Manipulation of host ubiquitination by the intracellular pathogen, *Legionella pneumophila*

Kathy Wong

Department of Biochemistry

McGill University, Montreal

April 2018

A thesis submitted to McGill University in partial fulfillment

of the requirements of the degree of Doctor of Philosophy

For Mom and Dad

Table of contents

List of figures

List of tables

List of abbreviations

2K – Y91K/L93K AnkB mutant

3K – Y91K/L93K/Y127K AnkB mutant

4K – Y91K/L93K/Y127K/L134K AnkB mutant

ADP-Rib-Ub – adenosine diphosphate ribosylated ubiquitin

Ank – ankyrin

ANKRA – ankyrin repeat family A

AR – ankyrin repeat

AdoMet MTase – S-adenosyl methionine-dependent methyl transferase

ADP – adenosine diphosphate

ADP-Rib-Ub – adenosine diphosphate ribosylated ubiquitin

AMP – adenosine monophosphate

ATP – adenosine triphosphate

ATPase – adenosine triphosphatase

AUC – analytical ultracentrifugation

BCL – B-cell lymphoma

bis-PNPP – bis(para-nitrophenyl) phosphate

BMDM – bone marrow-derived macrophages

C – chemical shift perturbation

cAMP – cyclic adenosine monophosphate

cGMP – cyclic guanosine monophosphate

CHD – Cys-His-Asp-

C_{max} – chemical shift perturbation at saturation

Cya – calmodulin-dependent adenylate cyclase

Dot/Icm – defective for organelle trafficking/intracellular multiplication

DUB – deubiquitinase

EDTA – ethylenediaminetetraacetic acid

ER – endoplasmic reticulum

FBS – fetal bovine serum

FBXO – F-box only

GAL8 – galectin-8

GDP – guanosine diphosphate

GEF – guanine exchange factor

GTP – guanosine triphosphate

HA – hemagglutinin

HD – histidine aspartate

HECT – Homologous to the E6AP Carboxyl Terminus

hMDM – human monocytes-derived macrophages

HSQC – ^{15}N-^{1}H heteronuclear single quantum correlation

IBR – In-Between-RING

IPTG – isopropyl β-D-1-thiogalactopyranoside

ISG15 – interferon-stimulated gene 15

K_d – dissociation constant

KLAMP – bacterial histidine kinase-like ATP-binding region-containing proteins and S-adenosylmethionine-dependent methyltransferase proteins

