* (Paris strain) was the only *ankB* allele to harbor the ΔA450 mutation (Fig. 10b). The data showed that 19 of 51 isolates (37.25%), of which 17 were environmental and 2 were of clinical origin, contained *ankB1* allele (*ank1/ankB-Paris*). We conclude that the *ankB1/ankB-Paris* allele is wide-spread and is predominant among environmental isolates. To better understand the forces that shaped *ankB* evolution, in particular the maintenance and spread of the *ankB1* allele in environmental *Legionella* isolates, we next analyzed the selective pressures acting on *ankB* codons and also on the *ankB1* branch using CODEML and a variety of other methods (materials and methods). We found preponderance of sites that were constrained by either negative selection or evolved neutrally. Comparisons of models M7 and M8 also suggested that at least 4 *ankB* codons had been subject to positive selection. However, we also found evidence of recombination in the *ankB* alignment (Fig. 12). Although the M7 vs M8 comparison should be robust to model violations introduced due to recombination, we sought additional evidence to verify positive selection using REL, FEL, IFEL and MEME methods (materials and methods). Each of these methods found significant statistical support for site-specific positive selection in *ankB* codons (data not shown). Branch site tests implemented in CODEML, and the GA branch test both provided statistical support for the hypotheses that the internal branch leading up to *ankB1/ankB-Paris* and its branch both had

experienced positive selection. Moreover, the NKYAP C-terminal motif itself was identified as the target of positive selection (Fig. 13).

a

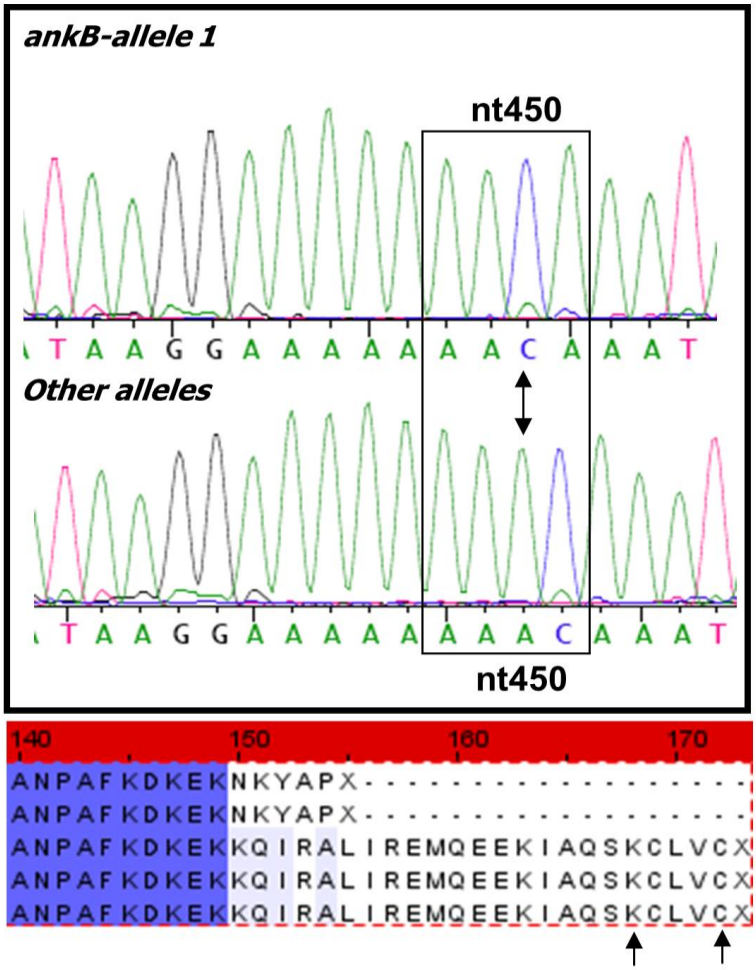

Fig. 10a: **Molecular evolution of *ankB*.** Representative sequence chromatograms indicating the frame shift mutation at nucleotide position 450 (*ΔA*) in *ankB1* compared to other *ankB* alleles. This mutation in *ankB*1 alters the reading frame and predicts a prematurely terminated AnkB protein at residue 154 and generation of a unique NKYAP sequence motif. Arrows mark the location of the Dot-Icm translocation signal.

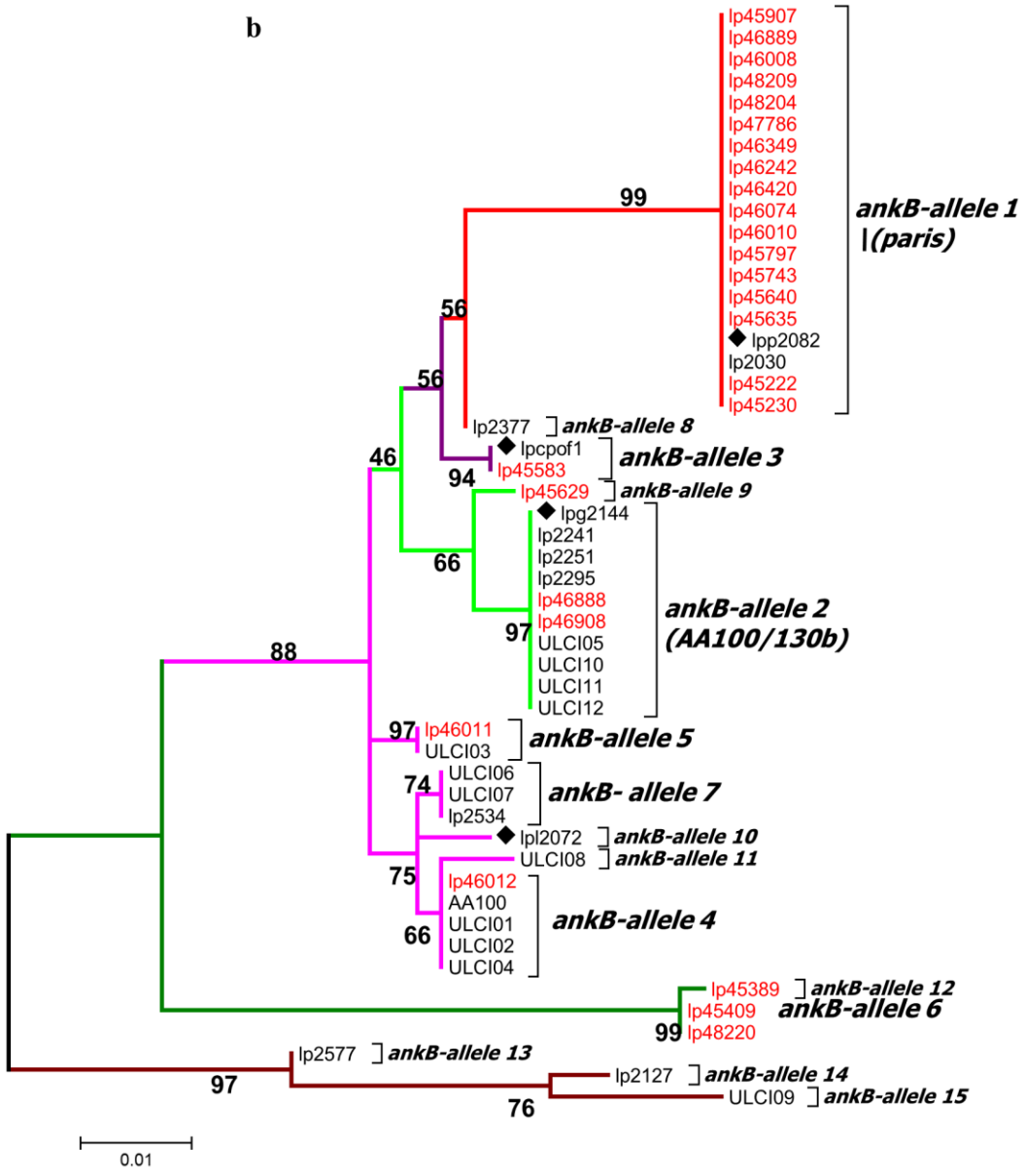

Fig. 10b. ML analysis of *ankB* alleles variously from clinical (in black text) and environmental (in red text) isolates. All alleles are indicated. Diamonds indicate *ankB* sequence from strains whose complete genomes has been determined. Bootstrap values are shown above the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

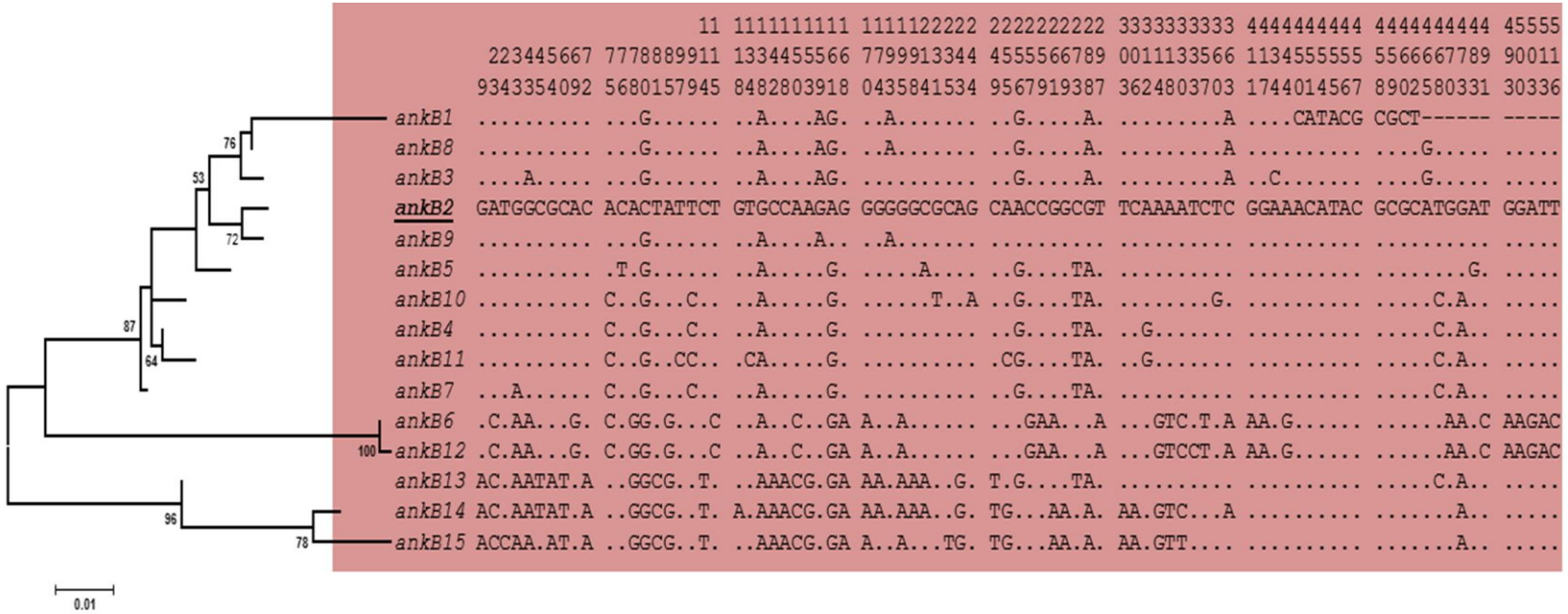

Fig. 11. Sequence polymorphism and divergence in *ankB* alleles. Polymorphisms unique to each *ankB* allele are shown. Site numbers for each variable position is shown in a vertical format. *ankB2* allele sequence was used as a reference. This analysis was done using MEGA 6.

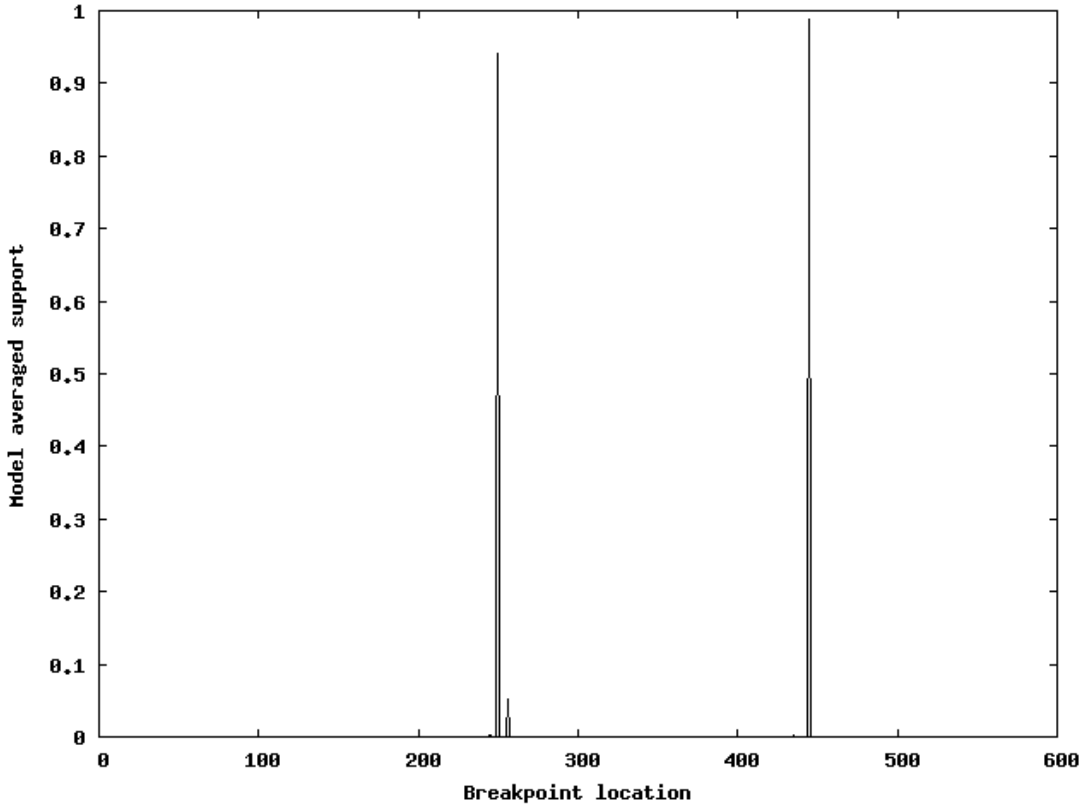

Fig. 12. Two recombination breakpoints were identified in *ankB* alignment. Statistical support for each breakpoint is shown. Breakpoint at position 250 was also significant for a Kishino Hasegawa test indicating the presence of 2 or more incongruent topologies for *ankB* alignment. This analysis was done using GARD.

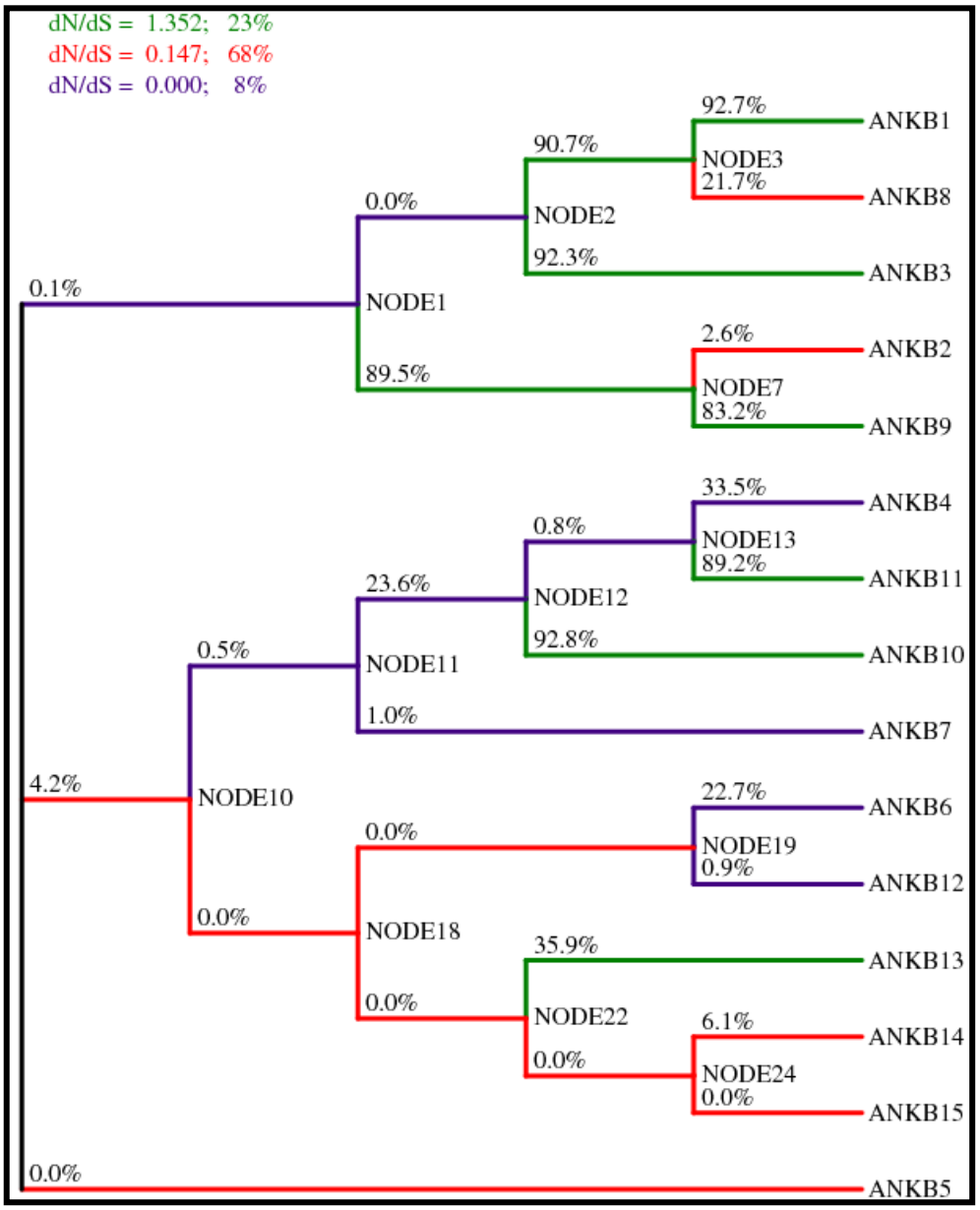

Fig. 13. *ankB* lineages have experienced variable selective pressures. An unscaled *ankB* phylogeny indicating the associated selective pressures for each branch is shown. This GA Branch analysis indicates a 92.7% and 90.7% support for positive selection on the *ankB1* branch and the internal node leading up to *ankB1*, respectively.

Decoration of the LCV with Polyubiquitinated Proteins Independent of the *ankB* Genotype

We next asked whether the altered C-terminus of AnkB variant encoded by *ankB*1 (Fig. 14a) either modified or significantly reduced the ability of AnkB1 strains to recruit polyubiquitinated proteins to the LCV. Overall, we found extensive variation in the ability of 23 distinct *Legionella* isolates to recruit polyubiquitinated proteins to the LCV (Fig. 14a-14b). Specifically, among AnkB1 strains, polyubiquitination varied from 30% to 54%. In contrast, six isolates with full-length *ankB* alleles were either similar to the Δ*ankB* mutant or the *ΔdotA* translocation-defective mutant in their ability to recruit polyubiquitinated proteins to the LCV (Fig. 14a-14b). However, a comparison of environmental and clinical isolates revealed a modest, but statistically significant difference (Student *t*-test, $P < 0.05$), in their ability to recruit polyubiquitinated proteins (Fig. 15). Thus, while the environmental isolates seem less capable than clinical isolates to recruit polyubiquitinated proteins to the LCV, the ability to recruit polyubiquitinated proteins itself seems independent of the *ankB* genotype.

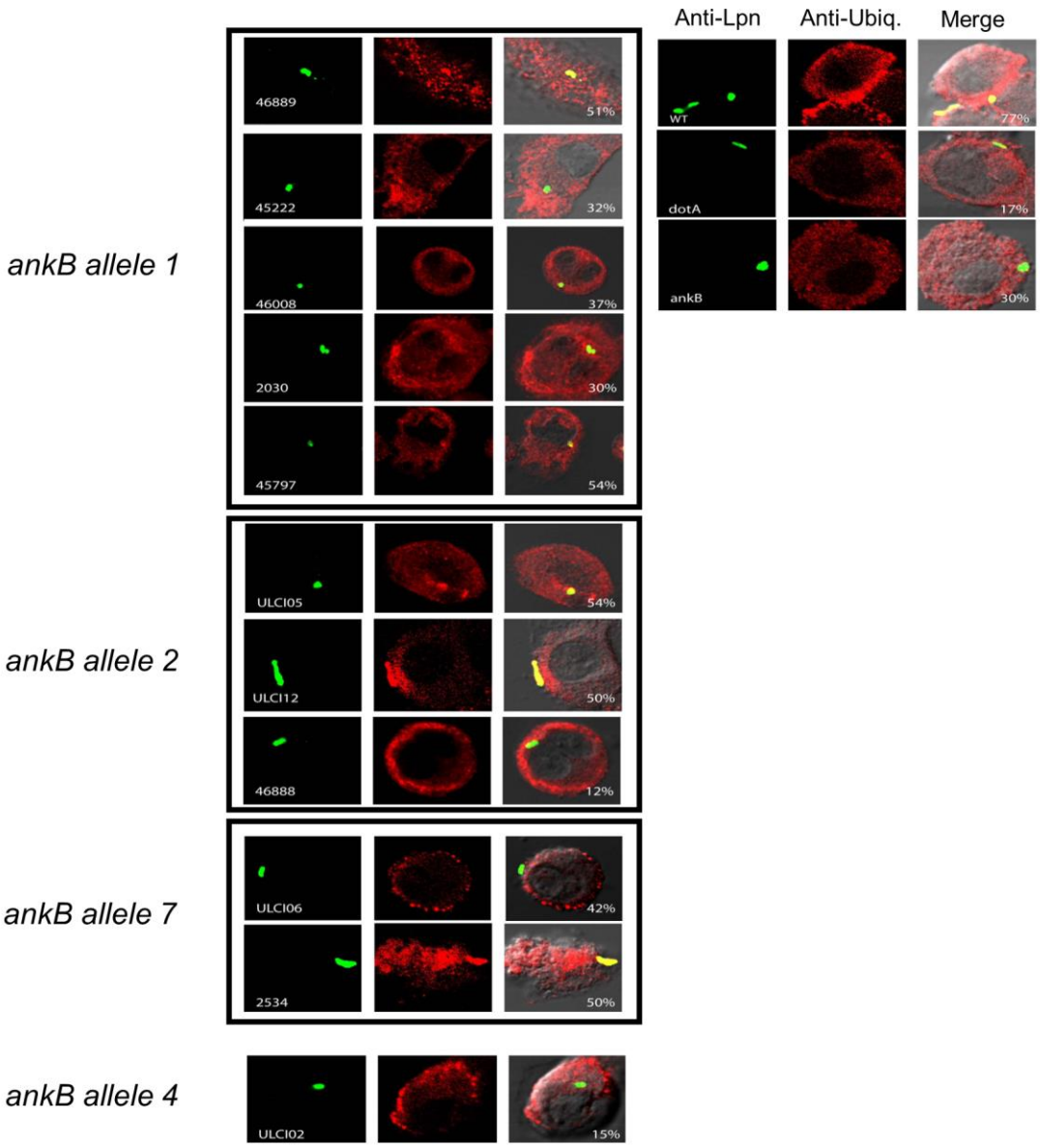

Fig. 14a. AnkB-genotype does not predict the ability to recruit polyubiquitinated proteins to the LCV. Representative confocal microscopy images of polyubiquitinated protein recruitment to the LCV among 23 different isolates of *L. pneumophila* expressing various *ankB* alleles. Percentages indicate the number of LCVs positive for ubiquitin. Images in the right panel were taken using the AA100 strain.

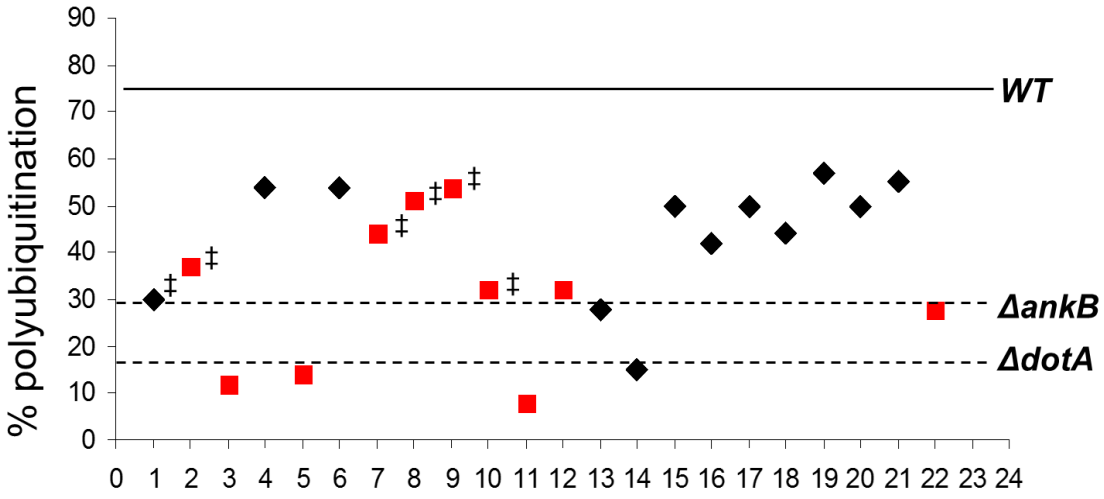

Fig 14b. Distribution of percent polyubiquitin recruitment among 23 different isolates of *L. pneumophila.* Environmental isolates are shown in red and clinical isolates are shown in black. ‡, indicates strains carrying the *ankB1* allele.

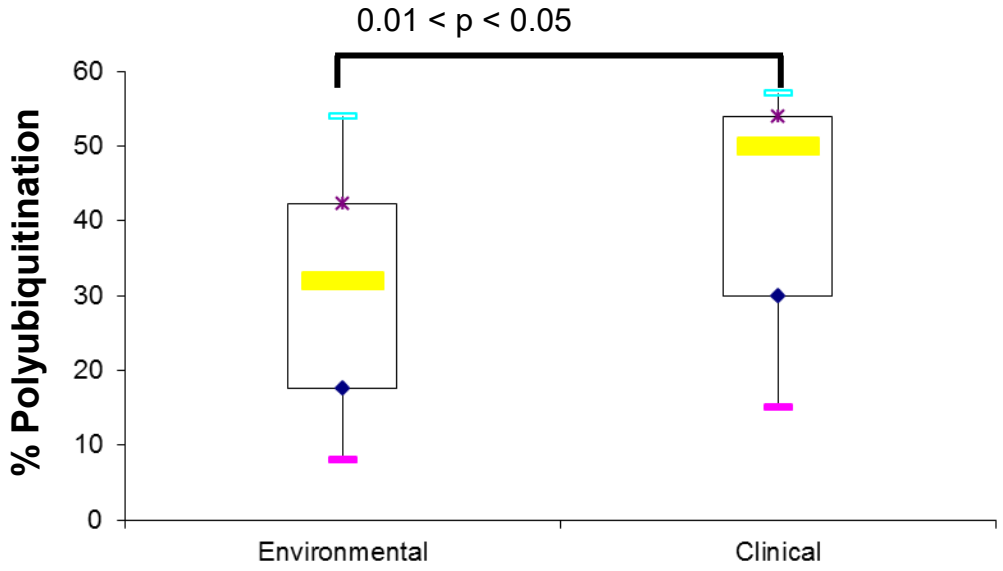

Fig. 15. Environmental and clinical isolates showed modest, but significant, difference in their ability to recruit polyubiquitinated proteins. Box-plots show % polyubquitination (ability to recruit polyubiquitinated proteins to the LCV) among 10 environmental and 22 clinical isolates. Statistical significance was analyzed via a two tailed Mann-Whitney nonparametric test. The Z-Score was 2.0111; the p-value is 0.04444. The critical value of U at p≤ 0.05 was 29. Therefore, the difference between clinical and environmental isolates is significant at p≤ 0.05.

Localization of AnkB-Paris to the LCV membrane

Ectopic expression of AnkB-AA100/130b within amoeba and HEK293T cells results in farnesylation-mediated anchoring to the plasma membrane of both evolutionarily distinct host cells [110, 111, 144, 145]. In contrast to AnkB-AA100/130b, AnkB-Paris ectopically expressed within A549 cells is enriched at the leading edge of lamellipodium formation and co-localizes with α-actinin [98]. Our data showed that in contrast to AnkB-AA100/130b (Fig. 16a), when AnkB-Paris is ectopically expressed in HEK293T cells, it did not localize to the plasma membrane; but it exhibited a punctate appearance and a perinuclear distribution, which is characteristic of sub-cellular localization of the ER (Fig. 16c). This perinuclear distribution is distinct from the diffuse cytosolic pattern characteristic of AnkB-C169A, which lacks the farnesylation motif (Fig. 16b). Interestingly, mutation of both lysines within the di-lysine motif of AnkB-Paris to arginine (AnkB-Paris $K^{149,151R}$) seemed to alter the distribution upon ectopic expression (Fig. 16d). Ectopically expressed AnkB-Paris tended to localize to the perinuclear ER region while AnkB-Paris^{K149,151R} tended to distribute throughout the cytoplasm, suggesting the loss of ER localization mediated by the ER retention motif (Fig. 16). Our data clearly show a distinct sub-cellular localization of AnkB-Paris and AnkB-AA100/130b.

While AnkB-AA100/130b is localized to the LCV membrane during infection by host-mediated farnesylation, sub-cellular location of AnkB-Paris during infection is not known. Since the farnesylation motif is missing from AnkB-Paris, we set out to determine sub-cellular localization of AnkB-Paris during

infection of hMDMs. We created an identical *ankB-Paris* allele and introduced it into the isogenic *ankB* null mutant of strain AA100/130b to determine its potential anchoring to the LCV membrane despite the lack of the farnesylation motif. At 2 hours post-infection, the LCVs were semi-purified from infected hMDMs. Prior to their permeabilization, the LCVs were labeled with anti-AnkB antibodies to detect AnkB on the cytosolic side of the LCV membrane, as we described previously [110]. Analyses by confocal microscopy showed that AnkB-AA100/130b was anchored to the cytosolic side of the LCV membrane of 80% of WT strain-containing LCVs (Fig. 17a-17b). As expected, complementation of the *ankB* mutant of strain AA100/130b with *ankB*-AA100/130b restored localization of AnkB to the LCV membrane similar to the wild type strain [110]. Interestingly, despite the lack of the farnesylation motif, complementation of the AA100/130b-derived *ankB* mutant with the *ankB-Paris* allele resulted in anchoring AnkB-Paris to the LCV membrane, similar to AnkB-AA100/130b. Therefore, despite lacking the farnesylation motif, which is indispensable for anchoring AnkB-AA00/130b to the ER-derived LCV membrane, AnkB-Paris is also anchored to the LCV membrane, and this is likely to be mediated the di-lysine ER retention motif.

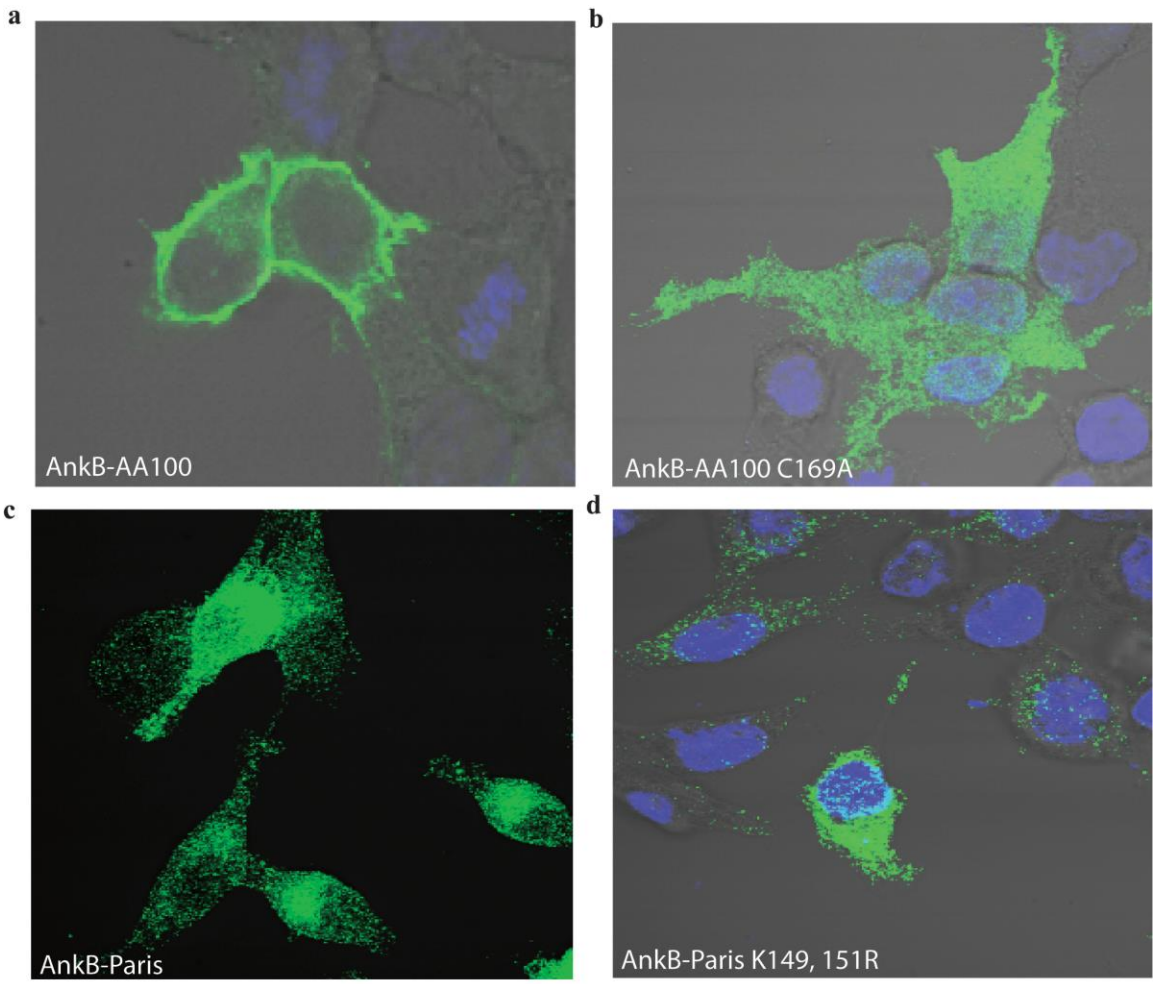

Fig. 16. Ectopically expressed AnkB-Paris localizes to the cytoplasm with a perinuclear distribution. a-d. Localization of AnkB-AA100, AnkB-Paris, AnkB-Paris $K^{149,151}R$, and AnkB-C169A in HEK293T cells transiently transfected with 3X Flag-tagged versions of each and stained with anti-flag antibodies and DAPI. Representative confocal images are shown.

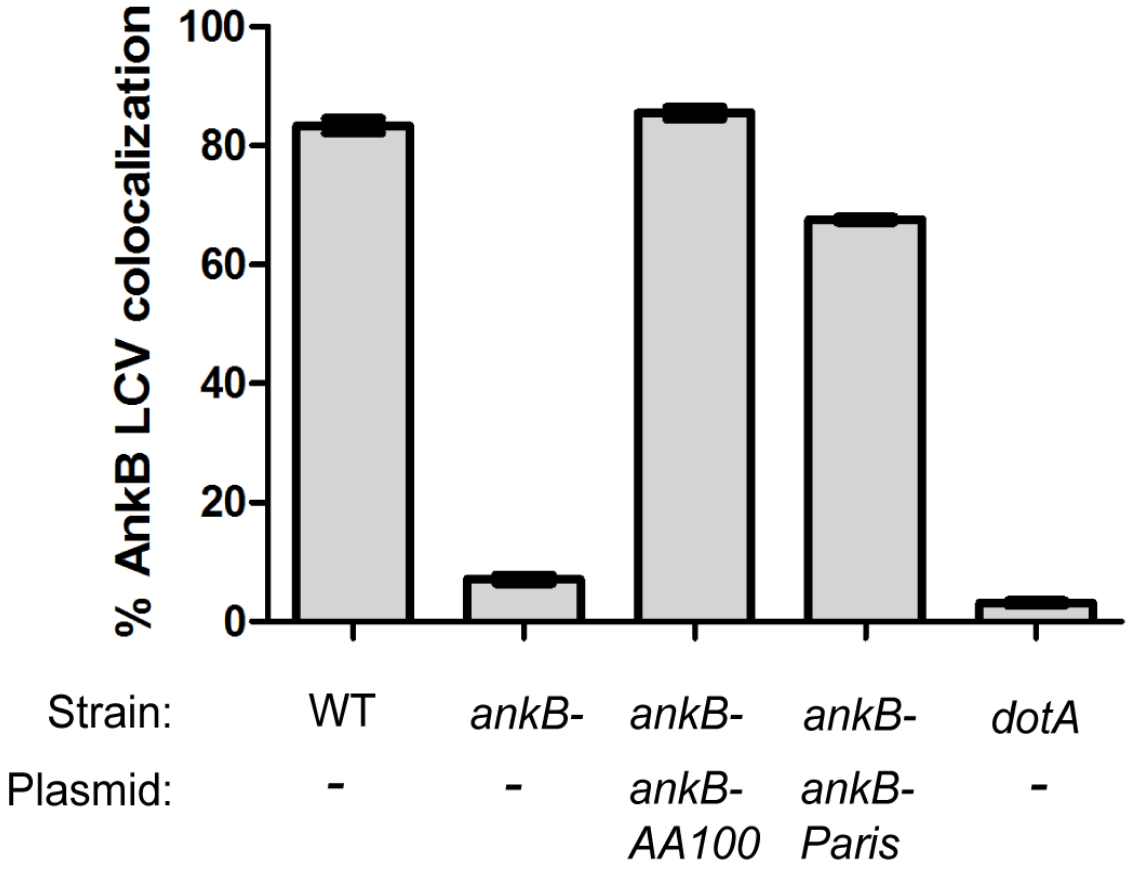

Fig. 17a. AnkB-Paris localizes to the LCV during infection. Macrophages infected with WT *L. pneumophila* AA100/130b, *ankB* mutant, or the *ankB* mutant complemented with either *ankB* from AA100/130b or *ankB* from Paris strain were fixed at 2 hours post-infection and stained with antibodies to AnkB and *Legionella* (Lpn). The percentage of bacteria staining positive for AnkB (mean ±1SD) was determined by analysis of 100 infected cells in triplicate. Data are representative of 2 independent experiments.

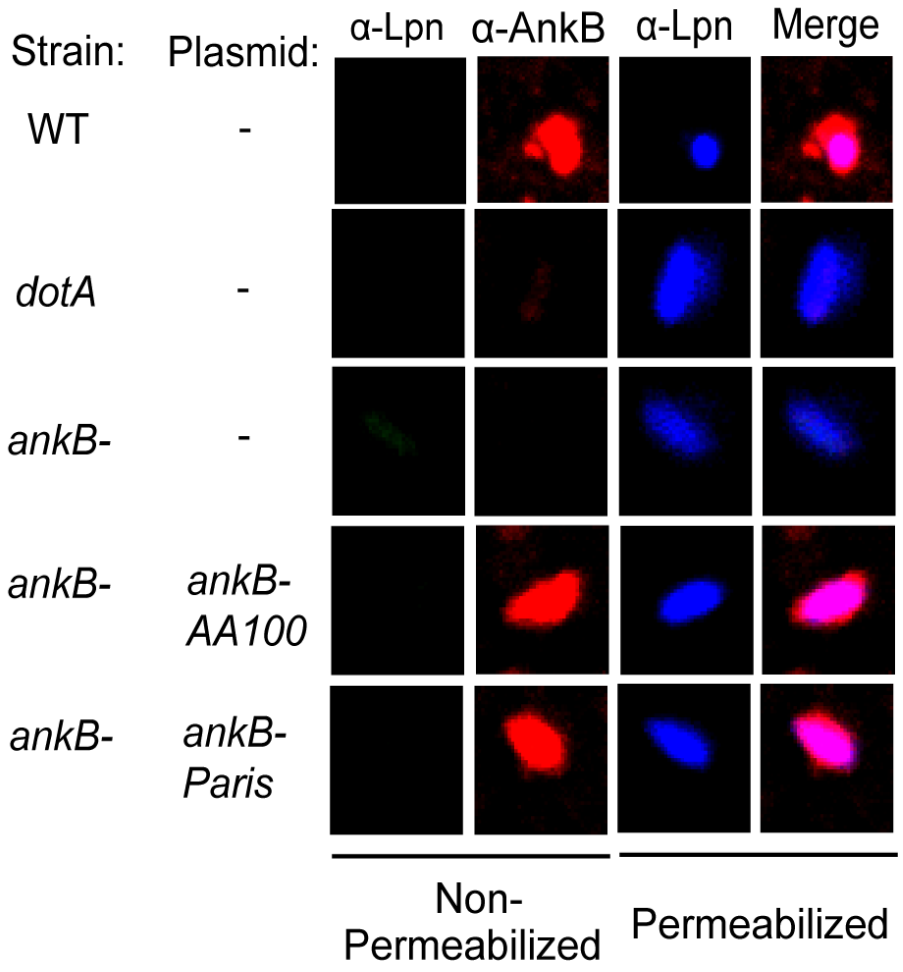

Fig. 17b. AnkB-Paris localizes to the LCV during infection. Representative confocal images of LCVs isolated from macrophages infected with the indicated strains. To differentiate between intact LCVs and extracellular bacteria, the LCVs were labeled prior to permeabilization with mouse anti-Lpn antisera and rabbit anti-AnkB antisera for 1h. LCVs were then permeabilized with -20°C methanol and counter-labeled with goat anti-Lpn antisera to detect intact LCVs. Abbreviations: *ankB-* (*ankB* null mutant in AA100/130b strain), *dotA (dotA* null mutant in AA100/130b strain), WT (wild type AA100/130b strain). Plasmids indicate the *ankB* allele used to complement the indicated strain.

Functional substitution of AnkB-AA100/130b by AnkB-Paris

Compared to AnkB-AA100/130b, the AnkB-Paris has a truncation of the last 18 amino acids. The crystal structure of AnkB indicates that the C-terminal truncation of AnkB-Paris eliminates a large portion of the third ankyrin repeat compared to AnkB-AA100/130b (Fig. 18) [146]. To determine if AnkB-Paris can functionally substitute for AnkB-AA100/130b strain, we complemented the *ankB* null mutant of the AA100/130b strain with the *ankB-Paris* allele and assessed intracellular replication and decoration of the LCV with polyubiquitinated proteins within human monocytes-derived macrophages (hMDMs) [110, 144]. We assessed polyubiquitination of the vacuole at 2 hours post-infection by confocal microscopy. The data showed that 20% of *ankB* mutant-containing LCVs were decorated with polyubiquitinated proteins (Fig. 19a). In contrast, approximately 75% of strain AA100/130b-containing LCVs were decorated with polyubiquitinated proteins. Despite truncation of the third ankyrin domain, the *ankB-Paris* allele fully complemented the AA100/130b isogenic *ankB* null mutant for accumulation of polyubiquitinated proteins, similar to the wild type strain.

We assessed the ability of AnkB-Paris to restore intracellular replication of the *ankB* mutant of strain AA100/130b by determination of the frequency of formation of replicative vacuoles at 10 hours post-infection, by confocal microscopy (Fig. 19b). The majority of cells infected with the *ankB* mutant contained a single bacterium. In contrast, the majority of cells infected with WT bacteria contained 2-4 bacteria per cell and >20% of the LCVs harbored more than 5 bacteria per cell. Similarly, the *ankB* mutant complemented with the

ankB-Paris allele formed replicative vacuoles at a frequency similar to the WT strain (Fig. 19b). Taken together, these results indicate that AnkB-Paris is functionally equivalent to AnkB-AA100/130b in its ability to decorate the LCV with polyubiquitinated proteins and to power intracellular proliferation of *L. pneumophila*.

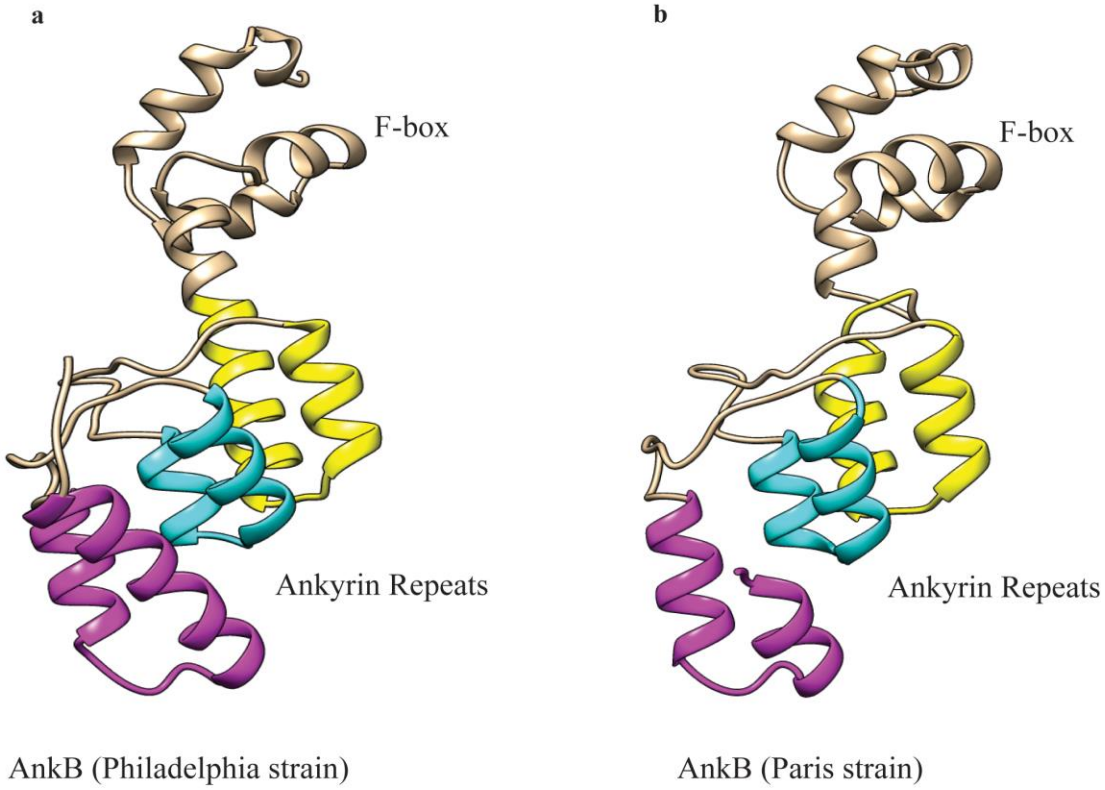

Fig. 18. The crystal structure of AnkB. a. The structure of AnkB from the Philadelphia strain spanning from Lys2 to Ala168 [146]. The F-box domain near the N-terminus is indicated. Three ankyrin repeats are present rather than the predicted two repeats based on the sequence (yellow, cyan, and magenta). **b.** Predicted structure of AnkB-Paris, which maintains all three ankyrin repeats by keeping the last half of the last repeat (magenta).

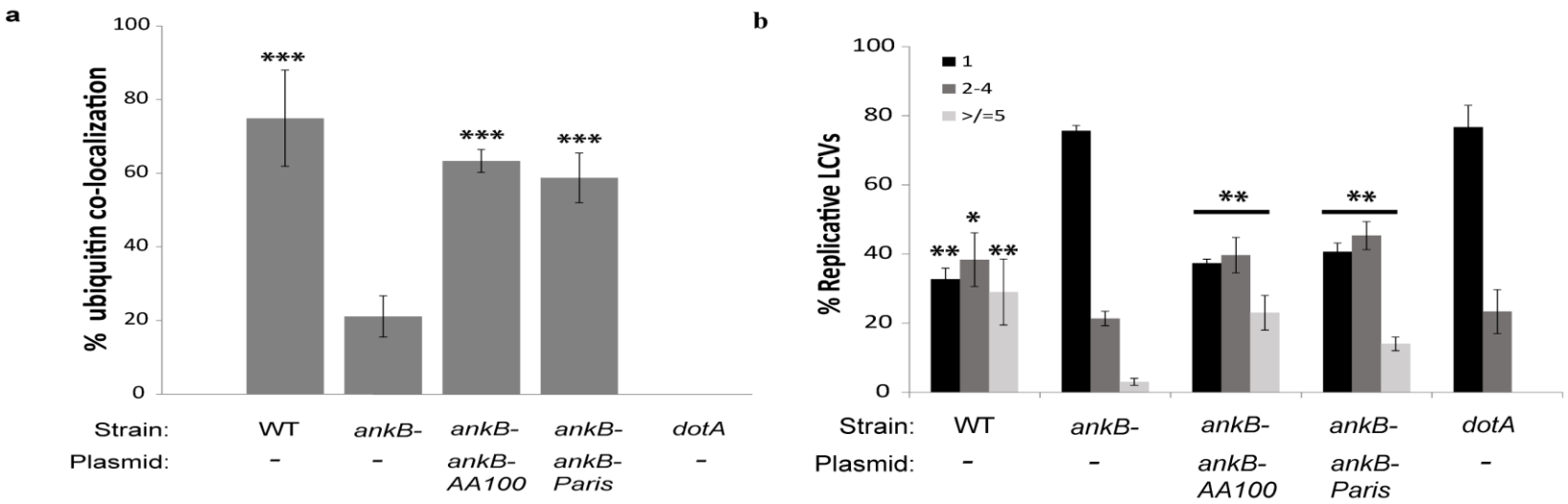

Fig. 19. AnkB-Paris complements the *ankB* mutant of strain AA100/130b. a. Co-localization of polyubiquitinated proteins with the LCV at 2 hours post-infection of hMDMs. Macrophages were infected with either wild type *L. pneumophila* strain AA100/130b, its isogenic *ankB* or *dotA* mutant, or the *ankB* mutant complemented with *ankB-Paris* or *ankB-AA100*. Numbers indicate the percentage of LCVs (±1SD) that co-localize with polyubiquitinated proteins. The data are based on analysis of 100 infected cells performed in triplicate and are representative of three independent experiments. **b.** At 10 hours post-infection hMDMs were fixed, stained with anti-Lpn, and analyzed by confocal microscopy. The number of bacteria per cell was determined and the data are based on analysis of 100 infected cells (mean ±1SD) performed in triplicate and are representative of three independent experiments. *p<0.05, **p<0.01, ***p<0.001 compared to corresponding value for *ankB* null mutant.

An indispensable role for the C-terminal Di-lysine Motif of AnkB-Paris in biological function

The C-terminal 5 residues of AnkB strain Paris ($K^{149}NK^{151}YAP$) resemble a eukaryotic di-lysine motif (KxKxx) responsible for ER-to-golgi retrograde protein trafficking and retention in the ER [113-115, 147]. The AnkB-AA100/130b is anchored to the LCV membrane by host-mediated farnesylation, which is essential for biological function. Therefore, we tested the hypotheses that the generated di-lysine ER-retention motif is also required for biological function of AnkB-Paris. We constructed single and double substitutions of lysine^{149} and lysine^{151} in the *ankB-Paris* allele with arginine. Since *L. pneumophila* effectors often have Dot/Icm translocation signals encoded in their C-terminus, we tested the ER retention motif AnkB-Paris^{K149R} and AnkB-Paris^{K151R} substitution mutants for Dot/Icm-mediated translocation using an adenylate cyclase reporter assay, as we described previously [71, 99, 110]. The *cya* reporter fusions of *ankB-Paris,* *ankB-Paris*^{K149R}, or *ankB-Paris*^{K151R} were transformed into either the WT strain AA100/130b or its isogenic translocation-deficient *dotA* mutant. After 2 hours of infection, cells were lysed and cAMP levels were determined via ELISA (Fig. 20a-20b). Cells infected with WT bacteria expressing the Cya-AnkB-Paris reporter fusion showed robust cAMP production compared to cells infected with *dotA* mutant bacteria expressing the same reporter fusion or cells infected with the WT strain expressing the catalytic domain of Cya alone. This indicates that AnkB-Paris is translocated by the AA100/130b strain. In contrast, substitution of either K^{149} or K^{151} completely abolished translocation of AnkB-Paris. These

results indicate that the two lysine residues in the putative di-lysine ER-retention motif are essential for translocation of AnkB-Paris during infection.

In eukaryotic cells, the di-lysine motif is recognized by the coatomer complex (COPI). Coatomer is a multiprotein complex composed of two subcomplexes that include a trimer of α-COP, β'-COP, and ε-COP and a tetramer composed of β-COP, γ-COP, δ-COP, and ζ-COP [148]. The α-COP and β'-COP subunits of coatomer are responsible for binding di-lysine motifs. We tested for a physical interaction between AnkB-Paris and α-COP or β'-COP *in vivo* by Co-IP but were unable to detect any interaction. It is possible that overexpression of multiple members of the COPI complex is required to detect a physical interaction with AnkB-Paris.

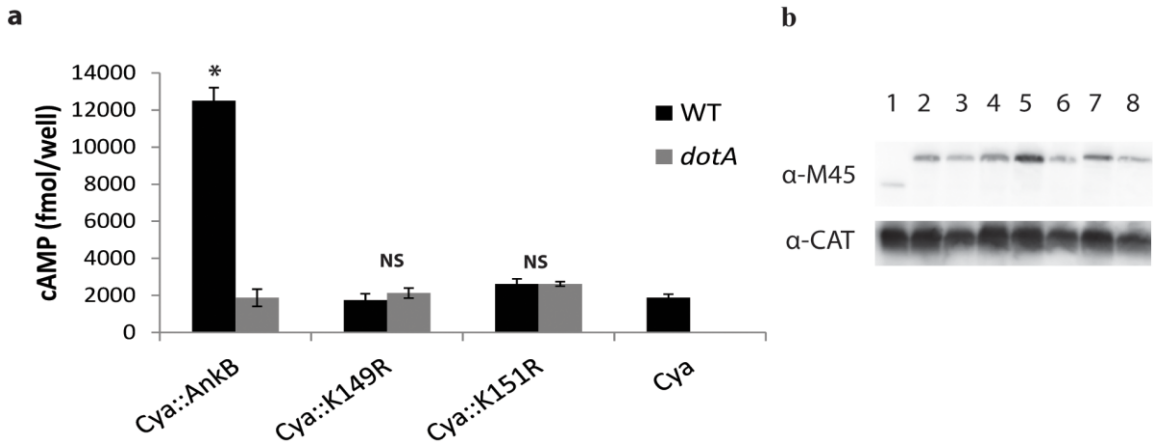

Figure 20. The putative di-lysine motif in the C-terminus of AnkB-Paris is essential for translocation by the Dot/Icm system. a. Translocation of AnkB-Paris into U937 cells at 2 hours post-infection by WT or *dotA* mutant bacteria expressing either Cya (negative control) or the indicated Cya::AnkB-Paris fusions. Data represent the mean cAMP concentration of 3 wells (±1SD). *p<0.005 compared to *dotA* harboring Cya::AnkB-Paris. **b.** Proteins derived from 1 x 10^8 bacteria were loaded onto an SDS-PAGE gel and expression of fusion constructs was detected by Western blot using an antibody to the M45 epitope present in all Cya fusions. Blots were re-probed with anti-CAT antibodies. Lanes 1: WT Cya, 2: *dotA* Cya-AnkB-Paris $K^{151}R$, 3: WT Cya-AnkB-Paris $K^{151}R$, 4: *dotA* Cya-AnkB-Paris $K^{149}R$, 5: WT Cya-AnkB-Paris $K^{149}R$, 6: *dotA* Cya-AnkB-Paris, 7: WT Cya-AnkB-Paris.

The Putative Di-lysine Motif of AnkB-Paris is Required for *in-trans* Rescue of the *ankB* Mutant

Since AnkB-Paris^{K149R} and AnkB-Paris^{K151R} are not translocated by the Dot/Icm system, we could not test the potential effect of these mutations on intracellular growth or decoration of the LCV with polyubiquitinated proteins. We have previously shown that the *ankB* null mutant of strain AA100/130b is rescued for intra-vacuolar growth within HEK293T ectopically expressing AnkB but not by the farnesylation-defective AnkB variant that has a substitution of the cysteine within the C-terminal CaaX farnesylation motif [110]. This is due to the ability of ectopically expressed AnkB to be farnesylated and anchored to the cytosolic side of the plasma membrane where polyubiquitinated proteins are assembled, while the farnesylation defective variant of AnkB is defective. Since this approach bypasses the need for translocation, we transfected 3X Flag-tagged versions of AnkB-Paris, AnkB-Paris^{K149R} and AnkB-Paris^{K151R} or 3X Flag vector control into HEK293T cells and then infected with the *ankB* mutants. At 10 hours post-infection, cells were fixed and examined for formation of replicative vacuoles using confocal microscopy. Our data showed that replication of the *ankB* mutant was efficiently *trans*-rescued by ectopically-expressed AnkB-Paris compared to the vector control (Fig. 21a-21b). In contrast, replication of the *ankB* mutant was not rescued in cells ectopically expressing AnkB-Paris^{K149R}, AnkB-Paris^{K151R}, or AnkB-Paris^{K149,151R} substitution mutants. Ectopically expressed was localized to the perinuclear ER region while the AnkB-Paris^{K149,151R} was distributed throughout the cytoplasm (Fig. 16c-16d). These data indicate that the putative

di-lysine ER-retention motif is indispensable for function of AnkB-Paris; likely through membrane anchoring to the ER-derived LCV membrane.