LD – Legionnaires' disease

LC3 – microtubule-associated protein 1 light chain 3

LCV – *Legionella*-containing vacuole

LIR – LC3-interacting region

lmj – *Leishmania major*

lpg – *Legionella pneumophila* Philadelphia-1 gene

Lpn – Legionella pneumophila

LPS – lipopolysaccharides

LRR – leucine-rich repeats

L_{tot} – total ligand concentration

LubX – *Legionella* U-box protein

mART – mono ADP-ribosyltransferase

MOI – multiplicity of infection

MR – molecular replacement

NAD – nicotinamide adenine dinucleotide

NADPH – nicotinamide adenine dinucleotide phosphate

NEDD8 – neural precursor cell expressed, developmentally downregulated 8

NMR – nuclear magnetic resonance

NS – not significant

P-Rib-Ub – phosphoribosylated ubiquitin

ParvB – Parvin B

PDE – phosphodiesterase

PE – phosphatidylethanolamine

PGE_2 – prostaglandin E2

PKA – protein kinase A

PMSF – phenylmethylsulfonyl fluoride

PNP – para-nitrophenol

P_{tot} – total concentration of labeled protein

RavZ – region allowing vacuole colocalization gene Z

RBR – Ring-in-Between-Ring

RING – Really Interesting New Gene

RMSD – root mean square deviation

ROI – reactive oxygen intermediate

Rtn4 – reticulon 4

SAD – single-wavelength anomalous dispersion

SCF – Skp1-Cullin-F-box

SD – standard deviation

Sde – SidE

SDS-PAGE – sodium dodecyl sulfate polyacrylamide gel electrophoresis

Sid – substrate of Icm/Dot transporter

SUMO – small ubiquitin-related modifier

T4(A/B)SS – type 4(A/B) secretion system

TCA – tricarboxylic acid

TRIM – tripartite motif

Ub – ubiquitin

Ulp – ubiquitin-like protease

WT – wild type

Abstract

Legionella pneumophila is the causative agent of Legionnaire's disease, an acute form of pneumonia. It is often transmitted by inhaling contaminated aerosols with symptoms resembling other forms of pneumonia, including fevers, chills, and coughs. Pathogenic gram-negative bacteria use specialized secretions systems that translocate bacterial proteins, termed effectors, directly into host cells to interact with host proteins and hijack eukaryotic biochemical processes for the benefit of the pathogen. The secreted effectors allow the bacterium to create a replicative niche, the *Legionella*-containing vacuole (LCV), and escape the host lysosomal pathway after phagocytosis. Currently, about 300 effectors of *L. pneumophila* have been identified. However, many are redundant and the functions of most remain unknown.

My studies focus on a subset of effectors that act mediate protein ubiquitination, a process that does not exist in bacteria. I first examined protein mimicry through Ankyrin B (AnkB). The *L. pneumophila* genome codes for a large number of eukaryotic-like proteins. AnkB contains an N-terminal F-box domain and a C-terminal Ank domain. F-box containing proteins form part of SCF E3 ubiquitin ligase complexes, which transfer ubiquitin from the E2 ubiquitin conjugating enzyme to a target. By recruiting poly-ubiquitinated substrates to be degraded into free amino acids, AnkB provides a source of energy. In eukaryotes, the F-box is typically associated with substrate interaction domains. However, the association of an ankyrin domain with an F-box is novel. Chapter two presents the first crystal structure of AnkB in complex with human Skp1. We identify a putative substrate-binding site, which we confirm by mutagenesis and *in vivo* functional assays to be critical for maturation of the LCV and pathogen survival.

Chapter three focuses on another Ank effector, AnkC. Despite overall similarity to previously characterized Ank repeats, AnkC uses its repeats as a novel dimerization module which

was observed in the crystal structure and confirmed by analytical ultracentrifugation. Interestingly, the dimerization interface exists along the back of the Ank groove, allowing dimerization to occur without interfering its ability to bind potential partners. Obtaining the Ank domain structure of AnkC is an important step toward elucidating its function.

I then studied a novel ubiquitin modification mediated by the SidE family of effectors. Capable of phosphoribosylation of ubiquitin, SidE proteins use a mono-ADP-ribosyl transferase (mART) domain and NAD to generate ADP-ribosylated ubiquitin. The SidE phosphodiesterase (PDE) domain then cleaves off AMP to attach phosphoribosylated ubiquitin to a substrate. The mechanisms of this reaction remain unsolved. In chapter four, I characterize lpg1496, a SidE-related member. We determine the structure of its three domains, a conserved PDE domain and two novel KLAMP domains. Through mutagenesis studies and NMR studies, we show that all three domains bind nucleotides with differing specificities and identify the key residues responsible. Finally, in chapter five, I demonstrate that lpg1496 does not function in the same reaction as the SidE family, although it still functions as a general PDE. I reveal that the PDE and mART activities of the SidE proteins are able to complement each other in trans.

This thesis encompasses the first structural characterizations of several *Legionella* effectors. Expanding knowledge on the molecular basis of host-pathogen interactions will contribute to a better understanding of the pathogenesis of *L. pneumophila*, and development of antivirulence drugs to treat Legionnaires' disease.