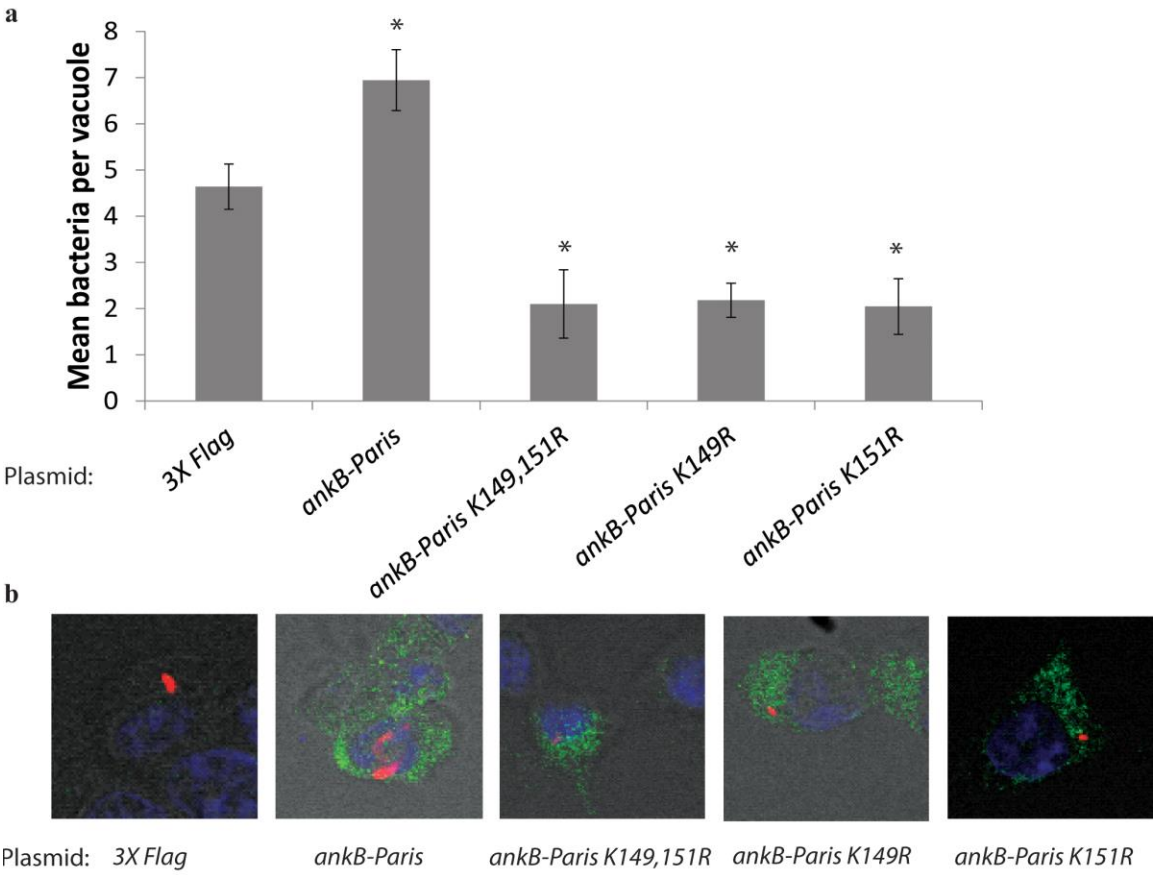

Figure 21. Requirement of the putative di-lysine motif in the C-terminus of ectopically expressed AnkB-Paris for *trans*-rescue of the *ankB* mutant. a. HEK293T cells were first transfected with plasmids encoding 3X-Flag empty vector, 3X-Flag AnkB-Paris, 3X-Flag AnkB-Paris $K^{149, 151}R$, 3X-Flag AnkB-Paris $K^{149}R$, or 3X-Flag AnkB-Paris $K^{151}R$ and then infected with the *ankB* mutant. Intracellular replication was analyzed at 10 hours post infection by confocal microscopy. The results are based on examination of 50 infected/transfected cells using three biological replicates. The mean number of bacteria per transfected HEK293T cell is shown. Error bars represent 1 standard deviation. *p<0.01 compared to cells transfected with empty vector. **b.** Representative confocal microscopy images of cells infected with the *ankB* mutant and expressing the indicated *3X Flag-ankB* fusion. Anti-flag staining is shown in green and anti-Lpn is shown in red.

Discussion

Among the more than 300 confirmed and predicted effectors of *L. pneumophila*, very few of them are required for intracellular proliferation and AnkB is one the effectors indispensable for intracellular proliferation [59]. These have been an emerging common theme of variations in the number of effectors and their paralogues among various strains and phenotypic differences between various strains associated with these differences [59, 149]. Loss of the AnkB AA100/130b effector results in a more severe intracellular defect in macrophage and amoeba and *in vivo* compared to AnkB-Paris, despite the observations that both function similarly in decorating the LCV with polyubiquitinated proteins [71, 98, 99]. Although host proteasomal degradation is essential for intracellular replication of the Philadelphia-derived Lp02 strain, its AnkB homologue does not contribute to decoration of the LCV with polyubiquitinated proteins or intracellular replication [95, 150]. This suggests that other F-box proteins or ubiquitin ligases, such as SidE and LubX, are involved in decorating the LCV with polyubiquitinated proteins [103, 151-153]. It is also becoming clear that *L. pneumophila* translocate deubiquitinases that remove ubiquitin from the modified protein, and variation in translocated deubiquitinases between various *L. pneumophila* isolates is likely to be a contributing factor for differences between them in polyubiquitination of the vacuoles and the effectors and metaeffectors (effectors of effectors) involved [152]. Whether metaeffectors of AnkB varies between various isolates is not known. In addition, modification of AnkB by K^{11}-linked polyubiquitination and by asparagine hydroxylation has been shown for

AnkB-AA100 but it is not known how that differs between isolates [141, 143]. Considering the phenotypic differences between isolates as a consequence of the loss of AnkB and the structural differences in AnkB between the two characterized strains Paris and AA100/130b, it is important to decipher the biological bases of these differences for one of the very few effectors required for intracellular proliferation of *L. pneumophila*.

Despite the frame shift mutation and deletion of the C-terminal CaaX farnesylation motif, AnkB-Paris (AnkB1) is required for decoration of the LCV with polyubiquitinated proteins [98]. Concurrently, a unique NKYAP sequence ER retention motif is generated at amino acids 150-154 of AnkB-Paris. The crystal structure of AnkB indicates that the third ankyrin repeat is truncated in AnkB-Paris [146]. Each ankyrin repeat domain is composed of two α-helices connected by a β-loop where the substrate binding domain is located [64]. Our phylogenetic data show that the *ankB1* allele is predominant among environmental isolates. Statistical support for positive selection in *ankB* codons and lineages, and variable effects of *ankB* genotype on recruitment of polyubiquitinated proteins suggest that AnkB may be functionally pleiotropic and may engage diverse cellular pathways triggered by various strains to ensure survival during intracellular residency. Other possibilities include differential regulation of *ankB* in different isolates; read through of the stop codon (encoding a modified aa in place of stop codon) resulting in a full-length functional AnkB similar to AnkB-AA100/130b. We conclude that positive selection acts on few *ankB* codons; that the *ankB1* allele itself is maintained in natural populations by

positive selection specifically on the NKYAP ER retention motif; and that the relatively high frequency of the *ankB1* allele in environmental isolates likely reflects a functionally advantageous trait conferred by the *ankB1* allele. The selective advantage to harbor the *ankB1* allele among environmental isolates of *L. pneumophila* could be due to a more efficient anchoring to the LCV membrane through the di-lysine ER retention motif compared to farnesylation in some unicellular hosts and/ or the third ankyrin domain that is truncated in AnkB1 does not interact with host targets in environmental host but interacts with a specific human target. It is also possible that other effectors expressed by various strains may compensate for the loss of the third ankyrin domain in the *ankB-Paris* allele. Identification of the AnkB-interacting targets and their interacting domains in AnkB should facilitate deciphering these possibilities.

Despite the lack of the farnesylation motif, AnkB-Paris is anchored to the cytosolic side of the LCV membrane. However, substitutions of the di-lysine ER retention motif results in failure to translocate the effector. This indicates an overlap in the signal for membrane anchoring and Dot/Icm-mediated translocation of the AnkB-Paris effector. Unfortunately, it is not possible to determine whether the di-lysine ER retention motif of AnkB-Paris was responsible for localization to the ER-derived LCV membrane, since the di-lysine ER-retention motif substitution in AnkB-Paris resulted in loss of translocation by the Dot/Icm system. In addition, the perinuclear ER-like distribution of ectopically-expressed AnkB-Paris is lost upon alternation of the di-lysine ER retention motif. Importantly, the *trans*-rescue of the *ankB* mutant within cells

ectopically expressing AnkB-Paris and the failure of the ER retention di-lysine mutant in *trans*-rescue clearly shows that the ER retention di-lysine motif is essential for the function of AnkB-Paris. This may not be surprising, since substitution of the cysteine within the CaaX farnesylation motif AnkB-AA100 results in a total loss of function of the effector in *trans*-rescue of the *ankB* mutant for intra-vacuolar proliferation [110]. We conclude that anchoring of AnkB variants to host membranes is essential for function, regardless of the mechanism of membrane anchoring by farnensylation or by the di-lysine ER-retention motif.

CHAPTER 3

STRUCTURAL MIMICRY BY A BACTERIAL F-BOX EFFECTOR HIJACKS THE HOST UBIQUITIN-PROTEASOME SYSTEM[2]

Results

Structure of AnkB/Skp1

The structure of the AnkB effector was determined in complex with its host partner, Skp1. It contains one molecule in the asymmetric unit with interpretable electron density for Pro2-Cys160 of Skp1 and Lys2-Ala165 of AnkB. The structure of AnkB resembles a step stool, with the F-box clasped into a groove formed by helices 5 – 8 of Skp1, and the ankyrin domain forming the next step (Fig. 23a). The F-box adopts a typical fold, with three α-helices in a right-handed superhelical organization. An overlay of the F-box of AnkB with other F-boxes reveals high similarity – an RMSD of 0.7 Å over 33 Cα atoms with a Skp1-Skp2 complex (PDB code 1FQV) (Fig. 23b) [85]. We also solved the structure of the isolated ankyrin domain (residues 54 to 168) to close to 1 Å resolution. The

[2] The work in this chapter was done in collaboration with Kathy Wong, Guennadi Kozlov, Miroslaw Cygler, and Kalle Gehring. They solved the crystal structure of AnkB. My contribution was the design, performance, and analysis of the experiments in figures 25 and 26, as well as the writing and editing of the manuscript. The contents of this chapter are from "Structural Mimicry by a Bacterial F Box Effector Hijacks the Host Ubiquitin-Proteasome System", by K. Wong, et al, 2017, *Structure, 25(2)*, p. 376-383.

domain is composed of three ankyrin repeats – one more than originally predicted – a short middle repeat (Pro97 to Lys116) flanked by two longer repeats (Ile54 to Lys81 and Pro131 to Glu161). Each repeat adopts a helix-turn-helix fold with connecting loops forming an L-shaped interaction surface typical of other ankyrin repeats (Fig. 22b and 23a) [154]. The majority of ankyrin domains contain four to seven repeats, while up to 34 repeats have been reported [155]. The largest sequence differences generally occur in the loop regions and confer binding specificity. Refinement statistics for both structures are shown in Table 3.

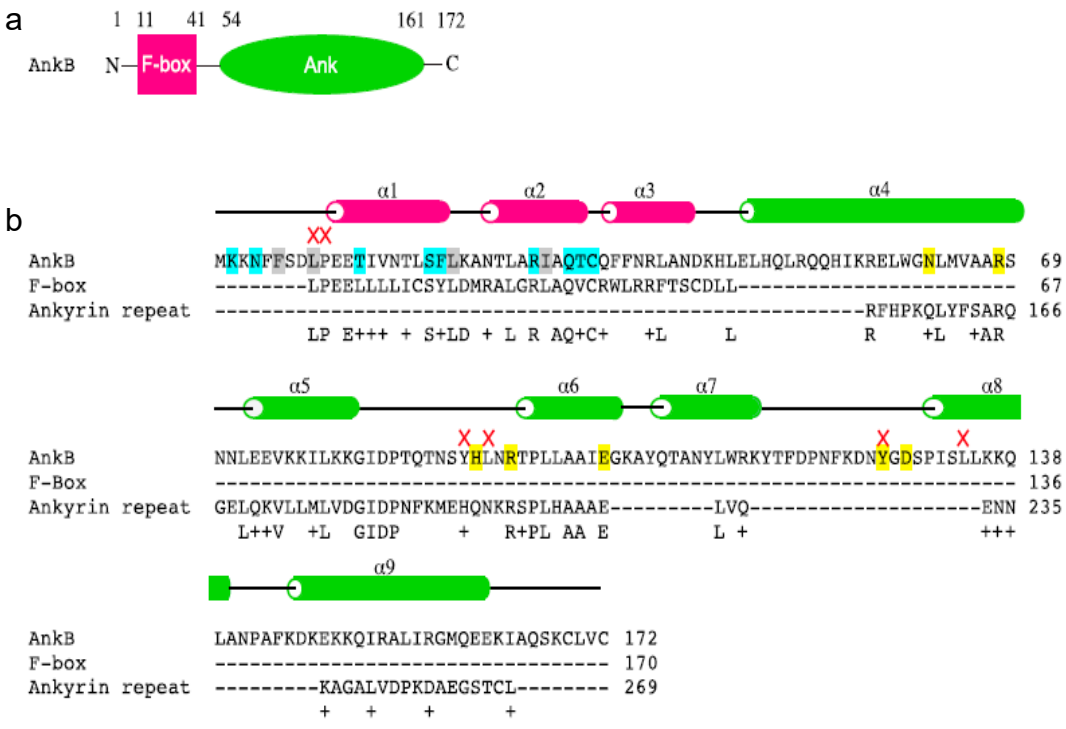

Fig. 22. Domain organization of AnkB a. AnkB is composed of two domains: an F-box (pink) and ankyrin repeats (green). **b.** Sequence alignment of AnkB with human proteins containing an F-box [EAW49753.1] and ankyrin repeats [AAH11608.2]. The secondary structure elements and important residues in AnkB are highlighted: *cyan,* hydrogen bonds with Skp1; *gray*, hydrophobic interactions with Skp1; *yellow,* putative substrate binding residues in ankyrin repeats; *red crosses,* mutations that prevent intracellular growth; *α*, alpha-helix.

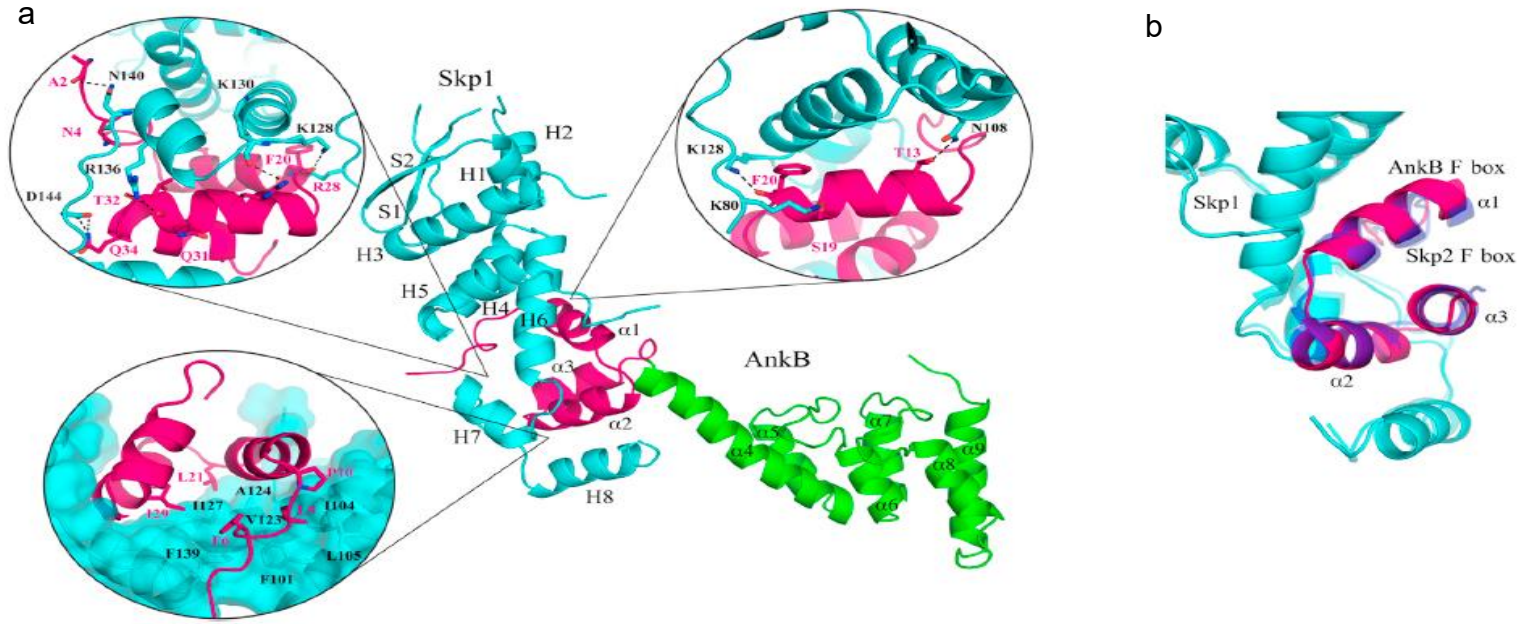

Fig. 23. Structure of AnkB as a component of the E3 ubiquitin ligase complex. a. Co-crystal of AnkB (*pink,* F-box; *green,* ankyrin repeats) and Skp1 (*cyan*). Hydrogen bonds between AnkB and human Skp1 are highlighted. Residues involved in the hydrophobic interaction surface between Skp1 and F-box are labeled. In addition to the four AnkB residues involved, Pro10 is also shown. Leu9 and Pro10 are highly conserved between F-boxes and abolish binding when mutated to alanine [99]. Skp1 secondary structure elements are labeled: *H*, alpha-helix; *S*, beta-sheet. AnkB secondary structure elements are labeled: *α*, alpha-helix. **b.** Comparison of AnkB (*pink*) and the human F-box protein Skp2 (*blue*) binding to Skp1 (*cyan*) (PDB code 1FQV) yields an RMSD of 0.7 Å over the F-box Cα atoms.

Table 3: Data collection and refinement statistics for AnkB

Data collection	54-168	1-168/Skp1(1-163)
Space group	$C222_1$	$P2_12_12_1$
Cell dimensions		
a, *b*, *c* (Å)	54.32, 80.49, 54.08	53.58, 57.04, 150.90
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	50-1.15 (1.17-1.15)[1]	50-2.85 (2.90-2.85)
R_{sym}	0.104 (0.435)	0.113 (0.622)
$I / \sigma I$	22.3 (3.8)	48.5 (6.42)
Completeness (%)	98.6 (97.9)	100 (100)
Redundancy	3.6 (3.6)	14.2 (14.5)
Refinement		
Resolution (Å)	45.0-1.15	75.45-2.85
No. reflections	39731	10735
R_{work} / R_{free}	0.173/0.191	0.219/0.275
No. atoms		
Protein	941	2498
B-factors		
Protein	12.79	40.40
Water	24.50	48.74
R.m.s deviations		
Bond lengths (Å)	0.016	0.008
Bond angles (°)	1.697	1.253
Ramachandran statistics (%)		
Most favored regions	100.0	95.7
Additional allowed regions	0.0	4.3

[1]Highest resolution shell is shown in parentheses.

Structural basis of AnkB-Skp1 binding

Full-length AnkB was insoluble when expressed without Skp1. The structure of the complex explains this phenomenon as the F-box is unlikely to fold without Skp1. A large hydrophobic surface formed by the N-terminal tail and helices 1 and 2 of AnkB interacts with helices 5, 6, and 7 of Skp1 (Fig. 23a). Multiple hydrogen bonds with helix 7 and its surrounding loops of Skp1 also stabilize the interaction. AnkB is insoluble without Skp1 to shield the hydrophobic surfaces and provide polar contacts.

A previous mutagenesis study showed that a mutation in the AnkB F-box domain leads to a defect in intracellular bacterial proliferation [99]. The L9A P10A mutant is unable to interact with host Skp1 and fails to decorate the LCV with polyubiquitinated proteins, a crucial source of carbon and energy for intracellular proliferation [99]. The leucine forms part of the hydrophobic interaction surface with Skp1, while the proline is responsible for initiating the first F-box α-helix (Fig. 23a). Both residues are highly conserved among F-box domains and their mutation to alanine likely prevents proper folding of the AnkB F-box domain [156].

Identification of the substrate-binding site on AnkB

We observed unusually well-ordered crystal contacts between the C-terminal tail of AnkB and the ankyrin domain of another molecule (Fig. 24a). The contacts also occurred in the crystals of the AnkB-Skp1 complex that adopt a different space group (data not shown). In the ankyrin domain crystal, a total of nine hydrogen bonds are formed between the backbone of the C-terminal tail,

Q_{160}EEKI, and the putative AnkB substrate-binding site. Additional side chain polar contacts contribute to the structuring of the peptide in the groove formed by the first two ankyrin repeats.

To validate the identification of the substrate-binding site, we ^{15}N-labeled the ankyrin domain and acquired NMR ^{15}N-^{1}H correlation spectra following a stepwise addition of peptides. Titrations of the ankyrin domain with a pentapeptide QEEKI derived from the AnkB C-terminus, resulted in several chemical shift changes, indicating weak but significant binding. We also tested the effects of N-terminal acetylation and C-terminal amidation and single amino acid substitutions to alanine but these had no significant impact on binding. Titration with a second peptide, PRLPTL, which binds to the ankyrin domain of ANKRA2 showed smaller shifts suggestive of weaker binding (Fig. 24b) [157].

Residues within the ankyrin domain of AnkB are essential for recruitment of polyubiquitinated proteins to the LCV

We selected four residues for mutagenesis that are predicted to be involved in substrate binding based on the AnkB crystal structure. Tyr91, Leu93 and Tyr127 form a hydrogen-bonding network connecting and stabilizing the loop residues. We also observed Leu134, located on the first helix of the last ankyrin repeat, is solvent exposed and potentially provides hydrophobic interactions with substrates.

To validate our structural prediction that Tyr91, Leu93, Tyr127 and Leu134 are important in the biological function of AnkB during infection, the residues were substituted with lysine. To verify that the lysine mutants were still correctly

folded, we acquired 1D NMR spectra of the mutants and wild-type ankyrin domain. The mutant spectra are similar to that of the native domain, indicating proper folding (data not shown). Human monocytes-derived macrophages (hMDM) were then infected with the wild-type strain (AA100/130b), the *ankB* null mutant, or the *ankB* mutant complemented with either a wild-type copy of *ankB* or one of the *ankB* mutant constructs. At 2 hours post-infection, the function of the AnkB variants was evaluated by assessment of recruitment of polyubiquitinated proteins using confocal microscopy. The data showed that approximately 55% of the LCVs of the wild-type strain-infected cells stained positively for polyubiquitin, whereas only 25% of the LCVs harboring the *ankB* null mutant were positive. Complementing the *ankB* mutant with a wild-type copy of *ankB* fully restored recruitment of polyubiquitinated proteins to the LCV with approximately 53% of the LCVs staining positively (Fig. 25a). The single lysine *ankB* mutants and the Y91K L93K double mutant were similarly functional in the complementation assay. In contrast, the Y91K L93K Y127K triple mutant and the Y91K L93K Y127K L134K quadruple mutant were defective in recruitment of polyubiquitinated proteins at levels similar to the *ankB* null mutant. This confirms the importance of the substrate binding site on the ankyrin domain for the recruitment of polyubiquitinated proteins to the LCV [141].

The ankyrin domain is required for intracellular replication

We determined if mutations of these four residues resulted in a replication defect of the bacteria within the LCV. hMDMs were infected with wild-type *L. pneumophila*, its isogenic *dotA* or *ankB* mutants, or the *ankB* mutant

complemented with a wild-type or mutated copy of *ankB*. At 10 hours post-infection, the *dotA* null mutant showed no replication and the *ankB* null mutant was markedly compromised compared to the wild-type strain (Fig. 25b). Complementation of the *ankB* mutant with a wild-type copy of *ankB* restored replication to wild-type levels. The single mutants Y91K, L93K, Y127K, L134K, and double mutant Y91K L93K were also effective in restoring growth. In contrast, complementation with the triple and quadruple mutations in the ankyrin domain showed a significant defect in replication. Similar results were obtained with the U937 macrophage cell line (data not shown). These data are in agreement with the decreased ubiquitination of the LCV observed in bacteria expressing *ankB* with the same triple and quadruple mutations. Residues Tyr91, Leu93, Tyr127, and Leu134 within the ankyrin repeats of AnkB are critical both for recruitment of ubiquitinated proteins to the LCV and for replication within hMDMs and U937 macrophages (Fig. 25-26).

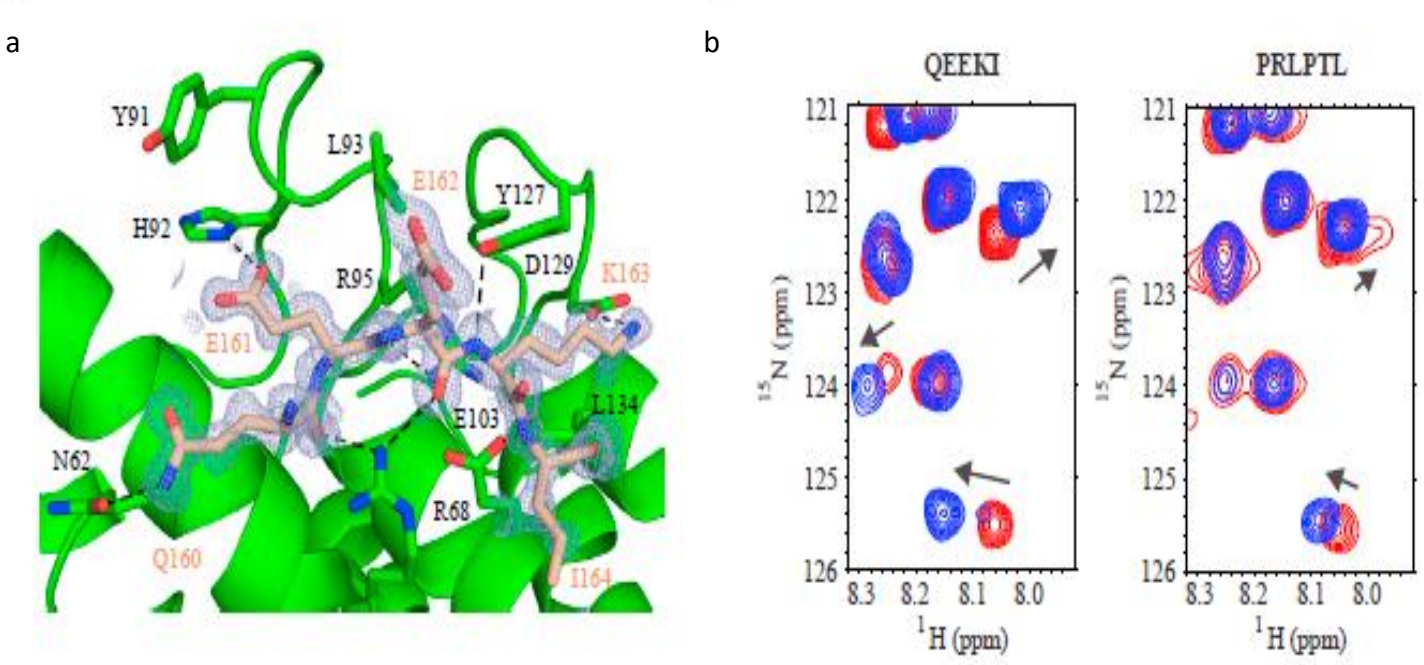

Fig. 24. Substrate binding by the ankyrin repeats. **a.** Crystal contacts in the AnkB ankyrin repeats mimic substrate binding. The C-terminal tail (residues 160 to 164, QEEKI) of one molecule (*wheat*) binds to the ankyrin repeats of another (*green*). Hydrogen bonds are indicated by dashed black lines. Residues involved in contacting the peptide and residues that were mutated for further functional studies are labeled in black. An omit map of the substrate is colored and labeled in wheat. See also Figure S1. **b.** Downfield region of HSQC spectra of the ^{15}N-labeled ankyrin domain show chemical shifts upon titration with the QEEKI peptide (C-terminal tail) and weaker binding upon titration with the PRLPTL peptide (negative control) at 0 mM (red) and 8 mM (blue).

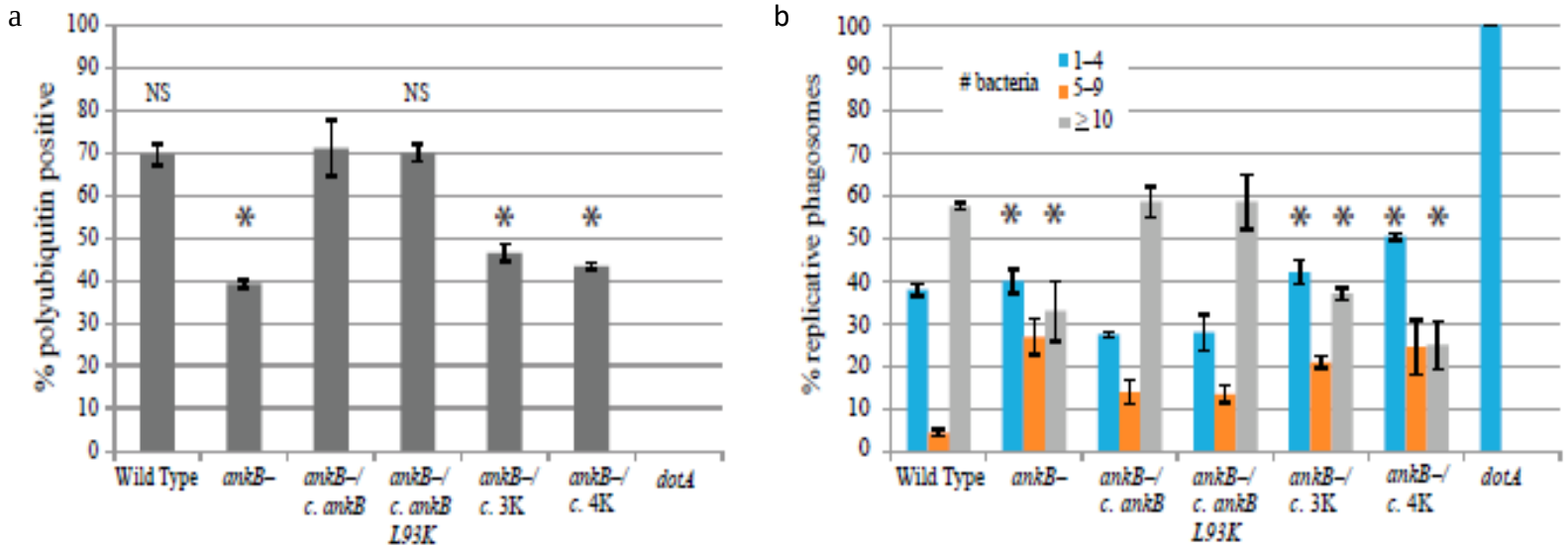

Fig. 25 Replication and Ubiquitylation Defect in hMDMs. a. Percentage of LCVs colocalizing with polyubiquitinated proteins by confocal microscopy at 2 hours post-infection. Human monocytes-derived macrophages (hMDMs) were infected with wild type *L. pneumophila, ankB* mutant, or the *ankB* mutant complemented with either a wild type copy of *ankB* or *ankB* containing the indicated single or multiple mutations. The data are representative of three independent experiments and are based on analysis of 100 infected cells per strain with each strain analyzed in duplicate. Error bars indicate +/- 1 SD. Abbreviations: 3K and 4K refer to *Y91K/L93K/Y127K* and *Y91K/L93K/Y127K/L134K*, respectively. *p<0.02 compared to *ankB* mutant complemented with wild type *ankB*. NS, not significant. See also Figure S2. **b.** hMDMs were infected with wild type, *dotA* mutant, *ankB* mutant, or the *ankB* mutant complemented with either wild type *ankB* or *ankB* containing the indicated mutations at an MOI of 1 followed by 1 hour treatment with gentamicin to kill extracellular bacteria. After 10 h, 100 infected cells were analyzed by confocal microscopy and the number of bacteria per cell was determined. The data are representative of two independent experiments with each strain analyzed in duplicate. Error bars indicate +/- 1 SD. Abbreviations: 3K, and 4K refer to *Y91K/L93K/Y127K* and *Y91K/L93K/Y127K/L134K*, respectively. *p <0.05 compared to *ankB* mutant complemented with wild type *ankB*.

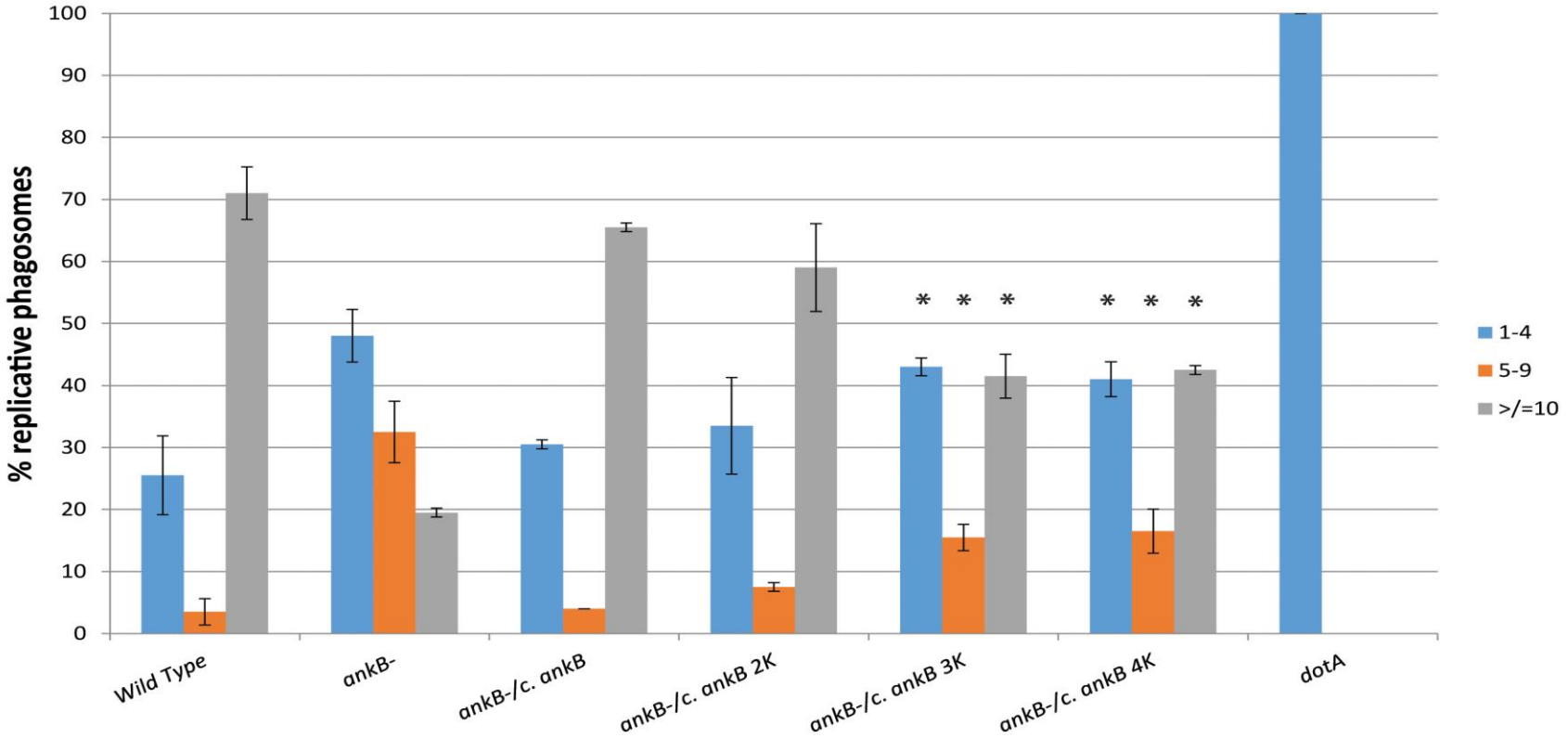

Fig. 26 Mutation of three or more key residues within the Ankyrin domain of AnkB causes a replication defect in U937 macrophages U937 cells were infected with wild type, *dotA* mutant, *ankB* mutant, or the *ankB* mutant complemented with either wild type *ankB* or *ankB* containing the indicated mutations at an MOI of 1 followed by 1 hour treatment with gentamicin to kill extracellular bacteria. After 10 h, 100 infected cells were analyzed by confocal microscopy and the number of bacteria per cell was counted. Each strain was analyzed in duplicate and the results are representative of two independent experiments. Error bars indicate +/- 1 SD. Abbreviations: 2K, 3K, and 4K refer to *ankB Y91K L93K*, *ankB Y91K L93K Y127K*, and *ankB Y91K L93K Y127K L134K*, respectively. *p <0.05 compared to *ankB* mutant complemented with wild type *ankB*.

Discussion

Here, we present the first structure of a bacterial F-box protein, the targeting subunits of SCF ligases. There are close to 70 F-box proteins in humans that are involved in a wide range of diseases. These proteins are composed of an F-box domain and a variable targeting domain which belongs to three main classes: WD40 domains, leucine-rich repeats, and other domains. AnkB represents a unique association of an F-box and ankyrin repeats that appears to be unique to a small number of lower eukaryotes, bacteria, and viruses [158]. Two other F-box effectors exist in the *Legionella* genome that could interact with the host SCF complex, but do not contain an ankyrin domain. One has a coiled coil domain (lpp2486) and the other consists of only an F-box (lpp0233). Previous mutagenesis studies have shown the importance of the AnkB F-box for acquisition of polyubiquitinated proteins to the LCV and bacterial proliferation [99]. The ankyrin domain of AnkB is likely involved in recruiting substrates for ubiquitination. The structures of a large number of ankyrin protein complexes have been determined and reveal a wide range of types of interactions. Generally, ankyrin domains use the inter-repeat loops and inner row of α-helices to bind other proteins; however, there is no consensus for the structure of the bound partner [154]. Ankyrin repeats can bind discontinuous protein surfaces, α-helices, and extended strands as observed for AnkB. The interactions of AnkB with its C-terminal tail most closely resemble the complex of ANKRA2 with a PxLPxI/L motif found in some histone deacetylases and other proteins we observed low affinity binding of AnkB to the peptide PRLPTL [157].

The ankyrin domain of AnkB shows broad specificity. This is typical of ankyrin domains that bind unfolded or extended peptide sequences and likely reflects the preponderance of AnkB interactions with the backbone atoms in the bound peptide. Fitting of the NMR titration curve suggested a dissociation constant (K_d) greater than 8 mM for the QEEKI peptide (Fig. 24b). Single amino acid substitutions in the peptide did not result in significant changes in the titration behavior, which is consistent with low specificity and a distributed binding interface. Similarly, single point mutations in the AnkB ankyrin domain did not perturb its function in poly-ubiquitination of LCVs and promoting *Legionella* proliferation (Fig. 25a and 25b).

In cells, AnkB is unlikely to bind its own tail or that of another AnkB molecule. The QEEKI motif extends from the final helix in the third ankyrin repeat and is unable to bind to the first two repeats. The interactions between two AnkB molecules is also unlikely as the QEEKI motif is separated by only four residues from the AnkB CaaX farnesylation site. The tethering of AnkB to the membrane would block access of the QEEKI of one molecule to the ankyrin domain of another. Sequence alignment of the AnkB gene between different strains reveals that the Paris strain homolog is a truncated version of the AnkB structure presented in this paper. While the last α-helix of the last ankyrin repeat is absent, the Paris homolog retains the last loop and half of the last repeat. From analysis of the crystal structures, this would suggest that the Paris homolog would still have a functional substrate-binding interface.

To date, two interacting partners of AnkB have been identified. Parvin B (ParvB), a target of AnkB ubiquitination was identified by a yeast two-hybrid screen and co-immunoprecipitation [98]. ParvB functions in regulating the actin cytoskeleton for cell adhesion and migration [159]. Solubility issues prevented us from detecting direct binding of ParvB to the purified AnkB ankyrin domain. More recently, TRIM21 was identified by coimmunoprecipitation as a partner of AnkB. TRIM21 attaches Lys11-linked polyubiquitin chains on Lys76 of AnkB without affecting AnkB stability [141].

Studies have elucidated two roles of AnkB in *Legionella* virulence. Following phagocytosis, *L. pneumophila* injects effector proteins into the host cell cytosol via the Dot/Icm secretion system. Considerable redundancy exists between effectors and loss of the *dotA* gene (equivalent to a knockout of all 300 effectors) gives rise to a much stronger ubiquitination and replication deficiency than the loss of only *ankB* (Fig. 25a-25b). Nonetheless, AnkB is effectively essential for virulence and acts as a linker to recruit the SCF complex to the LCV. Farnesylation of AnkB appears to be essential for its function; however, there is strain specificity as AnkB from the *Legionella* strain Paris lacks the CaaX motif but retains function [98, 110, 145]. By co-opting the host SCF complex, AnkB redirects host ubiquitination to the LCV and substrates selected by AnkB. We have built a model of AnkB in context of the SCF ubiquitination complex and the connected UbcH7 (E2 conjugating enzyme), by aligning AnkB onto the F-box of a Cul1-Rbx1-Skp1-Skp2 (PDB code 1LDK) and docking UbcH7 onto the Rbx1 RING domain based on a c-Cbl-UbcH7 structure (PDB code 1FBV) (Fig. 27)

[132, 160]. In the model, the active cysteine (Cys86) of the E2 points towards the putative substrate-binding site of the AnkB ankyrin repeats, positioning the substrate to receive ubiquitin. Lys^{48}-linked poly-ubiquitination of the LCV is a critical step in the maturation of the LCV and required to prevent fusion with the lysosome. We observed a strong correlation between loss of ubiquitination activity and loss of *Legionella* proliferation for the AnkB mutations tested. *Legionella* hijacks members of the secretory pathway to fuse endoplasmic reticulum (ER)-derived vesicles to the (LCV). As Lys^{48}-linked polyubiquitin chains are also associated with recruitment of the autophagy machinery, the reported association of AnkB with E3 ligases containing different chain specificities, such as TRIM21, is particularly interesting [141].

AnkB also plays a role in enriching the cytosolic pool of free amino acids through triggering Lys^{48}-linked polyubiquitination and increased protein turnover [120]. The levels of amino acids in the infected host cell are insufficient sources of carbon, nitrogen and energy for *L. pneumophila* [161]. AnkB promotes intra-vacuolar proliferation by ubiquitinating host proteins for their degradation into free amino acids [120, 162]. The growth defect of the *ankB* null mutant in both protozoan and eukaryotic cells can be rescued by supplementation with a mixture of free amino acids [118, 120]. AnkB likely functions by directly recruiting substrate proteins through the ankyrin domain. Mutating either the F-box or the ankyrin domain of AnkB results in the same phenotype, suggesting that the ability of *Legionella* to co-opt host E3 ubiquitin ligases through molecular mimicry plays a key role in pathogenesis.

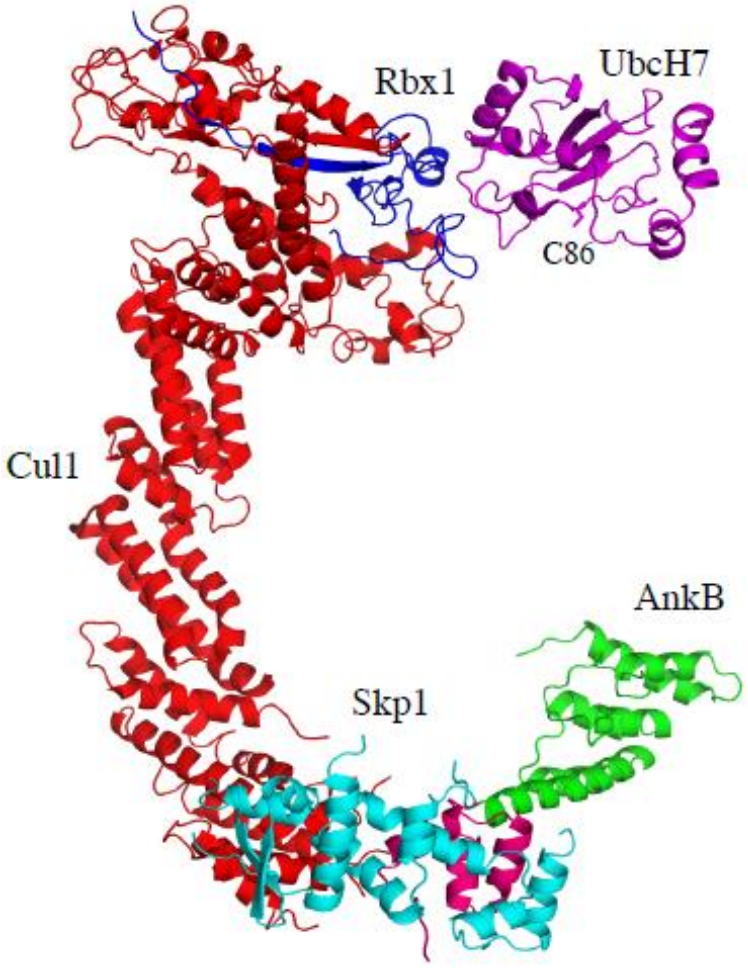

Fig. 27. Model in Context of Ubiquitylation. Model of AnkB as the F-box protein in the SCF E3 ubiquitin ligase complex (PBD code 1LDK) with an E2 conjugating enzyme (UbcH7) (PDB code 1FBV) docked. The active cysteine of UbcH7 (C86) is labeled. The color code is as follows: pink and green (F-box and Ankyrin domain of AnkB), cyan (Skp1), red (Cul1), blue (Rbx1), magenta (UbcH7).

CONCLUSIONS AND FUTURE DIRECTIONS

I identified a unique KNKYAP motif present within the C-terminus of AnkB-Paris that resembles an eukaryotic ER retention di-lysine motif. I show that this allele is predominant among environmental isolates and that the di-lysine motif is acted upon by positive selection. I also demonstrated that AnkB-Paris, like AnkB-AA100, is anchored to the cytosolic face of the LCV membrane, albeit by a mechanism distinct from farnesylation. The di-lysine motif is essential for the translocation of AnkB-Paris via the Dot/Icm system as mutations in either lysine of the motif abrogate translocation. Because of this, I was unable to determine if the di-lysine motif is essential for localization to the cytosolic face of the ER-derived LCV membrane. However, I showed the di-lysine motif is essential for *trans*-rescue of the *ankB* mutant in cells ectopically expressing AnkB-Paris, indicating a critical biological function of the motif that is distinct from translocation. The di-lysine motif is, most likely, acting as an ER retention motif anchoring AnkB-Paris to the ER-derived LCV membrane.

There are at least two alternative hypotheses that could explain the failure of the di-lysine mutants to *trans*-rescue the *ankB* mutant. First, the KNKYAP motif, which is unique to AnkB-Paris, may be crucial for anchoring the protein to the LCV surface by interacting with host and/or *L. pneumophila* proteins present on the LCV that are not involved in binding to di-lysine motifs. Mutation of the

lysine residues within C-terminus of AnkB-Paris may disrupt the interactions with these proteins resulting in loss of LCV localization and failure to *trans*-rescue the *ankB* mutant. Second, the di-lysine motif is contained within the last ankyrin repeat. Therefore, it is possible that substitution of these residues may impact the stability of the ankyrin domain and prevent AnkB-Paris from binding to its substrates. This would produce a loss of function in the *trans*-rescue experiment not by disruption of anchoring to the LCV membrane, but by loss of ability to bind and ubiquitylate substrates. Evidence in favor of the di-lysine hypothesis includes the striking similarity of the C-terminus of AnkB-Paris to the di-lysine consensus, the perinuclear distribution of AnkB-Paris upon ectopic expression, and the fact that the LCV is derived from ER and contains coatomer proteins known to bind di-lysine motifs.

Inside eukaryotic cells, the coatomer protein complex is responsible for binding di-lysine motifs and trafficking proteins from the Golgi to the ER [163, 164]. The coatomer complex is composed of a trimer (α-COP, β'-COP, ε-COP) and a tetramer (β-COP, γ-COP, δ-COP, and ζ-COP) [148]. I was unable to show an interaction between β'-COP and AnkB-Paris by co-IP in cells overexpressing both proteins. However, it is possible that overexpression of all members of the trimer subcomplex would be required for binding to AnkB-Paris as was shown for the di-lysine motif contained within diphtheria toxin [113]. Alternatively, we could test for binding to the coatomer subcomplex *in vitro* using fragments of AnkB-Paris fused to GST and immobilized on glutathione beads [147]. Since I have established that AnkB-Paris localizes to the ER-derived LCV membrane during

infection, we could also assess binding of AnkB-Paris to members of the coatomer complex by siRNA knockdown. In this scenario, cells would be treated with siRNA targeting one or more coatomer subunits and then infected with *L. pneumophila* expressing AnkB-Paris. Localization of AnkB-Paris to the LCV would be assessed by confocal microscopy. In support of this approach, coatomer subunits α, β, β', and γ were all shown to localize to the LCV in a study of the complete proteome of the WT LCV [162]. Such studies would provide additional evidence that the KNKYAP motif present in AnkB-Paris is functioning as a di-lysine motif.

Overall, I conclude that anchoring of AnkB variants to host membranes is essential for function, regardless of the mechanism of membrane anchoring by farnesylation or di-lysine ER retention. It will be interesting to explore the exact molecular interactions responsible for tethering the AnkB-Paris to the LCV membrane.

I identified residues within the three ankyrin domains of AnkB that are essential for function and likely involved in binding substrates. The crystal structure of AnkB revealed that it has a large hydrophobic surface in its N-terminus that interacts with helices 5, 6, and 7 of Skp1. The structure also revealed three ankyrin domains within AnkB. Each repeat adopts a helix-turn-helix fold with connecting loops in between that is typical of other ankyrin repeats. Three residues, Tyr91, Leu93, and Tyr127, were predicted to form a hydrogen-bonding network important for binding substrate proteins and one solvent-exposed hydrophobic residue, Leu134, potentially involved in making

hydrophobic interactions with substrates. Single or double substitutions in these residues had no effect on AnkB biological function in recruiting ubiquitylated proteins to the LCV or promoting intracellular replication. On the other hand, substituting three or four of these residues resulted in a complete loss of ubiquitylated protein recruitment and replication similar to an *ankB* null mutant. These data suggest that these residues are likely to be involved in binding to AnkB substrates that are ubiquitylated during infection. Once substrates of AnkB are identified, we will be able to confirm these key residues within AnkB are responsible for binding to its targets.

I sought to identify the specific host proteins targeted by AnkB for ubiquitylation. Despite using three different approaches, I was unable to identify AnkB-interacting proteins under our experimental conditions. This knowledge is important, as it may confirm the role of AnkB as a virulence factor involved in harvesting amino acids from the host. In this case, identification of substrates could reveal that AnkB targets host proteins rich in amino acids that are most favorable for energy production via the TCA cycle, such as Ser, Cys, or Ala. On the other hand, identification of specific substrates may reveal that AnkB is involved in modulating certain host cell pathways by altering the ubiquitylation state of key members.

Characterization of E3 ubiquitin ligase substrates in mammalian cells is challenging [165]. It is estimated that > 80% of the human proteome is subject to ubiquitin-mediated degradation, suggesting this process is involved in many cellular pathways [166]. There are approximately 600 E3 ligases in the human

genome and the substrates of these ligases are largely uncharacterized [140, 165]. Characterization of substrates is technically challenging for several reasons. First, E3 ligase activity is under tight control and deubiquitylases (DUB) act to reverse their activity making the entire process very dynamic [165]. Interactions between F-box protein and substrate are weak and transient with most encounters being unproductive since the dissociation rate is faster than the ubiquitylation rate [167]. In addition, there is significant redundancy as any particular substrate can be the target of multiple E3 ligases and one ligase can target multiple substrates depending on the cellular conditions [165]. Finally, the ubiquitylated substrates are rapidly degraded by the proteasome in many cases. These factors make isolation of substrates of a given E3 ligase challenging by traditional immunoprecipitation methods and may explain why I was unable to do so for AnkB.

A modified immunoprecipitation-based method has been successful in identifying F-box protein substrates [168]. Parallel adaptor capture proteomics (PAC) utilizes immunoaffinity-based purification of F-box substrates and couples it with the Comparative Proteomics Analysis Software Suite (CompPASS) [168]. The method involves transfecting cells with the F-box protein of interest and inhibiting the proteasome to increase the abundance of F-box-substrate complexes. The complexes are purified by immunoprecipitation and interacting proteins are identified by MS. The primary advantage of this approach is the use of the CompPASS software to identify high-confidence interacting proteins. This software compares results from an experiment to a table of 174 unrelated bait

proteins and calculates a probability that putative substrates are bona fide interacting proteins [168]. This approach successfully identified substrates of 19 different F-box proteins with an overall validation rate of 80% [168]. This approach could be readily applied to AnkB without the need for creating new constructs.

An alternative approach that does not rely on purification of F-box-substrate protein complexes is the BioID-based method, which is a proximity-based biotinylation method [135]. The E3 ligase of interest is fused to the *E. coli* biotin conjugating enzyme BirA, which can activate biotin but does not bind the activated molecule. As a result, the activated biotin diffuses away and reacts with any protein in close proximity, usually physically interacting proteins. Using this method, E3 ligase substrates can be efficiently biotinylated and purified on streptavidin under harsh lysis conditions [135]. This method overcomes the potential limitation, inherent to immunoaffinity-based methods, for transient low-affinity interactions encountered when working with E3 ligases and their substrates [135]. If AnkB interacts weakly or transiently with its substrate proteins, this method could result in identification of AnkB substrates that were not detected under our conditions.

A recently described method, thermophoresis, could be used for identification of novel AnkB interacting proteins. In this method, protein solutions are heated with a focused IR laser beam that causes a local movement of molecules either away from (positive) or toward (negative) the heating beam [169]. These movements are characteristic of the protein and can be altered by

changes in the protein's size, charge, or hydration shell, which can all change in response to binding to another protein [169]. This change in thermophoretic property can be measured and protein interactions detected at the picomolar level [169]. This technique could be applied to discovery of AnkB interacting proteins by testing fractions of cellular lysates for their binding to AnkB. Fractions could be further divided and purified to a degree that would allow identification of the proteins in the mixture that interact with AnkB. This technique could also be used to monitor for changes in binding of substrates to mutant versions of AnkB such as the ones described in chapter 3.

Overall, while I was not able to identify specific substrate proteins targeted for ubiquitylation by AnkB, I was able to identify key residues within the ankyrin domains that are likely responsible for binding specific substrates and promoting their ubiquitylation. I found that AnkB-Paris, which lacks the farnesylation motif found in the AA100/130b strain, contains a putative di-lysine ER retention motif. I show that the lysine residues within this motif are essential for it's biological function during infection. Similar to AnkB from the AA100/130b strain, AnkB-Paris localizes to the cytosolic face of the LCV most likely by virtue of its di-lysine ER retention motif.

REFERENCES

1. Fraser, D.W., et al., *Legionnaires' disease: description of an epidemic of pneumonia.* N.Engl.J.Med., 1977. **297**: p. 1189-1197.
2. McDade, J.E., et al., *Legionnaires' disease: isolation of a bacterium and demonstration of its role in other respiratory disease.* N.Engl.J.Med., 1977. **297**: p. 1197-1203.
3. Honigsbaum, M., *Legionnaires' disease: revisiting the puzzle of the century.* Lancet, 2016. **388**(10043): p. 456-7.
4. Jindal, S.K., *Koch's postulates - Pitfalls and relevance in the 21st century.* Indian J Tuberc, 2018. **65**(1): p. 6-7.
5. Den Boer, J.W. and E.P. Yzerman, *Diagnosis of Legionella infection in Legionnaires' disease.* Eur J Clin Microbiol Infect Dis, 2004. **23**(12): p. 871-8.
6. Mercante, J.W. and J.M. Winchell, *Current and emerging Legionella diagnostics for laboratory and outbreak investigations.* Clin Microbiol Rev, 2015. **28**(1): p. 95-133.
7. Rowbotham, T.J., *Current views on the relationships between amoebae, legionellae and man.* Isr.J.Med.Sci., 1986. **22**: p. 678-689.
8. Molofsky, A.B. and M.S. Swanson, *Differentiate to thrive: lessons from the Legionella pneumophila life cycle.* Mol Microbiol, 2004. **53**(1): p. 29-40.
9. Pierre, D.M., et al., *Diagnostic testing for Legionnaires' disease.* Ann Clin Microbiol Antimicrob, 2017. **16**(1): p. 59.
10. Cunha, B.A., A. Burillo, and E. Bouza, *Legionnaires' disease.* Lancet, 2015.
11. Phin, N., et al., *Epidemiology and clinical management of Legionnaires' disease.* The Lancet Infectious Diseases, 2014. **14**(10): p. 1011-1021.
12. *Legionellosis --- United States, 2000-2009.* MMWR Morb Mortal Wkly Rep, 2011. **60**(32): p. 1083-6.
13. Fields, B.S., R.F. Benson, and R.E. Besser, *Legionella and Legionnaires' disease: 25 years of investigation.* Clin Microbiol Rev, 2002. **15**(3): p. 506-26.
14. Marston, B.J., H.B. Lipman, and R.F. Breiman, *Surveillance for Legionnaires' disease: risk factors for mortality and morbidity.* Arch.Intern.Med., 1994. **154**: p. 2417-2422.
15. Blatt, S.P., et al., *Legionnaires' disease in human immunodeficiency virus-infected patients: eight cases and review.* Clin Infect Dis, 1994. **18**(2): p. 227-32.
16. Gutierrez Rodero, F., et al., *Legionnaires' disease in patients infected with human immunodeficiency virus.* Clin Infect Dis, 1995. **21**(3): p. 712-3.
17. Glick, T.H., et al., *Pontiac fever. An epidemic of unknown etiology in a health department: I. Clinical and epidemiologic aspects.* Am J Epidemiol, 1978. **107**(2): p. 149-60.
18. Xu, L. and Z.Q. Luo, *Cell biology of infection by Legionella pneumophila.* Microbes Infect, 2013. **15**(2): p. 157-67.