Resumé

Legionella pneumophila (*Lpn*) est l'agent infectieux qui cause la maladie du légionnaire, une forme de pneumonie aiguë. Elle est souvent transmise par l'inhalation d'aérosols contaminés. Les bactéries pathogènes à Gram-négatif utilisent des systèmes de sécrétion spécialisés qui transloquent des protéines bactériennes, nommées effecteurs, directement dans les cellules hôtes. Ces protéines interagissent avec les protéines de la cellule hôte et détournent les processus biochimiques cellulaires eucaryotiques à l'avantage du pathogène. Les effecteurs permettent aux bactéries de créer une niche de réplication, la vacuole contenant la *Legionella* (VCL), et d'échapper au système lysosomal. Présentement, autour de 300 effecteurs de *Lpn* ont été identifiés. Toutefois, plusieurs sont redondants et les fonctions de la majorité d'entre eux restent inconnues.

J'ai concentré mes efforts sur un sous-ensemble d'effecteurs qui agissent sur l'ubiquitination de l'hôte, un processus qui n'existe pas chez les bactéries. Premièrement, j'ai examiné le mimétisme de l'hôte dans le cas de AnkyrinB (AnkB). Le génome de *Lpn* contient un grand nombre de protéines semblables aux eucaryotes. AnkB contient un domaine F-box et un domaine Ank. Les protéines contenant un domaine F-box forment une partie des complexes d'ubiquitine ligase E3 SCF, qui transfèrent l'ubiquitine vers un substrat. En recrutant des substrats poly-ubiquitinés pour être dégradés en acides aminés libres, AnkB stimule la production d'énergie. Dans le chapitre deux, je présente la première structure cristallographique de AnkB en complexe avec la protéine humaine Skp1. Nous avons identifié un site putatif de liaison de substrat que nous avons confirmé par des essais de mutagenèse et des essais fonctionnels *in vivo* comme étant critique pour la maturation de la VCL et la survie du pathogène.

Le chapitre trois se concentre sur un autre effecteur Ank, soit AnkC. Malgré la similarité au domaine Ank caractérisé antérieurement, AnkC utilise ses répétitions comme un module de

dimérisation jusqu'ici inconnu. Ceci a été observé dans la structure cristallographique. Il est à noter que l'interface de dimérisation est à l'arrière de la rainure Ank, ce qui permet à la dimérisation de survenir sans interférer avec la capacité de lier des substrats potentiels. La fonction d'AnkC reste inconnue, mais la structure cristallographique de son domaine Ank est une étape importante pour la réalisation de cet objectif.

J'ai ensuite étudié une nouvelle modification de l'ubiquitine médiée par l'entremise de la famille d'effecteurs SidE. Les protéines SidE utilisent premièrement un domaine mono(ADP-ribosyl) transférase (mART) et NAD pour générer l'ubiquitine ADP-ribosylée. Ensuite, son domaine phosphodiestérase (PDE) sépare AMP pour attacher l'ubiquitine phosphoribosylée à un substrat. Le mécanisme de cette réaction reste à élucider. Dans le chapitre quatre, je caractérise lpg1496, une protéine orpheline apparentée a SidE. Nous avons identifié et déterminé la structure de ses trois domaines, un domaine PDE conservé et deux nouveaux domaines KLAMP. Nous avons démontré par des expériences de mutagenèse et de RMN que les trois domaines lient des nucléotides avec des spécificités différentes et nous avons identifié les résidus clés. Finalement, dans le chapitre 5, je démontre que lpg1496 n'utilise pas le même mécanisme d'action que la famille SidE, mais qu'il fonctionne tout de même comme une PDE générale. Je révèle aussi que les habiletés de PDE et de mART des protéines SidE peuvent se complémenter en trans.

Cette thèse comporte les premières caractérisations cristallographiques de plusieurs effecteurs de *Lpn*. L'accumulation de connaissances sur les interactions entre hôte et pathogène déjà un niveau moléculaire contribuera à une meilleure compréhension de la virulence de *Lpn* et mènera au développement de nouvelles thérapies pour la maladie du légionnaire.