19. Mandell, L.A., et al., *Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults.* Clin Infect Dis, 2007. **44 Suppl 2**: p. S27-72.
20. Benin, A.L., R.F. Benson, and R.E. Besser, *Trends in legionnaires disease, 1980-1998: declining mortality and new patterns of diagnosis.* Clin Infect Dis, 2002. **35**(9): p. 1039-46.
21. Heath, C.H., D.I. Grove, and D.F. Looke, *Delay in appropriate therapy of Legionella pneumonia associated with increased mortality.* Eur J Clin Microbiol Infect Dis, 1996. **15**(4): p. 286-90.
22. Fliermans, C.B., et al., *Ecological distribution of Legionella pneumophila.* Appl.Environ.Microbiol., 1981. **41**: p. 9-16.
23. Katz, S.M. and J.M. Hammel, *The effect of drying, heat, and pH on the survival of Legionella pneumophila.* Ann.Clin.Lab.Sci., 1987. **17**: p. 150-156.
24. Taylor, M., K. Ross, and R. Bentham, *Legionella, protozoa, and biofilms: interactions within complex microbial systems.* Microb Ecol, 2009. **58**(3): p. 538-47.
25. Bhopal, R.S., et al., *Proximity of the home to a cooling tower and risk of non-outbreak Legionnaires' disease.* BMJ., 1991. **302**: p. 378-383.
26. Atlas, R.M., J.F. Williams, and M.K. Huntington, *Legionella contamination of dental-unit waters.* Appl Environ Microbiol, 1995. **61**(4): p. 1208-13.
27. Bollin, G.E., et al., *Aerosols containing Legionella pneumophila generated by shower heads and hot-water faucets.* Appl.Environ.Microbiol., 1985. **50**: p. 1128-1131.
28. Guyard, C. and D.E. Low, *Legionella infections and travel associated legionellosis.* Travel Med Infect Dis, 2011. **9**(4): p. 176-86.
29. Newton, H.J., et al., *Molecular pathogenesis of infections caused by Legionella pneumophila.* Clin Microbiol Rev, 2010. **23**(2): p. 274-98.
30. Yamamoto, H., et al., *Factors stimulating propagation of legionellae in cooling tower water.* Appl Environ Microbiol, 1992. **58**(4): p. 1394-7.
31. Iervolino, M., B. Mancini, and S. Cristino, *Industrial Cooling Tower Disinfection Treatment to Prevent Legionella spp.* Int J Environ Res Public Health, 2017. **14**(10).
32. George, J.R., et al., *Amino acid requirements of Legionella pneumophila.* J.Clin.Microbiol., 1980. **11**: p. 286-291.
33. Pine, L., et al., *Development of a chemically defined liquid medium for growth of Legionella pneumophila.* J Clin Microbiol, 1979. **9**(5): p. 615-26.
34. Pine, L., et al., *Development of a chemically defined liquid medium for growth of Legionella pneumophila.* J.Clin.Microbiol., 1979. **9**: p. 615-626.
35. Ristroph, J.D., K.W. Hedlund, and S. Gowda, *Chemically defined medium for Legionella pneumophila growth.* J.Clin.Microbiol., 1981. **13**: p. 115-119.
36. Tesh, M.J., S.A. Morse, and R.D. Miller, *Intermediary metabolism in Legionella pneumophila: utilization of amino acids and other compounds as energy sources.* J.Bacteriol., 1983. **154**: p. 1104-1109.
37. Eisenreich, W. and K. Heuner, *The life stage-specific pathometabolism of Legionella pneumophila.* FEBS Lett, 2016. **590**(21): p. 3868-3886.
38. Oliva, G., T. Sahr, and C. Buchrieser, *The Life Cycle of L. pneumophila: Cellular Differentiation Is Linked to Virulence and Metabolism.* Frontiers in Cellular and Infection Microbiology, 2018. **8**(3).
39. Eylert, E., et al., *Isotopologue profiling of Legionella pneumophila: role of serine and glucose as carbon substrates.* J Biol Chem, 2010. **285**(29): p. 22232-43.

40. Hammer, B.K. and M.S. Swanson, *Co-ordination of legionella pneumophila virulence with entry into stationary phase by ppGpp.* Mol Microbiol, 1999. **33**(4): p. 721-31.
41. James, B.W., et al., *Poly-3-hydorxyburyrate in Legionella pneumophila, an energy source for survival in low-nutrient environments.* Appl.Environ.Microbiol., 1999. **65**: p. 822-827.
42. Gillmaier, N., et al., *Growth-related Metabolism of the Carbon Storage Poly-3-hydroxybutyrate in Legionella pneumophila.* J Biol Chem, 2016. **291**(12): p. 6471-82.
43. Horwitz, M.A., *Phagocytosis of the Legionnaires' disease bacterium (Legionella pneumophila) occurs by a novel mechanism: engulfment within a pseudopod coil.* Cell, 1984. **36**: p. 27-33.
44. Horwitz, M.A. and F.R. Maxfield, *Legionella pneumophila inhibits acidification of its phagosome in human monocytes.* J.Cell.Biol., 1984. **99**: p. 1936-1943.
45. Horwitz, M.A., *Formation of a novel phagosome by the Legionnaires' disease bacterium (Legionella pneumophila) in human monocytes.* J Exp Med, 1983. **158**(4): p. 1319-31.
46. Tilney, L.G., et al., *How the parasitic bacterium Legionella pneumophila modifies its phagosome and transforms it into rough ER: implications for conversion of plasma membrane to the ER membrane.* J Cell Sci, 2001. **114**(Pt 24): p. 4637-50.
47. Swanson, M.S. and R.R. Isberg, *Association of Legionella pneumophila with the macrophage endoplasmic reticulum.* Infect.Immun., 1995. **63**: p. 3609-3620.
48. Kagan, J.C. and C.R. Roy, *Legionella phagosomes intercept vesicular traffic from endoplasmic reticulum exit sites.* Nat Cell Biol, 2002. **4**: p. 945-954.
49. Clemens, D.L., B.Y. Lee, and M.A. Horwitz, *Deviant expression of rab5 on phagosomes containing the intracellular pathogens mycobacterium tuberculosis and legionella pneumophila is associated with altered phagosomal fate [In Process Citation].* Infect Immun, 2000. **68**(5): p. 2671-84.
50. Berger, K.H. and R.R. Isberg, *Two distinct defects in intracellular growth complemented by a single genetic locus in Legionella pneumophila.* Mol.Microbiol., 1993. **7**: p. 7-19.
51. Marra, A., et al., *Identification of a Legionella pneumophila locus required for intracellular multiplication in human macrophages.* Proc.Natl.Acad.Sci.USA., 1992. **89**: p. 9607-9611.
52. Ensminger, A.W., *Legionella pneumophila, armed to the hilt: justifying the largest arsenal of effectors in the bacterial world.* Curr Opin Microbiol, 2016. **29**: p. 74-80.
53. Hilbi, H., G. Segal, and H.A. Shuman, *Icm/Dot-dependent upregulation of phagocytosis by Legionella pneumophila.* Mol Microbiol, 2001. **42**(3): p. 603-17.
54. Kirby, J.E., et al., *Evidence for pore-forming ability by Legionella pneumophila.* Mol.Microbiol., 1998. **27**: p. 323-336.
55. Isaac, D.T. and R. Isberg, *Master manipulators: an update on Legionella pneumophila Icm/Dot translocated substrates and their host targets.* Future Microbiology, 2014. **9**(3): p. 343-359.
56. O'Connor, T.J., et al., *Minimization of the Legionella pneumophila genome reveals chromosomal regions involved in host range expansion.* Proc Natl Acad Sci U S A, 2011. **108**(36): p. 14733-40.
57. de Felipe, K.S., et al., *Evidence for acquisition of Legionella type IV secretion substrates via interdomain horizontal gene transfer.* J Bacteriol, 2005. **187**(22): p. 7716-26.
58. Hubber, A. and C.R. Roy, *Modulation of host cell function by Legionella pneumophila type IV effectors.* Annu Rev Cell Dev Biol, 2010. **26**: p. 261-83.
59. Burstein, D., et al., *Genomic analysis of 38 Legionella species identifies large and diverse effector repertoires.* Nat Genet, 2016. **48**(2): p. 167-75.

60. Gomez-Valero, L., et al., *Comparative and functional genomics of legionella identified eukaryotic like proteins as key players in host-pathogen interactions.* Front Microbiol, 2011. **2**: p. 208.
61. Cazalet, C., et al., *Evidence in the Legionella pneumophila genome for exploitation of host cell functions and high genome plasticity.* Nat Genet, 2004. **36**(11): p. 1165-73.
62. Ghosh, S. and T.J. O'Connor, *Beyond Paralogs: The Multiple Layers of Redundancy in Bacterial Pathogenesis.* Front Cell Infect Microbiol, 2017. **7**: p. 467.
63. So, E.C., et al., *Creating a customized intracellular niche: subversion of host cell signaling by Legionella type IV secretion system effectors.* Can J Microbiol, 2015. **61**(9): p. 617-35.
64. Al-Khodor, S., et al., *Functional diversity of ankyrin repeats in microbial proteins.* Trends Microbiol, 2010. **18**(3): p. 132-9.
65. Mosavi, L.K., et al., *The ankyrin repeat as molecular architecture for protein recognition.* Protein Sci, 2004. **13**(6): p. 1435-48.
66. Voth, D.E., *ThANKs for the repeat: Intracellular pathogens exploit a common eukaryotic domain.* Cell Logist, 2011. **1**(4): p. 128-132.
67. Kohl, A., et al., *Designed to be stable: crystal structure of a consensus ankyrin repeat protein.* Proc Natl Acad Sci U S A, 2003. **100**(4): p. 1700-5.
68. Cazalet, C. and C. Buchrieser, *[What do we learn from the genome of Legionella pneumophila?].* Med Sci (Paris), 2005. **21**(5): p. 455-7.
69. Habyarimana, F., et al., *Molecular characterization of the Dot/Icm-translocated AnkH and AnkJ eukaryotic-like effectors of Legionella pneumophila.* Infect Immun, 2010. **78**(3): p. 1123-34.
70. Habyarimana, F., et al., *Role for the Ankyrin eukaryotic-like genes of Legionella pneumophila in parasitism of protozoan hosts and human macrophages.* Environ Microbiol, 2008. **10**(6): p. 1460-74.
71. Al-Khodor, S., et al., *A Dot/Icm-translocated ankyrin protein of Legionella pneumophila is required for intracellular proliferation within human macrophages and protozoa.* Mol Microbiol, 2008. **70**(4): p. 908-23.
72. Kerscher, O., R. Felberbaum, and M. Hochstrasser, *Modification of proteins by ubiquitin and ubiquitin-like proteins.* Annu Rev Cell Dev Biol, 2006. **22**: p. 159-80.
73. Qiu, J. and Z.Q. Luo, *Hijacking of the Host Ubiquitin Network by Legionella pneumophila.* Front Cell Infect Microbiol, 2017. **7**: p. 487.
74. Komander, D. and M. Rape, *The ubiquitin code.* Annu Rev Biochem, 2012. **81**: p. 203-29.
75. Hoege, C., et al., *RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO.* Nature, 2002. **419**(6903): p. 135-41.
76. Haglund, K., et al., *Multiple monoubiquitination of RTKs is sufficient for their endocytosis and degradation.* Nat Cell Biol, 2003. **5**(5): p. 461-6.
77. Kwon, Y.T. and A. Ciechanover, *The Ubiquitin Code in the Ubiquitin-Proteasome System and Autophagy.* Trends Biochem Sci, 2017. **42**(11): p. 873-886.
78. Price, C.T. and Y.A. Kwaik, *Exploitation of Host Polyubiquitination Machinery through Molecular Mimicry by Eukaryotic-Like Bacterial F-Box Effectors.* Front Microbiol, 2010. **1**: p. 122.
79. Welchman, R.L., C. Gordon, and R.J. Mayer, *Ubiquitin and ubiquitin-like proteins as multifunctional signals.* Nat Rev Mol Cell Biol, 2005. **6**(8): p. 599-609.
80. Iwai, K., *Diverse roles of the ubiquitin system in NF-kappaB activation.* Biochim Biophys Acta, 2014. **1843**(1): p. 129-36.
81. Spence, J., et al., *Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain.* Cell, 2000. **102**(1): p. 67-76.

82. Huang, J., et al., *RAD18 transmits DNA damage signalling to elicit homologous recombination repair.* Nat Cell Biol, 2009. **11**(5): p. 592-603.
83. Zheng, N. and N. Shabek, *Ubiquitin Ligases: Structure, Function, and Regulation.* Annu Rev Biochem, 2017. **86**: p. 129-157.
84. Dove, K.K. and R.E. Klevit, *RING-Between-RING E3 Ligases: Emerging Themes amid the Variations.* J Mol Biol, 2017. **429**(22): p. 3363-3375.
85. Schulman, B.A., et al., *Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex.* Nature, 2000. **408**(6810): p. 381-6.
86. Wang, Z., et al., *Roles of F-box proteins in cancer.* Nat Rev Cancer, 2014. **14**(4): p. 233-47.
87. Skaar, J.R., J.K. Pagan, and M. Pagano, *SCF ubiquitin ligase-targeted therapies.* Nat Rev Drug Discov, 2014. **13**(12): p. 889-903.
88. Skaar, J.R., J.K. Pagan, and M. Pagano, *Mechanisms and function of substrate recruitment by F-box proteins.* Nat Rev Mol Cell Biol, 2013. **14**(6): p. 369-81.
89. Lucas, X. and A. Ciulli, *Recognition of substrate degrons by E3 ubiquitin ligases and modulation by small-molecule mimicry strategies.* Curr Opin Struct Biol, 2017. **44**: p. 101-110.
90. Kainulainen, M., et al., *NSs Virulence Factor of Rift Valley Fever Virus Engages the F-Box Proteins FBXW11 and beta-TRCP1 To Degrade the Antiviral Protein Kinase PKR.* J Virol, 2016. **90**(13): p. 6140-7.
91. Wang, C., et al., *Identification of FBL2 as a geranylgeranylated cellular protein required for hepatitis C virus RNA replication.* Mol Cell, 2005. **18**(4): p. 425-34.
92. Yao, I., et al., *SCRAPPER-dependent ubiquitination of active zone protein RIM1 regulates synaptic vesicle release.* Cell, 2007. **130**(5): p. 943-57.
93. Li, Y. and B. Hao, *Structural basis of dimerization-dependent ubiquitination by the SCF(Fbx4) ubiquitin ligase.* J Biol Chem, 2010. **285**(18): p. 13896-906.
94. Zeng, Z., et al., *Structural basis of selective ubiquitination of TRF1 by SCFFbx4.* Dev Cell, 2010. **18**(2): p. 214-25.
95. Ensminger, A.W. and R.R. Isberg, *E3 ubiquitin ligase activity and targeting of BAT3 by multiple Legionella pneumophila translocated substrates.* Infect Immun, 2010. **78**(9): p. 3905-19.
96. Ivanov, S.S. and C.R. Roy, *Modulation of ubiquitin dynamics and suppression of DALIS formation by the Legionella pneumophila Dot/Icm system.* Cell Microbiol, 2009. **11**(2): p. 261-78.
97. Angot, A., et al., *Exploitation of eukaryotic ubiquitin signaling pathways by effectors translocated by bacterial type III and type IV secretion systems.* PLoS Pathog, 2007. **3**(1): p. e3.
98. Lomma, M., et al., *The Legionella pneumophila F-box protein Lpp2082 (AnkB) modulates ubiquitination of the host protein parvin B and promotes intracellular replication.* Cell Microbiol, 2010. **12**(9): p. 1272-91.
99. Price, C.T., et al., *Molecular mimicry by an F-box effector of Legionella pneumophila hijacks a conserved polyubiquitination machinery within macrophages and protozoa.* PLoS Pathog, 2009. **5**(12): p. e1000704.
100. Kubori, T., et al., *Legionella metaeffector exploits host proteasome to temporally regulate cognate effector.* PLoS Pathog, 2010. **6**(12): p. e1001216.
101. Hsu, F., et al., *The Legionella effector SidC defines a unique family of ubiquitin ligases important for bacterial phagosomal remodeling.* Proc Natl Acad Sci U S A, 2014. **111**(29): p. 10538-43.

102. Puvar, K., et al., *Ubiquitin Chains Modified by the Bacterial Ligase SdeA Are Protected from Deubiquitinase Hydrolysis.* Biochemistry, 2017. **56**(36): p. 4762-4766.
103. Qiu, J., et al., *Ubiquitination independent of E1 and E2 enzymes by bacterial effectors.* Nature, 2016. **533**(7601): p. 120-4.
104. Wright, L.P. and M.R. Philips, *Thematic review series: lipid posttranslational modifications. CAAX modification and membrane targeting of Ras.* J Lipid Res, 2006. **47**(5): p. 883-91.
105. Ullah, N., M. Mansha, and P.J. Casey, *Protein Geranylgeranyltransferase Type 1 as a Target in Cancer.* Curr Cancer Drug Targets, 2016. **16**(7): p. 563-71.
106. Zhang, F.L. and P.J. Casey, *Protein prenylation: molecular mechanisms and functional consequences.* Annu Rev Biochem, 1996. **65**: p. 241-69.
107. Gelb, M.H., et al., *Therapeutic intervention based on protein prenylation and associated modifications.* Nat Chem Biol, 2006. **2**(10): p. 518-28.
108. Maurer-Stroh, S., S. Washietl, and F. Eisenhaber, *Protein prenyltransferases.* Genome Biol, 2003. **4**(4): p. 212.
109. Wang, M. and P.J. Casey, *Protein prenylation: unique fats make their mark on biology.* Nat Rev Mol Cell Biol, 2016. **17**(2): p. 110-22.
110. Price, C.T., et al., *Exploitation of conserved eukaryotic host cell farnesylation machinery by an F-box effector of Legionella pneumophila.* J Exp Med, 2010. **207**(8): p. 1713-26.
111. Price, C.T., et al., *Host-mediated post-translational prenylation of novel dot/icm-translocated effectors of legionella pneumophila.* Front Microbiol, 2010. **1**: p. 131.
112. Ivanov, S.S., et al., *Lipidation by the host prenyltransferase machinery facilitates membrane localization of Legionella pneumophila effector proteins.* J Biol Chem, 2010. **285**(45): p. 34686-98.
113. Trujillo, C., et al., *Essential lysine residues within transmembrane helix 1 of diphtheria toxin facilitate COPI binding and catalytic domain entry.* Mol Microbiol, 2010. **76**(4): p. 1010-9.
114. Custer, S.K., et al., *Dilysine motifs in exon 2b of SMN protein mediate binding to the COPI vesicle protein alpha-COP and neurite outgrowth in a cell culture model of spinal muscular atrophy.* Hum Mol Genet, 2013. **22**(20): p. 4043-52.
115. Gao, C., et al., *Retention mechanisms for ER and Golgi membrane proteins.* Trends Plant Sci, 2014. **19**(8): p. 508-15.
116. Price, C.T., et al., *Indispensable role for the eukaryotic-like ankyrin domains of the ankyrin B effector of Legionella pneumophila within macrophages and amoebae.* Infect Immun, 2010. **78**(5): p. 2079-88.
117. Schroeder, G.N., et al., *Legionella pneumophila strain 130b possesses a unique combination of type IV secretion systems and novel Dot/Icm secretion system effector proteins.* J Bacteriol, 2010. **192**(22): p. 6001-16.
118. Bruckert, W.M., C.T. Price, and Y. Abu Kwaik, *Rapid nutritional remodeling of the host cell upon attachment of Legionella pneumophila.* Infect Immun, 2014. **82**(1): p. 72-82.
119. Price, C.T. and Y. Abu Kwaik, *Exploitation of Host Polyubiquitination Machinery through Molecular Mimicry by Eukaryotic-Like Bacterial F-Box Effectors.* Front Microbiol, 2010. **1**: p. 122.
120. Price, C.T., et al., *Host proteasomal degradation generates amino acids essential for intracellular bacterial growth.* Science, 2011. **334**(6062): p. 1553-7.
121. Dalebroux, Z.D., R.L. Edwards, and M.S. Swanson, *SpoT governs Legionella pneumophila differentiation in host macrophages.* Mol Microbiol, 2009. **71**: p. 640-658.

122. Tamura, K., et al., *MEGA6: Molecular Evolutionary Genetics Analysis version 6.0.* Mol Biol Evol, 2013. **30**(12): p. 2725-9.
123. Yang, Z., *PAML 4: phylogenetic analysis by maximum likelihood.* Mol Biol Evol, 2007. **24**(8): p. 1586-91.
124. Xu, B. and Z. Yang, *PAMLX: a graphical user interface for PAML.* Mol Biol Evol, 2013. **30**(12): p. 2723-4.
125. Putty, K., et al., *Robustness of Helicobacter pylori infection conferred by context-variable redundancy among cysteine-rich paralogs.* PLoS One, 2013. **8**(3): p. e59560.
126. Delport, W., et al., *Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology.* Bioinformatics, 2010. **26**(19): p. 2455-7.
127. So, E.C., et al., *The Rab-binding profiles of bacterial virulence factors during infection.* J Biol Chem, 2016.
128. Otwinowski, Z. and W. Minor, *Processing of x-ray diffraction data collected in oscillation mode.* Meth.Enzymol., 1997. **276**: p. 307-326.
129. Sheldrick, G.M., *A short history of SHELX.* Acta Crystallogr A, 2008. **64**(Pt 1): p. 112-22.
130. Langer, G., et al., *Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7.* Nat Protoc, 2008. **3**(7): p. 1171-9.
131. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 355-67.
132. Zheng, N., et al., *Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex.* Nature, 2002. **416**(6882): p. 703-9.
133. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
134. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
135. Coyaud, E., et al., *BioID-based Identification of Skp Cullin F-box (SCF)beta-TrCP1/2 E3 Ligase Substrates.* Mol Cell Proteomics, 2015. **14**(7): p. 1781-95.
136. Rao, V.S., et al., *Protein-protein interaction detection: methods and analysis.* Int J Proteomics, 2014. **2014**: p. 147648.
137. Luhrmann, A., et al., *Inhibition of pathogen-induced apoptosis by a Coxiella burnetii type IV effector protein.* Proc Natl Acad Sci U S A, 2010. **107**(44): p. 18997-9001.
138. Datta, A., A. Kamthan, and M. Kamthan, *A simple protocol to detect interacting proteins by GST pull down assay coupled with MALDI or LC-MS/MS analysis.* 2015.
139. Shen, X., et al., *Targeting eEF1A by a Legionella pneumophila effector leads to inhibition of protein synthesis and induction of host stress response.* Cell Microbiol, 2009. **11**(6): p. 911-26.
140. Harper, J.W. and M.K. Tan, *Understanding cullin-RING E3 biology through proteomics-based substrate identification.* Mol Cell Proteomics, 2012. **11**(12): p. 1541-50.
141. Bruckert, W.M. and Y. Abu Kwaik, *Lysine11-linked polyubiquitination of the AnkB F-Box effector of Legionella pneumophila.* Infect Immun, 2015.
142. Mousnier, A., et al., *A new method to determine in vivo interactomes reveals binding of the Legionella pneumophila effector PieE to multiple rab GTPases.* MBio, 2014. **5**(4).
143. Price, C., et al., *Host FIH-Mediated Asparaginyl Hydroxylation of Translocated Legionella pneumophila Effectors.* Front Cell Infect Microbiol, 2017. **7**: p. 54.
144. Al-Quadan, T. and Y.A. Kwaik, *Molecular Characterization of Exploitation of the Polyubiquitination and Farnesylation Machineries of Dictyostelium Discoideum by the AnkB F-Box Effector of Legionella Pneumophila.* Front Microbiol, 2011. **2**: p. 23.

145. Al-Quadan, T., et al., *Anchoring of bacterial effectors to host membranes through host-mediated lipidation by prenylation: a common paradigm.* Trends Microbiol, 2011. **19**(12): p. 573-9.
146. Wang, K., et al., *Structural Mimicry by a Bacterial F Box Effector Hijacks the Host Ubiquitin-Proteasome System.* Structure, 2017. **doi: 10.1016/j.str.2016.12.015**(In Press).
147. Ma, W. and J. Goldberg, *Rules for the recognition of dilysine retrieval motifs by coatomer.* The EMBO Journal, 2013. **32**(7): p. 926-937.
148. Popoff, V., et al., *COPI budding within the Golgi stack.* Cold Spring Harb Perspect Biol, 2011. **3**(11): p. a005231.
149. Shi, X., et al., *Direct targeting of membrane fusion by SNARE mimicry: Convergent evolution of Legionella effectors.* Proc Natl Acad Sci U S A, 2016. **113**(31): p. 8807-12.
150. Dorer, M.S., et al., *RNA interference analysis of Legionella in Drosophila cells: exploitation of early secretory apparatus dynamics.* PLoS Pathog, 2006. **2**(4): p. e34.
151. Quaile, A.T., et al., *Molecular Characterization of LubX: Functional Divergence of the U-Box Fold by Legionella pneumophila.* Structure, 2015. **23**(8): p. 1459-69.
152. Urbanus, M.L., et al., *Diverse mechanisms of metaeffector activity in an intracellular bacterial pathogen, Legionella pneumophila.* Mol Syst Biol, 2016. **12**(12): p. 893.
153. Speir, M., et al., *Legionella pneumophila strain 130b evades macrophage cell death independent of the effector SidF in the absence of flagellin.* Front. Cell. Infect. Microbiol., 2017. **doi: 10.3389/fcimb.2017.00035**.
154. Parra, R.G., et al., *Structural and Energetic Characterization of the Ankyrin Repeat Protein Family.* PLoS Comput Biol, 2015. **11**(12): p. e1004659.
155. Li, J., A. Mahajan, and M.D. Tsai, *Ankyrin repeat: a unique motif mediating protein-protein interactions.* Biochemistry, 2006. **45**(51): p. 15168-78.
156. Kipreos, E.T. and M. Pagano, *The F-box protein family.* Genome Biol, 2000. **1**(5): p. REVIEWS3002.
157. Xu, C., et al., *Sequence-specific recognition of a PxLPxI/L motif by an ankyrin repeat tumbler lock.* Sci Signal, 2012. **5**(226): p. ra39.
158. Herbert, M.H., C.J. Squire, and A.A. Mercer, *Poxviral ankyrin proteins.* Viruses, 2015. **7**(2): p. 709-38.
159. Legate, K.R., et al., *ILK, PINCH and parvin: the tIPP of integrin signalling.* Nat Rev Mol Cell Biol, 2006. **7**(1): p. 20-31.
160. Zheng, N., et al., *Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases.* Cell, 2000. **102**(4): p. 533-9.
161. Price, C.T., A.M. Richards, and Y. Abu Kwaik, *Nutrient generation and retrieval from the host cell cytosol by intra-vacuolar Legionella pneumophila.* Front Cell Infect Microbiol, 2014. **4**: p. 111.
162. Bruckert, W.M. and Y. Abu Kwaik, *Complete and ubiquitinated proteome of the Legionella-containing vacuole within human macrophages.* J Proteome Res, 2015. **14**(1): p. 236-48.
163. Letourneur, F., et al., *Coatomer is essential for retrieval of dilysine-tagged proteins to the endoplasmic reticulum.* Cell, 1994. **79**(7): p. 1199-207.
164. Cosson, P. and F. Letourneur, *Coatomer interaction with di-lysine endoplasmic reticulum retention motifs.* Science, 1994. **263**(5153): p. 1629-31.
165. Iconomou, M. and D.N. Saunders, *Systematic approaches to identify E3 ligase substrates.* Biochem J, 2016. **473**(22): p. 4083-4101.
166. Yen, H.C., et al., *Global protein stability profiling in mammalian cells.* Science, 2008. **322**(5903): p. 918-23.

167. Pierce, N.W., et al., *Detection of sequential polyubiquitylation on a millisecond timescale.* Nature, 2009. **462**(7273): p. 615-9.
168. Tan, M.K., et al., *Parallel SCF adaptor capture proteomics reveals a role for SCFFBXL17 in NRF2 activation via BACH1 repressor turnover.* Mol Cell, 2013. **52**(1): p. 9-24.
169. Seidel, S.A., et al., *Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions.* Methods, 2013. **59**(3): p. 301-15.

APPENDIX 1

Potential AnkB interacting proteins from HEK293T

Table 4. Proteins identified in band 3 in HEK293T

Protein Name	Accession	Norm IBAQ	MW	MW on gel
Isoform 2 of Acidic leucine-rich nuclear phosphoprotein 32 family member B OS=Homo sapiens GN=ANP32B	sp|Q92688-2|AN32B_HUMAN (+1)	1.78E+07	22 kDa	25 kDa
Keratin, type I cytoskeletal 10 OS=Homo sapiens GN=KRT10 PE=1 SV=6	sp|P13645|K1C10_HUMAN	9.64E+06	59 kDa	25 kDa
Ankyrin-repeat protein B OS=Legionella pneumophila GN=ankB PE=4 SV=1	tr|A0A0A1EKG7|A0A0A1EKG7_LEGPN	6.81E+06	20 kDa	25 kDa
Cluster of Amino-terminal enhancer of split OS=Homo sapiens GN=AES PE=1 SV=4 (sp|Q08117|AES_HUMAN)	sp|Q08117|AES_HUMAN	3.48E+06	22 kDa	25 kDa
TATA box-binding protein-like protein 1 OS=Homo sapiens GN=TBPL1 PE=1 SV=1	sp|P62380|TBPL1_HUMAN	3.44E+06	21 kDa	25 kDa
ER membrane protein complex subunit 8 OS=Homo sapiens GN=EMC8 PE=1 SV=1	sp|O43402|EMC8_HUMAN (+1)	1.57E+06	24 kDa	25 kDa
Cluster of Keratin, type I cytoskeletal 14 OS=Homo sapiens GN=KRT14 PE=1 SV=4 (sp|P02533|K1C14_HUMAN)	sp|P02533|K1C14_HUMAN	1.37E+06	52 kDa	25 kDa
Isoform 2 of 60S ribosomal protein L15 OS=Homo sapiens GN=RPL15	sp|P61313-2|RL15_HUMAN (+3)	1.01E+06	17 kDa	25 kDa
Protein SETSIP OS=Homo sapiens GN=SETSIP PE=1 SV=1	sp|P0DME0|SETLP_HUMAN (+6)	7.66E+05	35 kDa	25 kDa
Cytochrome c1, heme protein, mitochondrial OS=Homo sapiens GN=CYC1 PE=1 SV=3	sp|P08574|CY1_HUMAN	5.51E+05	35 kDa	25 kDa

Protease-associated domain-containing protein 1 OS=Homo sapiens GN=PRADC1 PE=1 SV=1	sp|Q9BSG0|PADC1_HUMAN	7.18E+05	21 kDa	25 kDa
Transducin-like enhancer protein 1 OS=Homo sapiens GN=TLE1 PE=1 SV=2	sp|Q04724|TLE1_HUMAN (+2)	4.61E+05	83 kDa	25 kDa
DDB1- and CUL4-associated factor 7 OS=Homo sapiens GN=DCAF7 PE=1 SV=1	sp|P61962|DCAF7_HUMAN	4.18E+05	39 kDa	25 kDa
Proteasome subunit beta type-3 OS=Homo sapiens GN=PSMB3 PE=1 SV=2	sp|P49720|PSB3_HUMAN (+1)	3.21E+05	23 kDa	25 kDa
28 kDa heat- and acid-stable phosphoprotein OS=Homo sapiens GN=PDAP1 PE=1 SV=1	sp|Q13442|HAP28_HUMAN	2.11E+05	21 kDa	25 kDa
Isoform 2 of Coronin-1C OS=Homo sapiens GN=CORO1C	sp|Q9ULV4-2|COR1C_HUMAN (+3)	1.61E+05	54 kDa	25 kDa
Casein kinase I isoform epsilon OS=Homo sapiens GN=CSNK1E PE=1 SV=1	sp|P49674|KC1E_HUMAN	1.37E+05	47 kDa	25 kDa
Isoform 2 of Paraspeckle component 1 OS=Homo sapiens GN=PSPC1	sp|Q8WXF1-2|PSPC1_HUMAN (+2)	1.35E+05	46 kDa	25 kDa
Isoform 2 of Myc-associated zinc finger protein OS=Homo sapiens GN=MAZ	sp|P56270-2|MAZ_HUMAN	1.24E+05	51 kDa	25 kDa
GTP-binding nuclear protein Ran OS=Homo sapiens GN=RAN PE=1 SV=3	sp|P62826|RAN_HUMAN (+2)	7.27E+04	24 kDa	25 kDa

40S ribosomal protein S4, X isoform OS=Homo sapiens GN=RPS4X PE=1 SV=2	sp|P62701|RS4X_HUMAN	6.47E+04	30 kDa	25 kDa
Isoform 2 of Oxidation resistance protein 1 OS=Homo sapiens GN=OXR1	sp|Q8N573-2|OXR1_HUMAN (+5)	4.65E+04	94 kDa	25 kDa
Isoform 2 of Heat shock 70 kDa protein 1A OS=Homo sapiens GN=HSPA1A	sp|P0DMV8-2|HS71A_HUMAN (+2)	4.29E+04	64 kDa	25 kDa
Isoform 2 of Heat shock cognate 71 kDa protein OS=Homo sapiens GN=HSPA8	sp|P11142-2|HSP7C_HUMAN (+6)	3.78E+04	54 kDa	25 kDa
Isoform 2 of Keratin, type II cytoskeletal 78 OS=Homo sapiens GN=KRT78	sp|Q8N1N4-2|K2C78_HUMAN (+1)	7310	45 kDa	25 kDa

Table 5. Proteins identified in band 4 from HEK293T cells

Protein Name	**Accession**	**Norm IBAQ**	**MW**	**MW on gel**
Acidic leucine-rich nuclear phosphoprotein 32 family member A OS=Homo sapiens GN=ANP32A PE=1 SV=1	sp|P39687|AN32A_HUMAN (+1)	5.37E+07	29 kDa	30 kDa
Isoform 2 of Protein SET OS=Homo sapiens GN=SET	sp|Q01105-2|SET_HUMAN (+4)	1.64E+07	32 kDa	30 kDa
Small acidic protein OS=Homo sapiens GN=SMAP PE=1 SV=1	sp|O00193|SMAP_HUMAN	4.05E+06	20 kDa	30 kDa
Putative RNA-binding protein Luc7-like 2 OS=Homo sapiens GN=LUC7L2 PE=1 SV=2	sp|Q9Y383|LC7L2_HUMAN	2.37E+06	47 kDa	30 kDa
Cluster of Isoform Long of Proteasome subunit alpha type-1 OS=Homo sapiens	sp|P25786-2|PSA1_HUMAN [2]	1.80E+06	30 kDa	30 kDa

GN=PSMA1 (sp|P25786-2|PSA1_HUMAN)				
Isoform 2 of Nucleophosmin OS=Homo sapiens GN=NPM1	sp|P06748-2|NPM_HUMAN (+1)	1.34E+06	29 kDa	30 kDa
WD repeat-containing protein 82 OS=Homo sapiens GN=WDR82 PE=1 SV=1	sp|Q6UXN9|WDR82_HUMAN	9.71E+05	35 kDa	30 kDa
Isoform 2 of Protein FAM76A OS=Homo sapiens GN=FAM76A	sp|Q8TAV0-2|FA76A_HUMAN (+5)	9.45E+05	32 kDa	30 kDa
Isoform 2 of Protein FAM64A OS=Homo sapiens GN=FAM64A	sp|Q9BSJ6-2|FA64A_HUMAN (+2)	6.62E+05	26 kDa	30 kDa
Malonyl-CoA-acyl carrier protein transacylase, mitochondrial OS=Homo sapiens GN=MCAT PE=1 SV=2	sp|Q8IVS2|FABD_HUMAN	6.28E+05	43 kDa	30 kDa
RWD domain-containing protein 2B OS=Homo sapiens GN=RWDD2B PE=1 SV=1	sp|P57060|RWD2B_HUMAN	4.58E+05	36 kDa	30 kDa
Proteasome subunit alpha type-4 OS=Homo sapiens GN=PSMA4 PE=1 SV=1	sp|P25789|PSA4_HUMAN (+3)	3.76E+05	29 kDa	30 kDa
Ubiquitin carboxyl-terminal hydrolase isozyme L3 OS=Homo sapiens GN=UCHL3 PE=1 SV=1	sp|P15374|UCHL3_HUMAN (+2)	3.76E+05	26 kDa	30 kDa
Zinc finger BED domain-containing protein 3 OS=Homo sapiens GN=ZBED3 PE=1 SV=1	sp|Q96IU2|ZBED3_HUMAN	3.54E+05	25 kDa	30 kDa
Geminin OS=Homo sapiens GN=GMNN PE=1 SV=1	sp|O75496|GEMI_HUMAN (+3)	3.41E+05	24 kDa	30 kDa

WD repeat-containing protein 61 OS=Homo sapiens GN=WDR61 PE=1 SV=1	sp|Q9GZS3|WDR61_HUMAN	3.41E+05	34 kDa	30 kDa
Isoform 2 of ER membrane protein complex subunit 3 OS=Homo sapiens GN=EMC3	sp|Q9P0I2-2|EMC3_HUMAN (+1)	2.99E+05	25 kDa	30 kDa
Isoform 2 of Mitogen-activated protein kinase 1 OS=Homo sapiens GN=MAPK1	sp|P28482-2|MK01_HUMAN (+1)	2.86E+05	36 kDa	30 kDa
Isoform 2 of Ribonuclease P protein subunit p30 OS=Homo sapiens GN=RPP30	sp|P78346-2|RPP30_HUMAN (+2)	2.60E+05	36 kDa	30 kDa
Elongation factor 1-beta OS=Homo sapiens GN=EEF1B2 PE=1 SV=3	sp|P24534|EF1B_HUMAN	2.39E+05	25 kDa	30 kDa
Complement component 1 Q subcomponent-binding protein, mitochondrial OS=Homo sapiens GN=C1QBP PE=1 SV=1	sp|Q07021|C1QBP_HUMAN (+2)	2.20E+05	31 kDa	30 kDa
Isoform 2 of Na(+)/H(+) exchange regulatory cofactor NHE-RF2 OS=Homo sapiens GN=SLC9A3R2	sp|Q15599-2|NHRF2_HUMAN (+2)	1.85E+05	36 kDa	30 kDa
TP53-regulating kinase OS=Homo sapiens GN=TP53RK PE=1 SV=2	sp|Q96S44|PRPK_HUMAN	1.55E+05	28 kDa	30 kDa
Glutamate dehydrogenase 1, mitochondrial OS=Homo sapiens GN=GLUD1 PE=1 SV=2	sp|P00367|DHE3_HUMAN	1.51E+05	61 kDa	30 kDa
Isoform 2 of Cleavage and polyadenylation specificity factor subunit 7 OS=Homo sapiens GN=CPSF7	sp|Q8N684-2|CPSF7_HUMAN (+3)	1.51E+05	51 kDa	30 kDa

COP9 signalosome complex subunit 7a OS=Homo sapiens GN=COPS7A PE=1 SV=1	sp|Q9UBW8|CSN7A_HUMAN (+4)	1.36E+05	30 kDa	30 kDa
Isoform 2 of 2,4-dienoyl-CoA reductase, mitochondrial OS=Homo sapiens GN=DECR1	sp|Q16698-2|DECR_HUMAN (+4)	1.29E+05	35 kDa	30 kDa
Isoform 2 of Mitotic checkpoint protein BUB3 OS=Homo sapiens GN=BUB3	sp|O43684-2|BUB3_HUMAN (+2)	1.01E+05	37 kDa	30 kDa
Isoform 3 of Phosphatidylinositol 5-phosphate 4-kinase type-2 gamma OS=Homo sapiens GN=PIP4K2C	sp|Q8TBX8-3|PI42C_HUMAN (+1)	7.76E+04	45 kDa	30 kDa
Isoform 2 of Inhibitor of growth protein 4 OS=Homo sapiens GN=ING4	sp|Q9UNL4-2|ING4_HUMAN (+7)	7.01E+04	28 kDa	30 kDa
Isoform 2 of Plasminogen activator inhibitor 1 RNA-binding protein OS=Homo sapiens GN=SERBP1	sp|Q8NC51-2|PAIRB_HUMAN (+3)	4.61E+04	44 kDa	30 kDa
Pre-rRNA-processing protein TSR1 homolog OS=Homo sapiens GN=TSR1 PE=1 SV=1	sp|Q2NL82|TSR1_HUMAN	3.74E+04	92 kDa	30 kDa

Table 6. Proteins identified in band 5 from HEK293T cells

Protein Name	Accession	Norm IBAQ	MW	MW on gel
Translational activator of cytochrome c oxidase 1 OS=Homo sapiens GN=TACO1 PE=1 SV=1	sp|Q9BSH4|TACO1_HUMAN	5.42E+06	32 kDa	33 kDa
Isoform 3 of Acidic leucine-rich nuclear phosphoprotein 32 family member E OS=Homo sapiens GN=ANP32E	sp|Q9BTT0-3|AN32E_HUMAN (+2)	2.78E+06	25 kDa	33 kDa

Pyrroline-5-carboxylate reductase 3 OS=Homo sapiens GN=PYCRL PE=1 SV=3	sp|Q53H96|P5CR3_HUMAN (+1)	1.90E+06	29 kDa	33 kDa
Cluster of Isoform 2 of Mitogen-activated protein kinase 1 OS=Homo sapiens GN=MAPK1 (sp|P28482-2|MK01_HUMAN)	sp|P28482-2|MK01_HUMAN [2]	6.24E+05	36 kDa	33 kDa
40S ribosomal protein SA OS=Homo sapiens GN=RPSA PE=1 SV=4	sp|P08865|RSSA_HUMAN (+3)	5.69E+05	33 kDa	33 kDa
Casein kinase II subunit alpha OS=Homo sapiens GN=CSNK2A1 PE=1 SV=1	sp|P68400|CSK21_HUMAN (+2)	4.75E+05	45 kDa	33 kDa
Splicing factor U2AF 35 kDa subunit-like protein OS=Homo sapiens GN=U2AF1L5 PE=3 SV=1	sp|P0DN76|U2AF5_HUMAN	3.23E+05	28 kDa	33 kDa
Multiple myeloma tumor-associated protein 2 OS=Homo sapiens GN=MMTAG2 PE=1 SV=1	sp|Q9BU76|MMTA2_HUMAN	3.22E+05	29 kDa	33 kDa
Transcription initiation factor IIA subunit 1 OS=Homo sapiens GN=GTF2A1 PE=1 SV=1	sp|P52655|TF2AA_HUMAN (+1)	2.78E+05	42 kDa	33 kDa
Isoform 3 of Protein FRA10AC1 OS=Homo sapiens GN=FRA10AC1	sp|Q70Z53-3|F10C1_HUMAN	2.62E+05	37 kDa	33 kDa
Regulation of nuclear pre-mRNA domain-containing protein 1B OS=Homo sapiens GN=RPRD1B PE=1 SV=1	sp|Q9NQG5|RPR1B_HUMAN	2.50E+05	37 kDa	33 kDa
Developmentally-regulated GTP-binding protein 1 OS=Homo sapiens GN=DRG1 PE=1 SV=1	sp|Q9Y295|DRG1_HUMAN	2.11E+05	41 kDa	33 kDa

Isoform 2 of Elongation factor 1-delta OS=Homo sapiens GN=EEF1D	sp|P29692-2|EF1D_HUMAN (+14)	1.99E+05	71 kDa	33 kDa
Isoform 2 of F-actin-capping protein subunit beta OS=Homo sapiens GN=CAPZB	sp|P47756-2|CAPZB_HUMAN (+3)	1.95E+05	31 kDa	33 kDa
Pre-mRNA-splicing factor 38A OS=Homo sapiens GN=PRPF38A PE=1 SV=1	sp|Q8NAV1|PR38A_HUMAN	1.91E+05	37 kDa	33 kDa
Peptidyl-prolyl cis-trans isomerase-like 4 OS=Homo sapiens GN=PPIL4 PE=1 SV=1	sp|Q8WUA2|PPIL4_HUMAN	1.73E+05	57 kDa	33 kDa
Translin-associated protein X OS=Homo sapiens GN=TSNAX PE=1 SV=1	sp|Q99598|TSNAX_HUMAN	1.69E+05	33 kDa	33 kDa
Enhancer of mRNA-decapping protein 3 OS=Homo sapiens GN=EDC3 PE=1 SV=1	sp|Q96F86|EDC3_HUMAN	1.59E+05	56 kDa	33 kDa
Isoform 2 of Endonuclease III-like protein 1 OS=Homo sapiens GN=NTHL1	sp|P78549-2|NTH_HUMAN (+3)	1.29E+05	34 kDa	33 kDa
Isoform 2 of RNA pseudouridylate synthase domain-containing protein 3 OS=Homo sapiens GN=RPUSD3	sp|Q6P087-2|RUSD3_HUMAN (+1)	1.27E+05	37 kDa	33 kDa
BTB/POZ domain-containing protein KCTD12 OS=Homo sapiens GN=KCTD12 PE=1 SV=1	sp|Q96CX2|KCD12_HUMAN	1.20E+05	36 kDa	33 kDa
Heterogeneous nuclear ribonucleoprotein A0 OS=Homo sapiens GN=HNRNPA0 PE=1 SV=1	sp|Q13151|ROA0_HUMAN	1.18E+05	31 kDa	33 kDa

Cluster of Serine/threonine-protein phosphatase PP1-beta catalytic subunit OS=Homo sapiens GN=PPP1CB PE=1 SV=3 (sp|P62140|PP1B_HUMAN)	sp|P62140|PP1B_HUMAN	1.10E+05	37 kDa	33 kDa
Isoform 2 of 60S acidic ribosomal protein P0 OS=Homo sapiens GN=RPLP0	sp|P05388-2|RLA0_HUMAN (+2)	1.02E+05	27 kDa	33 kDa
Isoform 2 of V-type proton ATPase subunit E 1 OS=Homo sapiens GN=ATP6V1E1	sp|P36543-2|VATE1_HUMAN (+2)	1.01E+05	24 kDa	33 kDa
Isoform 2 of Septin-2 OS=Homo sapiens GN=SEPT2	sp|Q15019-2|SEPT2_HUMAN (+3)	8.81E+04	45 kDa	33 kDa
Ribose-5-phosphate isomerase OS=Homo sapiens GN=RPIA PE=1 SV=3	sp|P49247|RPIA_HUMAN	8.18E+04	33 kDa	33 kDa
Isoform 2 of Phosphatidylinositol 5-phosphate 4-kinase type-2 alpha OS=Homo sapiens GN=PIP4K2A	sp|P48426-2|PI42A_HUMAN (+1)	7.86E+04	40 kDa	33 kDa
Isoform 1 of RNA demethylase ALKBH5 OS=Homo sapiens GN=ALKBH5	sp|Q6P6C2-1|ALKB5_HUMAN (+1)	5.33E+04	52 kDa	33 kDa
Ribonucleoprotein PTB-binding 1 OS=Homo sapiens GN=RAVER1 PE=1 SV=1	tr|A0A087WZ13|A0A087WZ13_HUMAN (+1)	4.61E+04	78 kDa	33 kDa
FACT complex subunit SSRP1 OS=Homo sapiens GN=SSRP1 PE=1 SV=1	sp|Q08945|SSRP1_HUMAN (+1)	4.54E+04	81 kDa	33 kDa
Casein kinase II subunit alpha' OS=Homo sapiens GN=CSNK2A2 PE=1 SV=1	sp|P19784|CSK22_HUMAN	4.09E+04	41 kDa	33 kDa

ATP synthase subunit gamma, mitochondrial OS=Homo sapiens GN=ATP5C1 PE=1 SV=1	sp|P36542|ATPG_HUMAN	3.53E+04	33 kDa	33 kDa
Isoform 2 of Pre-B-cell leukemia transcription factor-interacting protein 1 OS=Homo sapiens GN=PBXIP1	sp|Q96AQ6-2|PBIP1_HUMAN (+1)	3.30E+04	78 kDa	33 kDa
Isoform 2 of Transducin-like enhancer protein 3 OS=Homo sapiens GN=TLE3	sp|Q04726-2|TLE3_HUMAN (+12)	2.96E+04	82 kDa	33 kDa
Erythrocyte band 7 integral membrane protein OS=Homo sapiens GN=STOM PE=1 SV=3	sp|P27105|STOM_HUMAN (+1)	2.37E+04	32 kDa	33 kDa
Isoform 3 of Ubiquitin carboxyl-terminal hydrolase 7 OS=Homo sapiens GN=USP7	sp|Q93009-3|UBP7_HUMAN (+1)	2.11E+04	126 kDa	33 kDa
Hornerin OS=Homo sapiens GN=HRNR PE=1 SV=2	sp|Q86YZ3|HORN_HUMAN	1.43E+04	282 kDa	33 kDa
Serum albumin OS=Homo sapiens GN=ALB PE=1 SV=2	sp|P02768|ALBU_HUMAN (+1)	7190	69 kDa	33 kDa

Table 7. Proteins identified in band 6 from HEK293T cells

Protein Name	**Accession**	**Norm IBAQ**	**MW**	**MW on gel**
Pyrroline-5-carboxylate reductase 2 OS=Homo sapiens GN=PYCR2 PE=1 SV=1	sp|Q96C36|P5CR2_HUMAN	7.62E+06	34 kDa	39 kDa
Isoform 3 of Pyrroline-5-carboxylate reductase 1, mitochondrial OS=Homo sapiens GN=PYCR1	sp|P32322-3|P5CR1_HUMAN (+1)	3.36E+06	36 kDa	39 kDa

Isoform 2 of RNA-binding protein with serine-rich domain 1 OS=Homo sapiens GN=RNPS1	sp|Q15287-2|RNPS1_HUMAN (+6)	7.88E+05	32 kDa	39 kDa
Splicing factor 3B subunit 4 OS=Homo sapiens GN=SF3B4 PE=1 SV=1	sp|Q15427|SF3B4_HUMAN	6.44E+05	44 kDa	39 kDa
Isoform 2 of Pyrroline-5-carboxylate reductase 3 OS=Homo sapiens GN=PYCRL	sp|Q53H96-2|P5CR3_HUMAN (+2)	5.93E+05	26 kDa	39 kDa
Isoform 2 of Protein FAM98A OS=Homo sapiens GN=FAM98A	sp|Q8NCA5-2|FA98A_HUMAN (+1)	5.77E+05	55 kDa	39 kDa
Phosphatidylinositol 5-phosphate 4-kinase type-2 beta OS=Homo sapiens GN=PIP4K2B PE=1 SV=1	sp|P78356|PI42B_HUMAN	5.44E+05	47 kDa	39 kDa
60S ribosomal protein L5 OS=Homo sapiens GN=RPL5 PE=1 SV=3	sp|P46777|RL5_HUMAN (+1)	4.32E+05	34 kDa	39 kDa
Zinc finger and SCAN domain-containing protein 26 OS=Homo sapiens GN=ZSCAN26 PE=1 SV=2	sp|Q16670|ZSC26_HUMAN (+2)	4.17E+05	55 kDa	39 kDa
Casein kinase I isoform alpha OS=Homo sapiens GN=CSNK1A1 PE=1 SV=2	sp|P48729|KC1A_HUMAN	3.94E+05	39 kDa	39 kDa
Isoform 3 of Ubiquitin carboxyl-terminal hydrolase 46 OS=Homo sapiens GN=USP46	sp|P62068-3|UBP46_HUMAN (+3)	3.22E+05	42 kDa	39 kDa
Isoform 2 of Eukaryotic peptide chain release factor subunit 1 OS=Homo sapiens GN=ETF1	sp|P62495-2|ERF1_HUMAN (+2)	2.93E+05	45 kDa	39 kDa
Isoform 3 of Serine hydroxymethyltransferase, mitochondrial OS=Homo sapiens GN=SHMT2	sp|P34897-3|GLYM_HUMAN (+1)	2.81E+05	53 kDa	39 kDa

Crk-like protein OS=Homo sapiens GN=CRKL PE=1 SV=1	sp|P46109|CRKL_HUMAN	2.54E+05	34 kDa	39 kDa
Isoform 2 of Microtubule-associated protein RP/EB family member 2 OS=Homo sapiens GN=MAPRE2	sp|Q15555-2|MARE2_HUMAN (+6)	2.47E+05	29 kDa	39 kDa
Inositol hexakisphosphate kinase 2 OS=Homo sapiens GN=IP6K2 PE=1 SV=2	sp|Q9UHH9|IP6K2_HUMAN	2.40E+05	49 kDa	39 kDa
Isoform 2 of Isovaleryl-CoA dehydrogenase, mitochondrial OS=Homo sapiens GN=IVD	sp|P26440-2|IVD_HUMAN (+2)	2.37E+05	43 kDa	39 kDa
Aurora kinase A OS=Homo sapiens GN=AURKA PE=1 SV=2	sp|O14965|AURKA_HUMAN	2.14E+05	46 kDa	39 kDa
RNA-binding protein NOB1 OS=Homo sapiens GN=NOB1 PE=1 SV=1	sp|Q9ULX3|NOB1_HUMAN	2.06E+05	47 kDa	39 kDa
Isoform 2 of Probable RNA-binding protein 23 OS=Homo sapiens GN=RBM23	sp|Q86U06-2|RBM23_HUMAN (+4)	1.96E+05	47 kDa	39 kDa
Isoform 4 of Casein kinase I isoform gamma-3 OS=Homo sapiens GN=CSNK1G3	sp|Q9Y6M4-4|KC1G3_HUMAN	1.33E+05	49 kDa	39 kDa
Isoform 2 of ATP-dependent RNA helicase DDX42 OS=Homo sapiens GN=DDX42	sp|Q86XP3-2|DDX42_HUMAN (+2)	1.30E+05	90 kDa	39 kDa
Lactadherin OS=Homo sapiens GN=MFGE8 PE=1 SV=2	sp|Q08431|MFGM_HUMAN	1.10E+05	43 kDa	39 kDa
Isoform 2 of Poly(rC)-binding protein 2 OS=Homo sapiens GN=PCBP2	sp|Q15366-2|PCBP2_HUMAN (+10)	1.02E+05	39 kDa	39 kDa
Zinc finger protein 696 OS=Homo sapiens GN=ZNF696 PE=2 SV=2	sp|Q9H7X3|ZN696_HUMAN (+1)	1.01E+05	41 kDa	39 kDa

tRNA (adenine(58)-N(1))-methyltransferase, mitochondrial OS=Homo sapiens GN=TRMT61B PE=1 SV=2	sp|Q9BVS5|TR61B_HUMAN	9.83E+04	53 kDa	39 kDa
E3 ubiquitin-protein ligase RING2 OS=Homo sapiens GN=RNF2 PE=1 SV=1	sp|Q99496|RING2_HUMAN (+1)	9.62E+04	38 kDa	39 kDa
Phenylalanine--tRNA ligase, mitochondrial OS=Homo sapiens GN=FARS2 PE=1 SV=1	sp|O95363|SYFM_HUMAN	9.49E+04	52 kDa	39 kDa
26S proteasome non-ATPase regulatory subunit 7 OS=Homo sapiens GN=PSMD7 PE=1 SV=2	sp|P51665|PSMD7_HUMAN	6.47E+04	37 kDa	39 kDa
Isoform 2 of ATP synthase subunit alpha, mitochondrial OS=Homo sapiens GN=ATP5A1	sp|P25705-2|ATPA_HUMAN (+1)	6.04E+04	54 kDa	39 kDa
Isoform 2 of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 3 OS=Homo sapiens GN=PFKFB3	sp|Q16875-2|F263_HUMAN (+6)	5.80E+04	59 kDa	39 kDa
AMME syndrome candidate gene 1 protein OS=Homo sapiens GN=AMMECR1 PE=1 SV=1	sp|Q9Y4X0|AMMR1_HUMAN	5.56E+04	35 kDa	39 kDa
Isoform 2 of DNA replication licensing factor MCM3 OS=Homo sapiens GN=MCM3	sp|P25205-2|MCM3_HUMAN (+1)	5.46E+04	96 kDa	39 kDa
Isoform C1 of Heterogeneous nuclear ribonucleoproteins C1/C2 OS=Homo sapiens GN=HNRNPC	sp|P07910-2|HNRPC_HUMAN (+8)	4.65E+04	32 kDa	39 kDa
Lupus La protein OS=Homo sapiens GN=SSB PE=1 SV=2	sp|P05455|LA_HUMAN	4.61E+04	47 kDa	39 kDa

Nucleolin OS=Homo sapiens GN=NCL PE=1 SV=3	sp|P19338|NUCL_HUMAN (+1)	4.46E+04	77 kDa	39 kDa
Protein KTI12 homolog OS=Homo sapiens GN=KTI12 PE=1 SV=1	sp|Q96EK9|KTI12_HUMAN	4.42E+04	39 kDa	39 kDa
Isoform 2 of Multifunctional protein ADE2 OS=Homo sapiens GN=PAICS	sp|P22234-2|PUR6_HUMAN (+2)	4.11E+04	48 kDa	39 kDa
T-complex protein 1 subunit beta OS=Homo sapiens GN=CCT2 PE=1 SV=4	sp|P78371|TCPB_HUMAN (+2)	3.95E+04	57 kDa	39 kDa
Isoform 2 of Septin-7 OS=Homo sapiens GN=SEPT7	sp|Q16181-2|SEPT7_HUMAN (+5)	3.46E+04	51 kDa	39 kDa
Elongation factor Tu, mitochondrial OS=Homo sapiens GN=TUFM PE=1 SV=2	sp|P49411|EFTU_HUMAN	2.06E+04	50 kDa	39 kDa
Isoform 2 of Polypeptide N-acetylgalactosaminyltransferase 2 OS=Homo sapiens GN=GALNT2	sp|Q10471-2|GALT2_HUMAN (+1)	2.00E+04	30 kDa	39 kDa
Isoform 1 of Growth factor receptor-bound protein 10 OS=Homo sapiens GN=GRB10	sp|Q13322-2|GRB10_HUMAN (+3)	1.49E+04	62 kDa	39 kDa
Isoform 2 of DnaJ homolog subfamily B member 1 OS=Homo sapiens GN=DNAJB1	sp|P25685-2|DNJB1_HUMAN (+1)	1.06E+04	27 kDa	39 kDa
DDB1- and CUL4-associated factor 13 OS=Homo sapiens GN=DCAF13 PE=1 SV=2	sp|Q9NV06|DCA13_HUMAN (+1)	6180	51 kDa	39 kDa
Isoform 2 of Sentrin-specific protease 6 OS=Homo sapiens GN=SENP6	sp|Q9GZR1-2|SENP6_HUMAN (+1)	4480	125 kDa	39 kDa