Acknowledgements

I would like to thank Dr. Kalle Gehring, for continuous guidance over the years, from my first years volunteering, to being an undergraduate honors student, and throughout my graduate studies. My research advisory committee members, Dr. Albert Berghuis, Dr. Samantha Gruenheid, Dr. Sébastian Faucher, who have given encouragement and constructive advice on my research. This thesis relied on the help of many collaborators. Dr. Miroslaw Cygler, the coordinator of the Montreal-Kingston Bacterial Structural Genomics Initiative, has supplied the lab with initial plasmids to begin structural studies. John D. Perpich from Dr. Abu Kwaik's lab, for the *in vivo* work in chapter two. I am grateful for financial support from Quebec Network for Research on Protein Structure, Function, and Engineering (PROTEO), Groupe de Recherche Axé sur la Structure des Protéines (GRASP), CIHR Canadian Graduate Scholarship, NSERC CREATE Training Program in Bionanomachines, and the Faculty of Medicine at McGill University.

After being a part of this lab for eight years, there are many people I am grateful for having met. Dr. Guennadi Kozlov, for supervising me throughout my graduate project, helping me develop crystallography skills, always being open for discussion and for having a much better memory than me when I am regularly tempted to repeat those malachite green assays. Thank you to Dr. Marie Ménade and Dr. Véronique Sauvé who had the patience to teach a second-year undergraduate student all the basic lab skills that set me up for the rest of my studies. In addition to helping me consistently with cloning issues, I especially thank Marie for allowing me to have colorful stationary that made keeping track of my research all the more fun. Thank you to Dr. Jingwei Xie for teaching me how to handle cells. Even though those localization assays will always be a mystery, they gave me a few extra slides for lab meeting and I had a great time working with you.

My first few years were a joy thanks to Irina Gulerez, Asparouh Lilov and Crystal Li. I am fortunate to have made lifelong friends and even though we are now mostly in different cities, I really cherish when we are able to catch up over food, coffee, board games, and that one thing that has become a tradition between me and Crystal. Irina, I miss our walks home together, no matter how quick of a walk it was. Asparouh, for translating my abstract to French. It is always fun whenever you visit and somehow a certain topic is brought up. Dr. Juliana Muñoz Escobar, for silly times, gossip times, serious scientific times, that are continuing post- your graduation. Thank you for telling me to go home and have a life on those late nights. Something that I have now passed on to Seby Chen. Seby, it is always humbling that you do not trust any stocks that I have made or touched. Dr. Meng Yang, for continuing to accompany me to the EM facility when I should not be asking anymore. Dr. Marjan Seirafi, I will always remember those talks we had about starting graduate studies. Many undergraduate students have contributed to my project over the years, spending precious studying hours helping me in the lab. Thank you to Michelle Zhang, Jean Luo, Si Jia Wang and Howard Li.

I had the opportunity to spend time in two exchange labs. From Dr. Sébastian Faucher's lab, Nilmini Mendis showed me how to work with *Legionella*. From Dr. Ralph Isberg's lab, Dr. Kristin Kotewicz spent many hours helping me generate mutant strains and checking plates. I am lucky to have made many memories and new friends, Karin Yoshida, Ila Anand and Stacie Clark.

Most importantly, I cannot imagine having the luxury to complete this chapter of my life without support from my loving family, Mom, Dad, Big Bro, Candy and Jonathon. Somehow four years away turned into nine.

Author contributions

This is a manuscript-based thesis consisting of three articles.

Chapter 2: Hijacking by an F-box effector

Wong, K., Perpich, J.D., Kozlov, G., Cygler, M., Abu Kwaik, Y., and Gehring, K. (2017). Structural mimicry by a bacterial F-box effector hijacks the host ubiquitin-proteasome system. Structure *25*, 376-383.

Structural work was designed, performed and analyzed by myself and Dr. Guennadi Kozlov. Initial constructs of AnkB were obtained from Dr. Miroslaw Cygler. Infection assays with hMDMs and U937 macrophages were performed by John D. Perpich in the lab of Dr. Yousef Abu Kwaik. Appropriate sections were written by myself and John D. Perpich. All authors contributed to the final version of the paper.