Table 8. Proteins identified in band 8 from HEK293T cells

Protein Name	Accession	Norm IBAQ	MW	MW on gel
2-hydroxyacylsphingosine 1-beta-galactosyltransferase OS=Homo sapiens GN=UGT8 PE=2 SV=2	sp|Q16880|CGT_HUMAN	1.55E+06	61 kDa	54 kDa
Lariat debranching enzyme OS=Homo sapiens GN=DBR1 PE=1 SV=2	sp|Q9UK59|DBR1_HUMAN	1.16E+06	62 kDa	54 kDa
F-box-like/WD repeat-containing protein TBL1XR1 OS=Homo sapiens GN=TBL1XR1 PE=1 SV=1	sp|Q9BZK7|TBL1R_HUMAN	1.00E+06	56 kDa	54 kDa
Isoform 2 of Creatine kinase U-type, mitochondrial OS=Homo sapiens GN=CKMT1A	sp|P12532-2|KCRU_HUMAN (+1)	8.59E+05	50 kDa	54 kDa
Isoform 2 of Replication initiator 1 OS=Homo sapiens GN=REPIN1	sp|Q9BWE0-4|REPI1_HUMAN (+3)	8.18E+05	70 kDa	54 kDa
Zinc finger protein ZIC 2 OS=Homo sapiens GN=ZIC2 PE=1 SV=2	sp|O95409|ZIC2_HUMAN (+1)	7.29E+05	55 kDa	54 kDa
Isoform 2 of Adenylosuccinate lyase OS=Homo sapiens GN=ADSL	sp|P30566-2|PUR8_HUMAN (+1)	6.93E+05	48 kDa	54 kDa
Cluster of Pyrroline-5-carboxylate reductase 2 OS=Homo sapiens GN=PYCR2 PE=1 SV=1 (sp|Q96C36|P5CR2_HUMAN)	sp|Q96C36|P5CR2_HUMAN	6.04E+05	34 kDa	54 kDa
Cluster of Isoform 2 of Transducin-like enhancer protein 3 OS=Homo sapiens GN=TLE3 (sp|Q04726-2|TLE3_HUMAN)	sp|Q04726-2|TLE3_HUMAN [9]	5.83E+05	82 kDa	54 kDa
Coronin-1B OS=Homo sapiens GN=CORO1B PE=1 SV=1	sp|Q9BR76|COR1B_HUMAN	4.96E+05	54 kDa	54 kDa

Isoform 2 of H/ACA ribonucleoprotein complex non-core subunit NAF1 OS=Homo sapiens GN=NAF1	sp|Q96HR8-2|NAF1_HUMAN (+1)	3.95E+05	42 kDa	54 kDa
Desmoglein-1 OS=Homo sapiens GN=DSG1 PE=1 SV=2	sp|Q02413|DSG1_HUMAN	3.94E+05	114 kDa	54 kDa
LanC-like protein 2 OS=Homo sapiens GN=LANCL2 PE=1 SV=1	sp|Q9NS86|LANC2_HUMAN	3.87E+05	51 kDa	54 kDa
Isoform 2 of Myc proto-oncogene protein OS=Homo sapiens GN=MYC	sp|P01106-2|MYC_HUMAN (+3)	3.13E+05	51 kDa	54 kDa
Suppressor of fused homolog OS=Homo sapiens GN=SUFU PE=1 SV=2	sp|Q9UMX1|SUFU_HUMAN	2.89E+05	54 kDa	54 kDa
Isoform 2 of Dynactin subunit 4 OS=Homo sapiens GN=DCTN4	sp|Q9UJW0-2|DCTN4_HUMAN (+2)	2.82E+05	46 kDa	54 kDa
Myotubularin-related protein 9 OS=Homo sapiens GN=MTMR9 PE=1 SV=1	sp|Q96QG7|MTMR9_HUMAN	2.58E+05	63 kDa	54 kDa
Isoform 3 of Src substrate cortactin OS=Homo sapiens GN=CTTN	sp|Q14247-3|SRC8_HUMAN (+1)	2.28E+05	57 kDa	54 kDa
Microfibrillar-associated protein 1 OS=Homo sapiens GN=MFAP1 PE=1 SV=2	sp|P55081|MFAP1_HUMAN	2.11E+05	52 kDa	54 kDa
Isoform 1 of Rab-3A-interacting protein OS=Homo sapiens GN=RAB3IP	sp|Q96QF0-2|RAB3I_HUMAN (+2)	1.84E+05	51 kDa	54 kDa
Isoform 2 of Signal recognition particle 54 kDa protein OS=Homo sapiens GN=SRP54	sp|P61011-2|SRP54_HUMAN (+2)	1.78E+05	50 kDa	54 kDa
Isoform 2 of Bifunctional arginine demethylase and lysyl-hydroxylase JMJD6 OS=Homo sapiens GN=JMJD6	sp|Q6NYC1-2|JMJD6_HUMAN (+5)	1.61E+05	43 kDa	54 kDa

Isoform 2 of Nucleolar and spindle-associated protein 1 OS=Homo sapiens GN=NUSAP1	sp|Q9BXS6-2|NUSAP_HUMAN (+6)	1.50E+05	49 kDa	54 kDa
Probable asparagine--tRNA ligase, mitochondrial OS=Homo sapiens GN=NARS2 PE=1 SV=3	sp|Q96I59|SYNM_HUMAN	1.48E+05	54 kDa	54 kDa
7SK snRNA methylphosphate capping enzyme OS=Homo sapiens GN=MEPCE PE=1 SV=1	sp|Q7L2J0|MEPCE_HUMAN	1.46E+05	74 kDa	54 kDa
DNA primase large subunit OS=Homo sapiens GN=PRIM2 PE=1 SV=2	sp|P49643|PRI2_HUMAN	1.29E+05	59 kDa	54 kDa
Isoform 2 of U4/U6.U5 tri-snRNP-associated protein 2 OS=Homo sapiens GN=USP39	sp|Q53GS9-2|SNUT2_HUMAN (+4)	1.08E+05	54 kDa	54 kDa
Isoform 2 of Protein SMG8 OS=Homo sapiens GN=SMG8	sp|Q8ND04-2|SMG8_HUMAN (+1)	7.63E+04	113 kDa	54 kDa
Parafibromin OS=Homo sapiens GN=CDC73 PE=1 SV=1	sp|Q6P1J9|CDC73_HUMAN	7.00E+04	61 kDa	54 kDa
Isoform 2 of Brain-specific angiogenesis inhibitor 1-associated protein 2 OS=Homo sapiens GN=BAIAP2	sp|Q9UQB8-2|BAIP2_HUMAN (+7)	6.87E+04	59 kDa	54 kDa
Isoform 2 of Myotubularin-related protein 12 OS=Homo sapiens GN=MTMR12	sp|Q9C0I1-2|MTMRC_HUMAN (+2)	5.16E+04	80 kDa	54 kDa
Suprabasin OS=Homo sapiens GN=SBSN PE=1 SV=2	sp|Q6UWP8|SBSN_HUMAN	3.76E+04	61 kDa	54 kDa
Isoform 3 of RNA polymerase II-associated factor 1 homolog OS=Homo sapiens GN=PAF1	sp|Q8N7H5-3|PAF1_HUMAN (+2)	3.29E+04	55 kDa	54 kDa

Isoform DPII of Desmoplakin OS=Homo sapiens GN=DSP	sp|P15924-2|DESP_HUMAN (+2)	1.62E+04	260 kDa	54 kDa
Transitional endoplasmic reticulum ATPase OS=Homo sapiens GN=VCP PE=1 SV=4	sp|P55072|TERA_HUMAN	7090	89 kDa	54 kDa

APPENDIX 2

Potential AnkB interacting proteins from U937 Cells

Table 9. Proteins identified in band 1 from U937 cells

Protein Name	Accession	Norm iBAQ	MW	MW on gel
Isoform 2 of Nucleophosmin OS=Homo sapiens GN=NPM1	sp|P06748-2|NPM_HUMAN (+2)	4.15E+06	29 kDa	17 kDa
40S ribosomal protein S18 OS=Homo sapiens GN=RPS18 PE=1 SV=3	sp|P62269|RS18_HUMAN	9.80E+05	18 kDa	17 kDa
Casein kinase II subunit alpha OS=Homo sapiens GN=CSNK2A1 PE=1 SV=1	sp|P68400|CSK21_HUMAN (+3)	4.65E+05	45 kDa	17 kDa
Nucleoplasmin-3 OS=Homo sapiens GN=NPM3 PE=1 SV=3	sp|O75607|NPM3_HUMAN	3.75E+05	19 kDa	17 kDa
Isoform SM-B of Small nuclear ribonucleoprotein-associated proteins B and B' OS=Homo sapiens GN=SNRPB	sp|P14678-2|RSMB_HUMAN (+1)	3.50E+05	24 kDa	17 kDa
60S ribosomal protein L12 OS=Homo sapiens GN=RPL12 PE=1 SV=1	sp|P30050|RL12_HUMAN	2.97E+05	18 kDa	17 kDa
Isoform 2 of Peptidyl-prolyl cis-trans isomerase H OS=Homo sapiens GN=PPIH	sp|O43447-2|PPIH_HUMAN (+2)	1.85E+05	14 kDa	17 kDa
40S ribosomal protein S15 OS=Homo sapiens GN=RPS15 PE=1 SV=2	sp|P62841|RS15_HUMAN (+7)	1.10E+05	17 kDa	17 kDa
Isoform 2 of 60S ribosomal protein L18 OS=Homo sapiens GN=RPL18	sp|Q07020-2|RL18_HUMAN (+5)	9.76E+04	18 kDa	17 kDa
Isoform 1 of Growth factor receptor-bound protein 10 OS=Homo sapiens GN=GRB10	sp|Q13322-2|GRB10_HUMAN (+3)	6.66E+04	62 kDa	17 kDa
Isoform 2 of DNA-3-methyladenine glycosylase OS=Homo sapiens GN=MPG	sp|P29372-2|3MG_HUMAN (+4)	1.22E+04	32 kDa	17 kDa
Isoform 2 of Paraspeckle component 1 OS=Homo sapiens GN=PSPC1	sp|Q8WXF1-2|PSPC1_HUMAN (+1)	9230	46 kDa	17 kDa

Isoform 2 of Obg-like ATPase 1 OS=Homo sapiens GN=OLA1	sp|Q9NTK5-2|OLA1_HUMAN (+2)	8430	28 kDa	17 kDa

Table 10. Proteins identified in band 2 from U937 cells

Protein Name	Accession	Norm iBAQ	MW	MW on gel
Isoform 2 of Acidic leucine-rich nuclear phosphoprotein 32 family member B OS=Homo sapiens GN=ANP32B	sp|Q92688-2|AN32B_HUMAN (+1)	5.94E+06	22 kDa	19 kDa
Nucleoplasmin-3 OS=Homo sapiens GN=NPM3 PE=1 SV=3	sp|O75607|NPM3_HUMAN	1.98E+06	19 kDa	19 kDa
Isoform 2 of U1 small nuclear ribonucleoprotein 70 kDa OS=Homo sapiens GN=SNRNP70	sp|P08621-2|RU17_HUMAN (+1)	6.30E+05	51 kDa	19 kDa
Ankyrin-repeat protein B OS=Legionella pneumophila GN=ankB PE=4 SV=1	tr|A0A0A1EKG7|A0A0A1EKG7_LEGPN	4.42E+05	20 kDa	19 kDa
Isoform 2 of Malignant T-cell-amplified sequence 1 OS=Homo sapiens GN=MCTS1	sp|Q9ULC4-2|MCTS1_HUMAN (+2)	2.88E+05	19 kDa	19 kDa
Isoform 2 of 60S ribosomal protein L18 OS=Homo sapiens GN=RPL18	sp|Q07020-2|RL18_HUMAN (+5)	1.09E+05	18 kDa	19 kDa
Protein SETSIP OS=Homo sapiens GN=SETSIP PE=1 SV=1	sp|P0DME0|SETLP_HUMAN (+6)	9.91E+04	35 kDa	19 kDa
Isoform 2 of ELAV-like protein 1 OS=Homo sapiens GN=ELAVL1	sp|Q15717-2|ELAV1_HUMAN	9.07E+04	39 kDa	19 kDa
Isoform 2 of DNA replication licensing factor MCM3 OS=Homo sapiens GN=MCM3	sp|P25205-2|MCM3_HUMAN (+1)	8.67E+04	96 kDa	19 kDa

Elongation factor Tu, mitochondrial OS=Homo sapiens GN=TUFM PE=1 SV=2	sp|P49411|EFTU_HUMAN	6.94E+04	50 kDa	19 kDa
Isoform 2 of E3 ubiquitin-protein ligase Hakai OS=Homo sapiens GN=CBLL1	sp|Q75N03-2|HAKAI_HUMAN (+1)	4.37E+04	54 kDa	19 kDa
Mitochondrial assembly of ribosomal large subunit protein 1 OS=Homo sapiens GN=MALSU1 PE=1 SV=1	sp|Q96EH3|MASU1_HUMAN	3.74E+04	26 kDa	19 kDa
Casein kinase II subunit beta OS=Homo sapiens GN=CSNK2B PE=1 SV=1	sp|P67870|CSK2B_HUMAN (+2)	3.25E+04	25 kDa	19 kDa
Nucleolin OS=Homo sapiens GN=NCL PE=1 SV=3	sp|P19338|NUCL_HUMAN	2.64E+04	77 kDa	19 kDa
Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15 OS=Homo sapiens GN=DHX15 PE=1 SV=2	sp|O43143|DHX15_HUMAN	2.52E+04	91 kDa	19 kDa
Zinc finger and BTB domain-containing protein 45 OS=Homo sapiens GN=ZBTB45 PE=2 SV=1	sp|Q96K62|ZBT45_HUMAN	2.07E+04	54 kDa	19 kDa
Mothers against decapentaplegic homolog 3 OS=Homo sapiens GN=SMAD3 PE=1 SV=1	sp|P84022|SMAD3_HUMAN	1.63E+04	48 kDa	19 kDa
Isoform 2 of RNA-binding protein 26 OS=Homo sapiens GN=RBM26	sp|Q5T8P6-2|RBM26_HUMAN (+5)	8680	111 kDa	19 kDa

Table 11. Proteins identified in band 3 from U937 cells

Protein Name	Accession	Norm iBAQ	MW	MW on gel
Casein kinase II subunit beta OS=Homo sapiens GN=CSNK2B PE=1 SV=1	sp|P67870|CSK2B_HUMAN (+2)	7.17E+05	25 kDa	25 kDa
Isoform 2 of Heterogeneous nuclear ribonucleoprotein A3 OS=Homo sapiens GN=HNRNPA3	sp|P51991-2|ROA3_HUMAN (+1)	3.94E+05	37 kDa	25 kDa
40S ribosomal protein SA OS=Homo sapiens GN=RPSA PE=1 SV=4	sp|P08865|RSSA_HUMAN (+2)	3.85E+05	33 kDa	25 kDa
Protein SETSIP OS=Homo sapiens GN=SETSIP PE=1 SV=1	sp|P0DME0|SETLP_HUMAN (+6)	2.25E+05	35 kDa	25 kDa
Cluster of Keratin, type I cytoskeletal 14 OS=Homo sapiens GN=KRT14 PE=1 SV=4 (sp|P02533|K1C14_HUMAN)	sp|P02533|K1C14_HUMAN	1.82E+05	52 kDa	25 kDa
Nucleolin OS=Homo sapiens GN=NCL PE=1 SV=3	sp|P19338|NUCL_HUMAN (+1)	1.56E+05	77 kDa	25 kDa
Elongation factor Tu, mitochondrial OS=Homo sapiens GN=TUFM PE=1 SV=2	sp|P49411|EFTU_HUMAN	9.69E+04	50 kDa	25 kDa
Lupus La protein OS=Homo sapiens GN=SSB PE=1 SV=2	sp|P05455|LA_HUMAN (+2)	8.48E+04	47 kDa	25 kDa
Proteasome subunit alpha type-5 OS=Homo sapiens GN=PSMA5 PE=1 SV=3	sp|P28066|PSA5_HUMAN	7.78E+04	26 kDa	25 kDa
Tyrosine--tRNA ligase, cytoplasmic OS=Homo sapiens GN=YARS PE=1 SV=4	sp|P54577|SYYC_HUMAN (+1)	7.37E+04	59 kDa	25 kDa

Isoform A1-A of Heterogeneous nuclear ribonucleoprotein A1 OS=Homo sapiens GN=HNRNPA1	sp|P09651-2|ROA1_HUMAN (+4)	7.26E+04	34 kDa	25 kDa
Splicing factor U2AF 35 kDa subunit-like protein OS=Homo sapiens GN=U2AF1L5 PE=3 SV=1	sp|P0DN76|U2AF5_HUMAN	3.81E+04	28 kDa	25 kDa
Na(+)/H(+) exchange regulatory cofactor NHE-RF1 OS=Homo sapiens GN=SLC9A3R1 PE=1 SV=4	sp|O14745|NHRF1_HUMAN (+1)	1.37E+04	39 kDa	25 kDa
sp|DHE3_BOVIN|	sp|DHE3_BOVIN| (+3)	9410	62 kDa	25 kDa
Isoform 2 of DNA replication licensing factor MCM3 OS=Homo sapiens GN=MCM3	sp|P25205-2|MCM3_HUMAN (+2)	0	96 kDa	25 kDa

Table 12. Proteins identified in band 4 from U937 cells

Protein Name	Accession	Norm iBAQ	MW	MW on gel
Cluster of Protein SET OS=Homo sapiens GN=SET PE=1 SV=3 (sp|Q01105|SET_HUMAN)	sp|Q01105|SET_HUMAN [2]	1.65E+07	33 kDa	34 kDa
Casein kinase II subunit alpha OS=Homo sapiens GN=CSNK2A1 PE=1 SV=1	sp|P68400|CSK21_HUMAN (+3)	1.53E+06	45 kDa	34 kDa
Cluster of Isoform 2 of Nucleosome assembly protein 1-like 1 OS=Homo sapiens GN=NAP1L1 (sp|P55209-2|NP1L1_HUMAN)	sp|P55209-2|NP1L1_HUMAN [11]	1.48E+06	43 kDa	34 kDa
40S ribosomal protein SA OS=Homo sapiens GN=RPSA PE=1 SV=4	sp|P08865|RSSA_HUMAN (+2)	7.86E+05	33 kDa	34 kDa
Casein kinase II subunit alpha' OS=Homo sapiens GN=CSNK2A2 PE=1 SV=1	sp|P19784|CSK22_HUMAN	4.55E+05	41 kDa	34 kDa

Isoform 2 of Glyceraldehyde-3-phosphate dehydrogenase OS=Homo sapiens GN=GAPDH	sp|P04406-2|G3P_HUMAN (+2)	3.68E+05	32 kDa	34 kDa
Eukaryotic translation initiation factor 2 subunit 1 OS=Homo sapiens GN=EIF2S1 PE=1 SV=3	sp|P05198|IF2A_HUMAN	3.64E+05	36 kDa	34 kDa
Isoform 2 of Transducin-like enhancer protein 3 OS=Homo sapiens GN=TLE3	sp|Q04726-2|TLE3_HUMAN (+11)	2.72E+05	82 kDa	34 kDa
Isoform 2 of Heterogeneous nuclear ribonucleoprotein R OS=Homo sapiens GN=HNRNPR	sp|O43390-2|HNRPR_HUMAN (+2)	2.47E+05	71 kDa	34 kDa
WD repeat-containing protein 5 OS=Homo sapiens GN=WDR5 PE=1 SV=1	sp|P61964|WDR5_HUMAN	1.39E+05	37 kDa	34 kDa
Isoform 2 of Heterogeneous nuclear ribonucleoprotein Q OS=Homo sapiens GN=SYNCRIP	sp|O60506-2|HNRPQ_HUMAN (+4)	1.38E+05	66 kDa	34 kDa
Obg-like ATPase 1 OS=Homo sapiens GN=OLA1 PE=1 SV=2	sp|Q9NTK5|OLA1_HUMAN (+1)	8.51E+04	45 kDa	34 kDa
Nucleolin OS=Homo sapiens GN=NCL PE=1 SV=3	sp|P19338|NUCL_HUMAN	8.07E+04	77 kDa	34 kDa
X-ray repair cross-complementing protein 6 OS=Homo sapiens GN=XRCC6 PE=1 SV=2	sp|P12956|XRCC6_HUMAN	5.33E+04	70 kDa	34 kDa
Tyrosine--tRNA ligase, cytoplasmic OS=Homo sapiens GN=YARS PE=1 SV=4	sp|P54577|SYYC_HUMAN (+1)	3.94E+04	59 kDa	34 kDa
Beta-adrenergic receptor kinase 1 OS=Homo sapiens GN=ADRBK1 PE=1 SV=2	sp|P25098|ARBK1_HUMAN	3.69E+04	80 kDa	34 kDa

Cluster of Beta-actin-like protein 2 OS=Homo sapiens GN=ACTBL2 PE=1 SV=2 (sp|Q562R1|ACTBL_HUMAN)	sp|Q562R1|ACTBL_HUMAN [2]	3.33E+04	42 kDa	34 kDa
Isoform 2 of Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 2 OS=Homo sapiens GN=AGAP2	sp|Q99490-2|AGAP2_HUMAN (+3)	2.32E+04	91 kDa	34 kDa
Cluster of Casein kinase I isoform gamma-2 OS=Homo sapiens GN=CSNK1G2 PE=1 SV=1 (sp|P78368|KC1G2_HUMAN)	sp|P78368|KC1G2_HUMAN	2.23E+04	47 kDa	34 kDa
DNA replication licensing factor MCM5 OS=Homo sapiens GN=MCM5 PE=1 SV=5	sp|P33992|MCM5_HUMAN (+1)	1.62E+04	82 kDa	34 kDa
Isoform 3 of Phosphatidylinositol 5-phosphate 4-kinase type-2 gamma OS=Homo sapiens GN=PIP4K2C	sp|Q8TBX8-3|PI42C_HUMAN (+1)	1.16E+04	45 kDa	34 kDa
Isoform 2 of Splicing factor 1 OS=Homo sapiens GN=SF1	sp|Q15637-2|SF01_HUMAN (+6)	1.09E+04	69 kDa	34 kDa
Isoform 2 of Exosome component 10 OS=Homo sapiens GN=EXOSC10	sp|Q01780-2|EXOSX_HUMAN (+1)	7480	98 kDa	34 kDa
Isoform 2 of Phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 1 OS=Homo sapiens GN=INPP5D	sp|Q92835-2|SHIP1_HUMAN (+3)	5880	133 kDa	34 kDa
RNA-binding protein 33 OS=Homo sapiens GN=RBM33 PE=1 SV=3	sp|Q96EV2|RBM33_HUMAN	3050	130 kDa	34 kDa

Table 13. Proteins identified in band 5.1 from U937 cells

Protein Name	Accession	Norm iBAQ	MW	MW on gel
Cluster of Protein SET OS=Homo sapiens GN=SET PE=1 SV=3 (sp|Q01105|SET_HUMAN)	sp|Q01105|SET_HUMAN [2]	2.05E+07	33 kDa	35 kDa
Casein kinase II subunit alpha OS=Homo sapiens GN=CSNK2A1 PE=1 SV=1	sp|P68400|CSK21_HUMAN (+3)	6.46E+06	45 kDa	35 kDa
Casein kinase II subunit alpha' OS=Homo sapiens GN=CSNK2A2 PE=1 SV=1	sp|P19784|CSK22_HUMAN	2.75E+06	41 kDa	35 kDa
Isoform 2 of Nucleosome assembly protein 1-like 1 OS=Homo sapiens GN=NAP1L1	sp|P55209-2|NP1L1_HUMAN (+13)	1.19E+06	43 kDa	35 kDa
Cluster of Isoform 2 of Heterogeneous nuclear ribonucleoprotein R OS=Homo sapiens GN=HNRNPR (sp|O43390-2|HNRPR_HUMAN)	sp|O43390-2|HNRPR_HUMAN [3]	4.83E+05	71 kDa	35 kDa
Ankyrin-repeat protein B OS=Legionella pneumophila GN=ankB PE=4 SV=1	tr|A0A0A1EKG7|A0A0A1EKG7_LEGPN	4.33E+05	20 kDa	35 kDa
Eukaryotic translation initiation factor 2 subunit 1 OS=Homo sapiens GN=EIF2S1 PE=1 SV=3	sp|P05198|IF2A_HUMAN	4.02E+05	36 kDa	35 kDa
40S ribosomal protein SA OS=Homo sapiens GN=RPSA PE=1 SV=4	sp|P08865|RSSA_HUMAN (+3)	3.37E+05	33 kDa	35 kDa
Isoform 2 of Nucleosome assembly protein 1-like 4 OS=Homo sapiens GN=NAP1L4	sp|Q99733-2|NP1L4_HUMAN (+5)	3.23E+05	44 kDa	35 kDa
Isoform 2 of Heterogeneous nuclear ribonucleoprotein L OS=Homo sapiens GN=HNRNPL	sp|P14866-2|HNRPL_HUMAN (+3)	3.03E+05	51 kDa	35 kDa

Elongation factor Tu, mitochondrial OS=Homo sapiens GN=TUFM PE=1 SV=2	sp|P49411|EFTU_HUMAN	2.31E+05	50 kDa	35 kDa
LanC-like protein 2 OS=Homo sapiens GN=LANCL2 PE=1 SV=1	sp|Q9NS86|LANC2_HUMAN	1.73E+05	51 kDa	35 kDa
Tyrosine--tRNA ligase, cytoplasmic OS=Homo sapiens GN=YARS PE=1 SV=4	sp|P54577|SYYC_HUMAN	1.64E+05	59 kDa	35 kDa
Isoform 2 of Heterogeneous nuclear ribonucleoprotein D0 OS=Homo sapiens GN=HNRNPD	sp|Q14103-2|HNRPD_HUMAN (+7)	1.52E+05	36 kDa	35 kDa
Nucleolin OS=Homo sapiens GN=NCL PE=1 SV=3	sp|P19338|NUCL_HUMAN	1.11E+05	77 kDa	35 kDa
DDB1- and CUL4-associated factor 7 OS=Homo sapiens GN=DCAF7 PE=1 SV=1	sp|P61962|DCAF7_HUMAN	1.10E+05	39 kDa	35 kDa
Phosphatidylinositol 5-phosphate 4-kinase type-2 beta OS=Homo sapiens GN=PIP4K2B PE=1 SV=1	sp|P78356|PI42B_HUMAN	1.08E+05	47 kDa	35 kDa
Isoform 2 of Phosphatidylinositol 5-phosphate 4-kinase type-2 alpha OS=Homo sapiens GN=PIP4K2A	sp|P48426-2|PI42A_HUMAN (+1)	9.98E+04	40 kDa	35 kDa
Casein kinase I isoform gamma-2 OS=Homo sapiens GN=CSNK1G2 PE=1 SV=1	sp|P78368|KC1G2_HUMAN	9.57E+04	47 kDa	35 kDa
Obg-like ATPase 1 OS=Homo sapiens GN=OLA1 PE=1 SV=2	sp|Q9NTK5|OLA1_HUMAN (+1)	8.26E+04	45 kDa	35 kDa
Isoform 2 of Methylmalonate-semialdehyde dehydrogenase [acylating],	sp|Q02252-2|MMSA_HUMAN (+1)	6.41E+04	56 kDa	35 kDa

mitochondrial OS=Homo sapiens GN=ALDH6A1				
DNA replication licensing factor MCM5 OS=Homo sapiens GN=MCM5 PE=1 SV=5	sp|P33992|MCM5_HUMAN (+1)	5.73E+04	82 kDa	35 kDa
Probable asparagine--tRNA ligase, mitochondrial OS=Homo sapiens GN=NARS2 PE=1 SV=3	sp|Q96I59|SYNM_HUMAN	5.05E+04	54 kDa	35 kDa
Eukaryotic translation initiation factor 5B OS=Homo sapiens GN=EIF5B PE=1 SV=4	sp|O60841|IF2P_HUMAN (+1)	3.92E+04	139 kDa	35 kDa
Isoform 2 of Homer protein homolog 3 OS=Homo sapiens GN=HOMER3	sp|Q9NSC5-2|HOME3_HUMAN (+3)	2.56E+04	39 kDa	35 kDa
Isoform 2 of Splicing factor 1 OS=Homo sapiens GN=SF1	sp|Q15637-2|SF01_HUMAN (+6)	1.73E+04	69 kDa	35 kDa
Mothers against decapentaplegic homolog 4 OS=Homo sapiens GN=SMAD4 PE=1 SV=1	sp|Q13485|SMAD4_HUMAN (+1)	1.53E+04	60 kDa	35 kDa
Isoform 3 of Phosphatidylinositol 5-phosphate 4-kinase type-2 gamma OS=Homo sapiens GN=PIP4K2C	sp|Q8TBX8-3|PI42C_HUMAN (+1)	1.48E+04	45 kDa	35 kDa
X-ray repair cross-complementing protein 6 OS=Homo sapiens GN=XRCC6 PE=1 SV=2	sp|P12956|XRCC6_HUMAN (+1)	1.10E+04	70 kDa	35 kDa
E3 ubiquitin-protein ligase CBL OS=Homo sapiens GN=CBL PE=1 SV=2	sp|P22681|CBL_HUMAN (+2)	9750	100 kDa	35 kDa

Table 14. Proteins identified in band 5.2 from U937 cells

Protein Name	Accession	Norm iBAQ	MW	MW on gel
Cluster of Protein SET OS=Homo sapiens GN=SET PE=1 SV=3 (sp|Q01105|SET_HUMAN)	sp|Q01105|SET_HUMAN [2]	1.74E+07	33 kDa	37 kDa
Putative RNA-binding protein Luc7-like 2 OS=Homo sapiens GN=LUC7L2 PE=1 SV=2	sp|Q9Y383|LC7L2_HUMAN	9.32E+06	47 kDa	37 kDa
Elongation factor 1-alpha 1 OS=Homo sapiens GN=EEF1A1 PE=1 SV=1	sp|P68104|EF1A1_HUMAN (+1)	7.53E+06	50 kDa	37 kDa
sp|TRYP_PIG|	sp|TRYP_PIG|	6.08E+06	24 kDa	37 kDa
Cluster of Isoform 2 of Nucleosome assembly protein 1-like 1 OS=Homo sapiens GN=NAP1L1 (sp|P55209-2|NP1L1_HUMAN)	sp|P55209-2|NP1L1_HUMAN [11]	2.22E+06	43 kDa	37 kDa
sp|K1C9_HUMAN|	sp|K1C9_HUMAN| (+1)	1.98E+06	62 kDa	37 kDa
Casein kinase II subunit alpha OS=Homo sapiens GN=CSNK2A1 PE=1 SV=1	sp|P68400|CSK21_HUMAN (+3)	1.73E+06	45 kDa	37 kDa
Non-POU domain-containing octamer-binding protein OS=Homo sapiens GN=NONO PE=1 SV=4	sp|Q15233|NONO_HUMAN	1.09E+06	54 kDa	37 kDa
Splicing factor 3A subunit 1 OS=Homo sapiens GN=SF3A1 PE=1 SV=1	sp|Q15459|SF3A1_HUMAN	6.26E+05	89 kDa	37 kDa
Isoform Short of Splicing factor, proline- and glutamine-rich OS=Homo sapiens GN=SFPQ	sp|P23246-2|SFPQ_HUMAN (+1)	5.98E+05	72 kDa	37 kDa
Na(+)/H(+) exchange regulatory cofactor NHE-RF1 OS=Homo sapiens GN=SLC9A3R1 PE=1 SV=4	sp|O14745|NHRF1_HUMAN	5.36E+05	39 kDa	37 kDa

Glutamate dehydrogenase 1, mitochondrial OS=Homo sapiens GN=GLUD1 PE=1 SV=2	sp|P00367|DHE3_HUMAN	3.60E+05	61 kDa	37 kDa
40S ribosomal protein SA OS=Homo sapiens GN=RPSA PE=1 SV=4	sp|P08865|RSSA_HUMAN (+2)	2.92E+05	33 kDa	37 kDa
Isoform 2 of Methylmalonate-semialdehyde dehydrogenase [acylating], mitochondrial OS=Homo sapiens GN=ALDH6A1	sp|Q02252-2|MMSA_HUMAN (+1)	2.24E+05	56 kDa	37 kDa
Chitinase-3-like protein 1 OS=Homo sapiens GN=CHI3L1 PE=1 SV=2	sp|P36222|CH3L1_HUMAN	2.12E+05	43 kDa	37 kDa
Isoform 2 of Heterogeneous nuclear ribonucleoprotein D0 OS=Homo sapiens GN=HNRNPD	sp|Q14103-2|HNRPD_HUMAN (+7)	2.02E+05	36 kDa	37 kDa
Fructose-bisphosphate aldolase A OS=Homo sapiens GN=ALDOA PE=1 SV=2	sp|P04075|ALDOA_HUMAN (+2)	1.65E+05	39 kDa	37 kDa
Isoform 2 of Isovaleryl-CoA dehydrogenase, mitochondrial OS=Homo sapiens GN=IVD	sp|P26440-2|IVD_HUMAN (+2)	9.23E+04	43 kDa	37 kDa
DNA replication licensing factor MCM5 OS=Homo sapiens GN=MCM5 PE=1 SV=5	sp|P33992|MCM5_HUMAN (+1)	9.15E+04	82 kDa	37 kDa
Elongation factor Tu, mitochondrial OS=Homo sapiens GN=TUFM PE=1 SV=2	sp|P49411|EFTU_HUMAN	9.06E+04	50 kDa	37 kDa
Isoform 2 of Mitotic checkpoint protein BUB3 OS=Homo sapiens GN=BUB3	sp|O43684-2|BUB3_HUMAN (+2)	8.85E+04	37 kDa	37 kDa
Nucleolin OS=Homo sapiens GN=NCL PE=1 SV=3	sp|P19338|NUCL_HUMAN	8.38E+04	77 kDa	37 kDa

NAD-dependent malic enzyme, mitochondrial OS=Homo sapiens GN=ME2 PE=1 SV=1	sp|P23368|MAOM_HUMAN	6.84E+04	65 kDa	37 kDa
Obg-like ATPase 1 OS=Homo sapiens GN=OLA1 PE=1 SV=2	sp|Q9NTK5|OLA1_HUMAN (+1)	6.35E+04	45 kDa	37 kDa
Actin, cytoplasmic 1 OS=Homo sapiens GN=ACTB PE=1 SV=1	sp|P60709|ACTB_HUMAN (+1)	6.01E+04	42 kDa	37 kDa
Isoform 1 of Growth arrest-specific protein 7 OS=Homo sapiens GN=GAS7	sp|O60861-1|GAS7_HUMAN (+2)	5.74E+04	47 kDa	37 kDa
Isoform MBP-1 of Alpha-enolase OS=Homo sapiens GN=ENO1	sp|P06733-2|ENOA_HUMAN (+1)	5.69E+04	37 kDa	37 kDa
LanC-like protein 1 OS=Homo sapiens GN=LANCL1 PE=1 SV=1	sp|O43813|LANC1_HUMAN (+1)	5.57E+04	45 kDa	37 kDa
LanC-like protein 2 OS=Homo sapiens GN=LANCL2 PE=1 SV=1	sp|Q9NS86|LANC2_HUMAN	5.29E+04	51 kDa	37 kDa
Isoform 10 of Abl interactor 1 OS=Homo sapiens GN=ABI1	sp|Q8IZP0-10|ABI1_HUMAN (+12)	5.20E+04	43 kDa	37 kDa
Isoform 2 of NAD kinase OS=Homo sapiens GN=NADK	sp|O95544-2|NADK_HUMAN (+3)	4.89E+04	63 kDa	37 kDa
Probable asparagine--tRNA ligase, mitochondrial OS=Homo sapiens GN=NARS2 PE=1 SV=3	sp|Q96I59|SYNM_HUMAN	4.70E+04	54 kDa	37 kDa
Isoform 2 of Phosphatidylinositol 5-phosphate 4-kinase type-2 alpha OS=Homo sapiens GN=PIP4K2A	sp|P48426-2|PI42A_HUMAN (+1)	4.59E+04	40 kDa	37 kDa
Isoform 2 of Interferon regulatory factor 2-binding protein 2 OS=Homo sapiens GN=IRF2BP2	sp|Q7Z5L9-2|I2BP2_HUMAN (+1)	4.18E+04	59 kDa	37 kDa

U3 small nucleolar RNA-associated protein 18 homolog OS=Homo sapiens GN=UTP18 PE=1 SV=3	sp|Q9Y5J1|UTP18_HUMAN	3.96E+04	62 kDa	37 kDa
DNA polymerase beta OS=Homo sapiens GN=POLB PE=1 SV=3	sp|P06746|DPOLB_HUMAN (+2)	3.94E+04	38 kDa	37 kDa
Tyrosine--tRNA ligase, cytoplasmic OS=Homo sapiens GN=YARS PE=1 SV=4	sp|P54577|SYYC_HUMAN (+1)	2.53E+04	59 kDa	37 kDa
Beta-adrenergic receptor kinase 1 OS=Homo sapiens GN=ADRBK1 PE=1 SV=2	sp|P25098|ARBK1_HUMAN	2.48E+04	80 kDa	37 kDa
Isoform 3 of Bifunctional lysine-specific demethylase and histidyl-hydroxylase MINA OS=Homo sapiens GN=MINA	sp|Q8IUF8-3|MINA_HUMAN (+1)	2.09E+04	24 kDa	37 kDa
Tyrosyl-DNA phosphodiesterase 1 OS=Homo sapiens GN=TDP1 PE=1 SV=2	sp|Q9NUW8|TYDP1_HUMAN (+2)	1.89E+04	68 kDa	37 kDa
Paired amphipathic helix protein Sin3a OS=Homo sapiens GN=SIN3A PE=1 SV=2	sp|Q96ST3|SIN3A_HUMAN (+1)	1.88E+04	145 kDa	37 kDa
Isoform 2 of Septin-2 OS=Homo sapiens GN=SEPT2	sp|Q15019-2|SEPT2_HUMAN (+3)	9780	45 kDa	37 kDa
Phosphatidylinositol 5-phosphate 4-kinase type-2 beta OS=Homo sapiens GN=PIP4K2B PE=1 SV=1	sp|P78356|PI42B_HUMAN	7790	47 kDa	37 kDa

Table 15. Proteins identified in band 6 from U937 cells

Protein Name	Accession	Norm iBAQ	MW	MW on gel
Protein SET OS=Homo sapiens GN=SET PE=1 SV=3	sp|Q01105|SET_HUMAN	1.42E+07	33 kDa	43 kDa
Isoform 2 of Cleavage and polyadenylation specificity factor subunit 6 OS=Homo sapiens GN=CPSF6	sp|Q16630-2|CPSF6_HUMAN (+3)	3.14E+06	63 kDa	43 kDa
Isoform 2 of Nucleosome assembly protein 1-like 1 OS=Homo sapiens GN=NAP1L1	sp|P55209-2|NP1L1_HUMAN (+10)	2.99E+06	43 kDa	43 kDa
Cluster of Nucleolin OS=Homo sapiens GN=NCL PE=1 SV=3 (sp|P19338|NUCL_HUMAN)	sp|P19338|NUCL_HUMAN	1.36E+06	77 kDa	43 kDa
Tyrosine--tRNA ligase, cytoplasmic OS=Homo sapiens GN=YARS PE=1 SV=4	sp|P54577|SYYC_HUMAN	1.07E+06	59 kDa	43 kDa
Isoform 2 of Nucleosome assembly protein 1-like 4 OS=Homo sapiens GN=NAP1L4	sp|Q99733-2|NP1L4_HUMAN (+3)	9.00E+05	44 kDa	43 kDa
Isoform 2 of Heterogeneous nuclear ribonucleoprotein L OS=Homo sapiens GN=HNRNPL	sp|P14866-2|HNRPL_HUMAN (+3)	7.06E+05	51 kDa	43 kDa
Probable cysteine--tRNA ligase, mitochondrial OS=Homo sapiens GN=CARS2 PE=1 SV=1	sp|Q9HA77|SYCM_HUMAN	3.95E+05	62 kDa	43 kDa
Lupus La protein OS=Homo sapiens GN=SSB PE=1 SV=2	sp|P05455|LA_HUMAN	2.83E+05	47 kDa	43 kDa

Phenylalanine--tRNA ligase, mitochondrial OS=Homo sapiens GN=FARS2 PE=1 SV=1	sp|O95363|SYFM_HUMAN	2.75E+05	52 kDa	43 kDa
Probable asparagine--tRNA ligase, mitochondrial OS=Homo sapiens GN=NARS2 PE=1 SV=3	sp|Q96I59|SYNM_HUMAN	2.39E+05	54 kDa	43 kDa
Eukaryotic translation initiation factor 5B OS=Homo sapiens GN=EIF5B PE=1 SV=4	sp|O60841|IF2P_HUMAN (+1)	2.03E+05	139 kDa	43 kDa
Isoform 2 of Methyltransferase-like protein 17, mitochondrial OS=Homo sapiens GN=METTL17	sp|Q9H7H0-2|MET17_HUMAN (+3)	1.86E+05	50 kDa	43 kDa
Isoform 3 of Phosphatidylinositol 5-phosphate 4-kinase type-2 gamma OS=Homo sapiens GN=PIP4K2C	sp|Q8TBX8-3|PI42C_HUMAN (+1)	1.85E+05	45 kDa	43 kDa
E3 ubiquitin-protein ligase CBL OS=Homo sapiens GN=CBL PE=1 SV=2	sp|P22681|CBL_HUMAN (+2)	1.20E+05	100 kDa	43 kDa
Coiled-coil domain-containing protein 6 OS=Homo sapiens GN=CCDC6 PE=1 SV=2	sp|Q16204|CCDC6_HUMAN	1.15E+05	53 kDa	43 kDa
Tyrosyl-DNA phosphodiesterase 1 OS=Homo sapiens GN=TDP1 PE=1 SV=2	sp|Q9NUW8|TYDP1_HUMAN (+2)	1.03E+05	68 kDa	43 kDa
Isoform 4 of Squamous cell carcinoma antigen recognized by T-cells 3 OS=Homo sapiens GN=SART3	sp|Q15020-4|SART3_HUMAN (+2)	9.59E+04	106 kDa	43 kDa
Eukaryotic initiation factor 4A-III OS=Homo sapiens GN=EIF4A3 PE=1 SV=4	sp|P38919|IF4A3_HUMAN	7.59E+04	47 kDa	43 kDa

Beta-adrenergic receptor kinase 1 OS=Homo sapiens GN=ADRBK1 PE=1 SV=2	sp|P25098|ARBK1_HUMAN	7.56E+04	80 kDa	43 kDa
Nucleolar protein 58 OS=Homo sapiens GN=NOP58 PE=1 SV=1	sp|Q9Y2X3|NOP58_HUMAN	7.28E+04	60 kDa	43 kDa
Isoform 5 of Protein disulfide-isomerase A6 OS=Homo sapiens GN=PDIA6	sp|Q15084-5|PDIA6_HUMAN	4.56E+04	53 kDa	43 kDa
Coronin-1A OS=Homo sapiens GN=CORO1A PE=1 SV=4	sp|P31146|COR1A_HUMAN (+3)	3.44E+04	51 kDa	43 kDa
Endoplasmic reticulum resident protein 44 OS=Homo sapiens GN=ERP44 PE=1 SV=1	sp|Q9BS26|ERP44_HUMAN	0	47 kDa	43 kDa

Table 16. Proteins identified in band 8 from U937 cells

Protein Name	Accession	Norm iBAQ	MW	MW on gel
X-ray repair cross-complementing protein 5 OS=Homo sapiens GN=XRCC5 PE=1 SV=3	sp|P13010|XRCC5_HUMAN	6.68E+05	83 kDa	65 kDa
Aspartate--tRNA ligase, mitochondrial OS=Homo sapiens GN=DARS2 PE=1 SV=1	sp|Q6PI48|SYDM_HUMAN	2.54E+05	74 kDa	65 kDa
Replication protein A 70 kDa DNA-binding subunit OS=Homo sapiens GN=RPA1 PE=1 SV=2	sp|P27694|RFA1_HUMAN	1.01E+05	68 kDa	65 kDa
Squamous cell carcinoma antigen recognized by T-cells 3 OS=Homo sapiens GN=SART3 PE=1 SV=1	sp|Q15020|SART3_HUMAN (+1)	6.94E+04	110 kDa	65 kDa

Staphylococcal nuclease domain-containing protein 1 OS=Homo sapiens GN=SND1 PE=1 SV=1	sp|Q7KZF4|SND1_HUMAN	6.43E+04	102 kDa	65 kDa
X-ray repair cross-complementing protein 6 OS=Homo sapiens GN=XRCC6 PE=1 SV=2	sp|P12956|XRCC6_HUMAN (+1)	5.32E+04	70 kDa	65 kDa
Bromodomain-containing protein 4 OS=Homo sapiens GN=BRD4 PE=1 SV=2	sp|O60885|BRD4_HUMAN	3.72E+04	152 kDa	65 kDa
Stress-70 protein, mitochondrial OS=Homo sapiens GN=HSPA9 PE=1 SV=2	sp|P38646|GRP75_HUMAN	2.37E+04	74 kDa	65 kDa
Asparagine--tRNA ligase, cytoplasmic OS=Homo sapiens GN=NARS PE=1 SV=1	sp|O43776|SYNC_HUMAN	1.85E+04	63 kDa	65 kDa
Ubiquitin-associated protein 2 OS=Homo sapiens GN=UBAP2 PE=1 SV=1	sp|Q5T6F2|UBAP2_HUMAN	1.20E+04	117 kDa	65 kDa

APPENDIX 3

Potential AnkB interacting proteins two-step immunoprecipitation

Table 17. Potential AnkB interacting proteins identified in region 1 (related to fig. 6)

Protein	MW kDa	AnkB			Control		
		Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
Desmoglein-1	114	1	1		2	4	3
Cluster of Actin, cytoplasmic 1	42			8			
Isoform 2 of E3 ubiquitin-protein ligase HUWE1	480	2	7				
Alpha-2-macroglobulin	163	12					
Complement C3	187	8					
Serum albumin	69	5		2			
Polyubiquitin-B	26	1	2		2	2	
Immunoglobulin heavy constant alpha	38	2		1			
Isoform DSPIa of Desmoplakin	279					1	
Serotransferrin	77	3					
Isoform 2 of Fibrinogen alpha chain	70	2					
Isoform 2 of Glyceraldehyde-3-phosphate dehydrogenase	32			2			
Immunoglobulin kappa constant	12	2					
Immunoglobulin heavy constant gamma 1	36	3					
Isoform 2 of Haptoglobin	38	1					
Isoform 2 of Gelsolin	81			2			
Isoform 2 of Immunoglobulin heavy constant mu	52	3					
Hemoglobin subunit beta	16			2			
Apolipoprotein A-I	31	1					

Table 17 (continued). Potential AnkB interacting proteins identified in region 1 (related to fig. 6)

Protein	MW kDa	AnkB			Control		
		Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
Filaggrin-2	248	1					
Isoform H14 of Myeloperoxidase	74			2			
Alpha-enolase	47			1			
Annexin A1	39			2			
Isoform Gamma-A of Fibrinogen gamma chain	49	1					
Fibrinogen beta chain	56	1					
Immunoglobulin lambda-like polypeptide	23	1					
Cluster of Isoform 2 of Semenogelin-1	45						1
Isoform 2 of Transketolase	69			2			
Immunoglobulin heavy constant gamma 2	36	1					

Table 18. Potential AnkB interacting proteins identified in region 2 (related to fig. 6)

Protein	MW kDa	AnkB			Control		
		Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
Isoform 2 of E3 ubiquitin-protein ligase HUWE1	480	4	4	12			
Ubiquitin-40S ribosomal protein S27a	18	3	8	3			1
Isoform DSPIa of Desmoplakin	279		7				
Desmoglein-1	114	3	5	1			
Junction plakoglobin	82		6				
Isoform 1B of Desmocollin-1	94	1	3				
Isoform 2 of Glyceraldehyde-3-phosphate dehydrogenase	32		2				
Hornerin	282		2				
POTE ankyrin domain family member F	121		1				

Table 19. Potential AnkB interacting proteins identified in region 3 (related to fig. 6)

Protein	MW kDa	AnkB			Control		
		Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
Isoform 12 of Titin	3816				1		
Isoform 2 of Protein piccolo	531	1					
Polyubiquitin-B	26		2				
Isoform 2 of Methylosome protein 50	30					2	
Desmoglein-1	114					1	

Table 20. Potential AnkB interacting proteins identified in region 4 (related to fig. 6)

Protein	MW kDa	AnkB			Control		
		Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
Desmoplakin	332			21	2		
Desmoglein-1	114	1		10	4		
Ubiquitin-60S ribosomal protein L40	15	3	4	3	1	1	
AnkB of *L. pneumophila*	35	4	6	6			
40S ribosomal protein S3	27	2	4		1	3	
Junction plakoglobin	82			11			
Isoform 1B of Desmocollin-1	94			5			
ATP synthase subunit gamma, mitochondrial	33		2				
Isoform 2 of Glyceraldehyde-3-phosphate dehydrogenase	32			4			
Isoform 1 of Plakophilin-1	80			4			
U1 small nuclear ribonucleoprotein A	31						3
Isoform 2 of Annexin A2	40			2			
Corneodesmosin	52			2			
Filaggrin-2	248	1					
Protein-glutamine gamma-glutamyltransferase E	77			2			
Hornerin	282			2			
Isoform 2 of Extracellular matrix protein 1	46			1			
Bleomycin hydrolase	53			1			
Isoform 3 of Protein KIAA0100	45						1

APPENDIX 4

Permission to Reproduce Published Articles

1. Structural Mimicry by a Bacterial F Box Effector Hijacks the Host Ubiquitin-Proteasome System.

2. Divergent evolution of Di-lysine ER retention vs. farnesylation motif mediated anchoring of the AnkB virulence effector to the Legionella-containing vacuolar membrane.

CURRICULUM VITA

John David Perpich, Pharm.D., Ph.D.
1009 Ambridge Dr
Louisville, KY 40207
Cell: 843.324.0824, john.perpich@gmail.com

EDUCATION

1994 – 1999 **Bachelor of Science (Medical Technology)**
With Honors, Michigan State University; East Lansing, MI

1999 – 2001 **Master of Science (Clinical Laboratory Sciences)**
Michigan State University; East Lansing, MI
Mentor: Leonel Mendoza

2003 – 2007 **Doctor of Pharmacy**
Medical University of South Carolina (MUSC); Charleston, SC
Completed Advanced Clinical Track

2013 – 2018 **Doctor of Philosophy (Microbiology & Immunology)**
University of Louisville; Louisville, KY
Mentor: Yousef Abu Kwaik

PUBLICATIONS & ABSTRACTS

Publications **Maintaining the Transcription Factor SpoIIID Level Late during Sporulation Causes Spore Defects in *Bacillus subtilis***
Wang L, **Perpich J**, Driks A, Kroos L
J Bacteriol. 2007 Oct;189(20):7302-9. Epub 2007 Aug 10.

One Perturbation of the Mother Cell Gene Regulatory Suppresses the Effects of Another During Sporulation of *Bacillus subtilis*
Wang L, **Perpich J**, Driks A, Kroos L
J Bacteriol. 2007 Dec;189(23):8467-73. Epub 2007 Sep 21.

Ixabepilone (Ixempra®): First in a new class of chemotherapy agents used for metastatic breast cancer.
March 2008 edition of The Capsule Intermountain Medical Center hospital newsletter.

Abiraterone Acetate in Castrate-Recurrent Prostate Cancer
John Perpich and Bradley Atkinson
J Adv Pract Oncol, 2011; 2:390-395

Structural Mimicry by a Bacterial F Box Effector Hijacks the Host Ubiquitin-Proteasome System
Wong K, **Perpich JD**, Kozlov G, Cygler M, Abu Kwaik Y, Gehring K. Structure. 2017 Feb 7;25(2):376-383. doi: 10.1016/j.str.2016.12.015. Epub 2017 Jan 19.

Divergent Evolution of Di-lysine ER Retention vs. Farnesylation Motif-mediated Anchoring of the AnkB Virulence Effector to the *Legionella*-containing Vacuolar Membrane
Perpich JD, Kalia A, Price CTD, Jones SC, Wong K, Gehring K, Kwaik YA
Sci Rep. 2017 Jul 11;7(1):5123. doi: 10.1038/s41598-017-05211-5.

Abstracts/Posters

Cloning and Characterization of the Chitin Synthase Gene in the Human and Mammalian Pathogen *Pythium insidiosum*
Perpich JD, Herr RA, and Mendoza AL
99th Annual General Meeting of the American Society for Microbiology; Chicago, IL May 2000

Genetic Regulatory Networks in *Bacillus subtilis*
Perpich JD and Kroos L
Genes in Development and Disease Focus Group, Michigan State University; East Lansing, MI April 2003

Clinical and Economic Analysis of Linezolid vs. Comparators in the Treatment of Methicillin-resistant *Staphylococcus aureus* (MRSA) Infections in Clinical Practice
Perpich JD, Benson, J
American Society of Health Systems Pharmacists (ASHP) Midyear Clinical Meeting, December 2007

Divergent Evolution of Di-lysine ER Retention vs. Farnesylation Motif-mediated Anchoring of the AnkB Effector to the *Legionella*-containing Vacuolar Membrane
Perpich JD, Kalia A, Price CTD, Jones SC, Wong K, Gehring K, Kwaik YA
24th Annual Midwest Microbial Pathogenesis Conference, University of Notre Dame; Notre Dame, IN August 2017.

RESIDENCY PROJECTS

2007-2008 **Clinical and Economic Analysis of Linezolid vs. Comparators in the Treatment of Methicillin-resistant Staphylococcus Aureus (MRSA) Infections in Clinical Practice**
PGY1 Research Project, Intermountain Medical Center; Salt Lake City, UT

2008-2009 **Safety, Tolerability, and Cost Analysis of Rapid Rituximab Infusions**
PGY2 Research Project, M. D. Anderson Cancer Center; Houston, TX

PROFESSIONAL EXPERIENCE

2005 – 2007 **Nuclear Pharmacy Intern**
Low Country Diagnostics; Charleston, SC

2007 – 2008 **PGY1 Pharmacy Residency- ASHP Accredited**
Intermountain Medical Center; Salt Lake City, UT

2008 – 2009 **PGY2 Oncology Pharmacy Residency- ASHP Accredited**
The University of Texas M. D. Anderson Cancer Center; Houston, TX

2009 – 2011 **Clinical Pharmacy Specialist**
The University of Texas M. D. Anderson Cancer Center; Houston, TX

2011 – 2012 **Oncology Pharmacist**
Norton Cancer Institute; Louisville, KY

TEACHING EXPERIENCE

1999 – 2001 **Teaching Assistant**
Hematology Laboratory for Medical Technology undergraduates
Medical Technology Program
Michigan State University; East Lansing, MI

June-Aug 2000 **Teaching Assistant**
Molecular Biology Laboratory for international scholars
Michigan State University; East Lansing, MI

Jan-April 2007 **Group Tutor**
Biochemistry for Doctor of Pharmacy candidates
Medical University of South Carolina; Charleston, SC

2014 – 2015 **Graduate Teaching Academy**
Participated in nine 2-hour workshops to enhance teaching skills
University of Louisville; Louisville, KY

2016 - 2017 **Guest Lecturer**
Research Methods course for doctoral students
Department of Microbiology and Immunology
University of Louisville; Louisville, KY

2017 – 2018 **Active Learning Sessions**
Developed and facilitated active learning in General Microbiology
Teaching Innovation Learning Lab (TILL)
University of Louisville; Louisville, KY

PRESENTATIONS

May 2006 **Potentially Inappropriate Prescribing in the Elderly**
Physician In-service. University of Louisville Hospital; Louisville, KY

June 2006 **Utility of Stimulant Medications in the Surgical Intensive Care Unit**
Nursing In-service.
University of Louisville Hospital; Louisville, KY

August 2006 **Use of ^{90}Y-ibritumomab tiuxetan, Zevalin®, in the Treatment of Non-Hodgkin's Lymphoma**
Grand Rounds, Medical University of South Carolina; Charleston, SC

August 2006 **Quadrivalent Human Papilloma Virus (HPV) Vaccine, Gardasil®, for the Prevention of HPV-Associated Cancers**
Monograph Presentation to the Pharmacy and Therapeutics Committee. Medical University of South Carolina; Charleston, SC

October 2007 **Sitagliptin (Januvia®) for Type 2 Diabetes**
Prepared and Presented a Drug Monograph to the Pharmacy and Therapeutics Committee, Intermountain Health Care, PGY-1 Residency; Salt Lake City, UT

April 2008 **Review of Immunology**
Professional development seminar at Intermountain Medical Center;
Salt Lake City, UT

March 2009 **Management of *bcr-abl*+ Leukemias**
ACPE Continuing Education Seminar, M. D. Anderson Cancer Center; Houston, TX

2016 – 2018 **Host Protein Targets of the AnkB Effector of *Legionella pneumophila***
Department of Microbiology and Immunology Seminar Series, University of Louisville; Louisville, KY