Chapter 3: Ankyrin repeats as a dimerization module

Kozlov, G.*, Wong, K.*, Wang, W., Skubák, P., Muñoz-Escobar, J., Liu, Y., Siddiqui, N., Pannu, N.S., and Gehring, K. (2017) Ankyrin repeats as a dimerization module. Biochem Biophys Res Commun doi: 10.1016/j.bbrc.2017.11.135 * co-first authors

Cloning of constructs and crystallization trials were performed by myself, Dr. Guennadi Kozlov, Wenxuan Wang, and Yue Liu. Dr. Juliana Muñoz-Escobar collected the diffraction data at the Advanced Photon Source facility in Chicago. Use of a novel program, PRASA, to help solve the structure of AnkC, was performed by Dr. Pavol Skubàk and Dr. Navraj S. Pannu. Analytical ultracentrifugation data was collected and analyzed by Dr. Nadeem Siddiqui. Appropriate sections were written by myself, Dr. Kozlov, Dr. Skubàk, Dr. Pannu and Dr. Siddiqui. All authors contributed to the final version of the paper.

Chapter 4: An orphan SidE-related member

Wong, K., Kozlov, G., Zhang, Y., and Gehring, K. (2015) Structure of the *Legionella* effector, lpg1496, suggests a role in nucleotide metabolism. J Biol Chem *290*, 24727-37.

I performed and analyzed NMR titration experiments and crystallized six of the seven reported structures. Dr. Guennadi Kozlov assigned the backbone amides of KLAMP2 using NMR. Yinglu Zhang crystallized KLAMP2 in the $P2_1$ space group. I designed and performed activity assays. The manuscript was written by myself and Dr. Kozlov and edited by Dr. Gehring.

Original scholarship

Chapter 2: Hijacking by an F-box effector

I determined the crystal structure of the ankyrin repeats domain of AnkB, and full length AnkB in complex with its human binding partner, Skp1. This revealed how *Legionella* is able to hijack host ubiquitination machinery. From the structure, several residues were chosen based on their potential to affect AnkB function. I generated the mutants and confirmed their stability by NMR. These mutants were then given to Dr. Kwaik's lab for *in vivo* testing, confirming their importance in survival of the pathogen.

Chapter 3: Ankyrin repeats as a dimerization module

I cloned constructs used for the crystallization of AnkC and was involved in initial crystallization trials. I helped in analysis of the data, and in preparation of the manuscript.

Chapter 4: An orphan SidE-related member

I determined the crystal structure of all three domains of lpg1496, of KLAMP1 bound to 3',5'-cAMP and of the PDE domain bound to ADP. This revealed the molecular determinants of nucleotide binding by each of the three domains. The KLAMP domains were identified as novel folds. I performed NMR titrations of the two KLAMP domains with a collection of different nucleotides to determine differing specificities. I performed malachite green assays using lpg1496 and showed that it does not have activity against common nucleotides.

Chapter 1: Introduction

1.1 Legionnaires' disease

During the 58th convention of the American Legion in Philadelphia of 1976, a flu-like disease spread and infected 182 legionnaires (Fraser et al., 1977). Of those infected, 29 died. Half a year later, the gram-negative *Legionella* bacterium was finally identified in isolates obtained from the air conditioning system of the hotel that housed the convention (McDade et al., 1977). Common symptoms include bradycardia, coughing, diarrhea and confusion. Since its discovery, *Legionella* is increasingly recognized as a severe cause of both community-acquired pneumonia and nosocomial (hospital) acquired pneumonia (Cunha and Cunha, 2017). The bacterium is usually transmitted through contaminated water sources in the form of aerosols and causes two forms of infection: Legionnaires' disease (LD), an atypical pneumonia, and the milder, more uncommon, Pontiac fever, a fever not presenting with pneumonia (McDade et al., 1977). However, there has been one reported case of human-to-human transmission in Portugal between a mother and her son who had contracted LD 300 km away (Correia et al., 2016).

Patients with a healthy immune response system are able to effectively eliminate the pathogen. Mainly the elderly and immunocompromised individuals are at risk of developing a severe form of LD (Shin, 2012). Current treatment involves broad-spectrum antibiotics. Azithromycin targets the 50S subunit of the bacterial ribosome, blocking progression of bacterial protein synthesis (Jelic and Antolovic, 2016). Rifampin is often used in combination therapy in more severe cases of LD, inhibiting bacterial RNA polymerase, preventing DNA transcription (Wehrli, 1983). Even with effective bactericidal drugs, the mortality rate of LD still exists above 10% (Descours et al., 2017).

1.2 Life cycle of *Legionella pneumophila*

Currently, at least 58 different species of *Legionella* have been discovered. About half of those species have been associated with clinical cases. Over 90% of reported LD cases in the United States are caused by *L. pneumophila* (*Lpn*), making it the most well-studied species (Fields et al., 2002). *L. micdadei* is the second most common causative agent of LD in the United States and Europe (Reingold et al., 1984; Roig et al., 2003). In Australia, *L. longbeachae* causes about half of reported LD cases (Doyle and Heuzenroeder, 2002).

Lpn replicates naturally in a diverse range of amoeba such as *Acanthomeoba polyphaga* and *Dictyostelium discoideum* (Boamah et al., 2017). However, the focus of this section will be on its intracellular replication within the alveolar macrophages of human hosts. Once phagocytosed, avirulent bacteria are effectively delivered to the lysosome, obtaining early and late markers such as Rab5, Rab7 and LAMP-1 (Allombert et al., 2013). The phagolysosome is an acidic environment (~pH 4.5) filled with hydrolytic enzymes to destroy the pathogen (**Fig. 1.1**) (Scott et al., 2003). Virulent *Lpn* inhibits fusion of the phagosome with bactericidal lysosomes and creates the *Legionella*-containing vacuole (LCV). Within minutes of uptake, vesicles originating from the endoplasmic reticulum (ER), and mitochondria are recruited to the newly formed vacuole. This replicative niche has a pH around 6.1, is eventually decorated with ribosomes and within this environment, *Lpn* is able to replicate efficiently (**Fig. 1.1**) (Horwitz and Maxfield, 1984; Swanson and Isberg, 1995). When nutrients become limiting, the bacteria transform into a mature infectious form and are released from the original host cell to infect neighboring cells (Robertson et al., 2014).

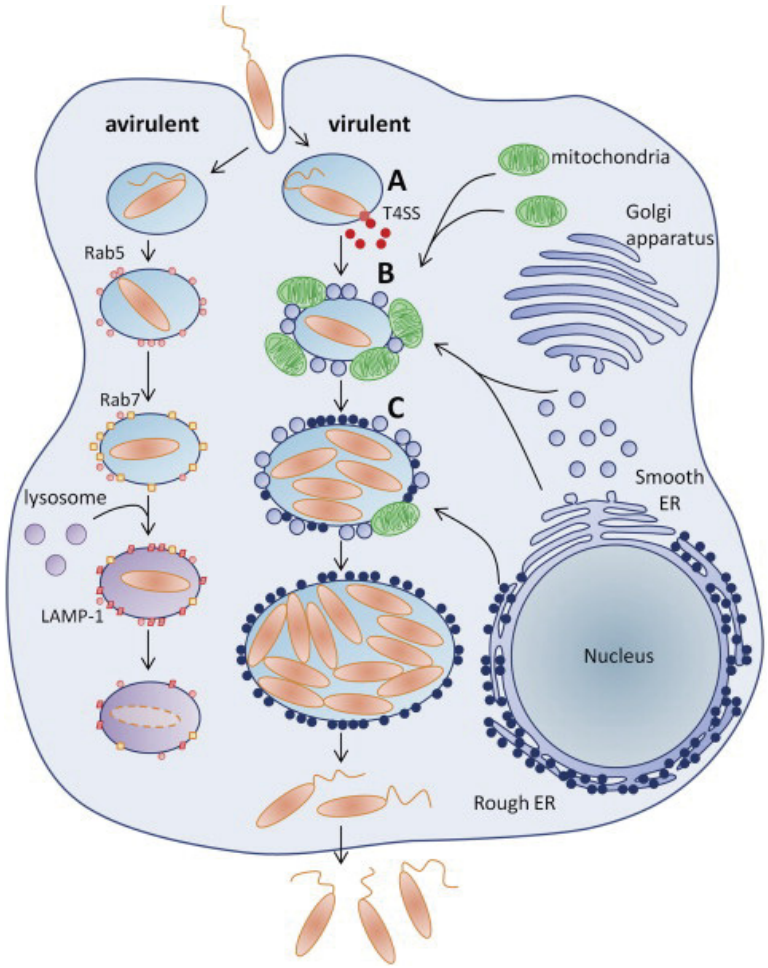


Figure 1.1 Life cycle of *Legionella pneumophila*. *Lpn* is phagocytosed by a host macrophage. Avirulent bacteria, such as Dot/Icm deficient strains (see 1.2.1) fuse with lysosomes to form an acidic phagolysosome and are cleared by hydrolytic enzymes. Virulent bacteria avoid lysosomal fusion by secreting virulence factors through their T4SS (type 4 secretion system) and recruiting organelles and vesicles from the Golgi and ER to form a *Legionella*-containing vacuole conducive to intracellular replication. When nutrients are depleted, avirulent *Lpn* are released from the host cell to infect neighboring cells. Adapted from (Allombert et al., 2013).

1.3 The Dot/Icm Type 4 secretion system

Intracellular replication of *Lpn* is dependent on a functional Dot/Icm (defective for organelle trafficking/intracellular multiplication) Type 4 secretion system (T4SS). This system allows the bacteria to translocate its own bacterial proteins, termed effectors, into the host cell to

aid in the infection process. T4SSs are found in multiple bacterial species including *Helicobacter pylori* and *Escherichia coli*, and can be divided into two major subfamilies, type 4A and 4B. The T4ASS system is composed of 12 subunits, termed VirB1 to B11 and VirD4. A channel through the inner and outer membranes is formed by VirB3, VirB6, VirB8, and VirB7, VirB9, VirB10, respectively. Energy is generated by VirB1, VirB4, VirB11 and VirD4. A pilus for direct contact with the host is composed of VirB2 and VirB5 (Chandran Darbari and Waksman, 2015).

The type 4B system (T4BSS) is present in *Lpn*. It is a more complex and unique system with 26 dot/icm proteins involved. A similar system exists only in two other pathogens, *Coxiella burnetii*, the causative agent of Q fever, and the anthropod pathogen *Rickettsiella grylli* (Nagai and Kubori, 2011). This secretion system is essential for replication in both amoeba and macrophage hosts, where a *dotA* mutant results in insufficient remodeling of the LCV for escape from lysosomal fusion (Segal and Shuman, 1999). The only sequence similarity between the two T4SSs exists in the Trbl domain of VirB10 and DotG. The 4B core complex consists of an outer membrane channel formed by DotC, DotD and DotH to cross the periplasm, into an inner membrane channel formed by DotG and DotT. Energy is generated by the DotB, DotL and DotI ATPases (Nagai and Kubori, 2011).

An *in-situ* structure of the *Legionella* T4BSS was recently solved by electron cryotomography. While this system shares little sequence similarity with T4ASS, its overall structure is highly similar to a previously reported *Escherichia coli* T4ASS VirB3 to VirB10 negative stain reconstruction. The T4ASS consists of a smaller periplasmic complex, linked by a stalk to a larger inner-membrane complex with two barrel-shaped densities reaching into the cytoplasm (**Fig. 1.2A**) (Low et al., 2014). The *Lpn* T4BSS is approximately twice as large. A "hat" density inserted into the outer membrane attached to α and β densities is attached to a stem and a

stalk decorated with γ densities in the periplasm. "Wing" densities extend from the inner membrane into the periplasm. Multiple rod-like densities were also observed below the inner membrane, into the cytoplasm (**Fig. 1.2B**) (Ghosal et al., 2017).

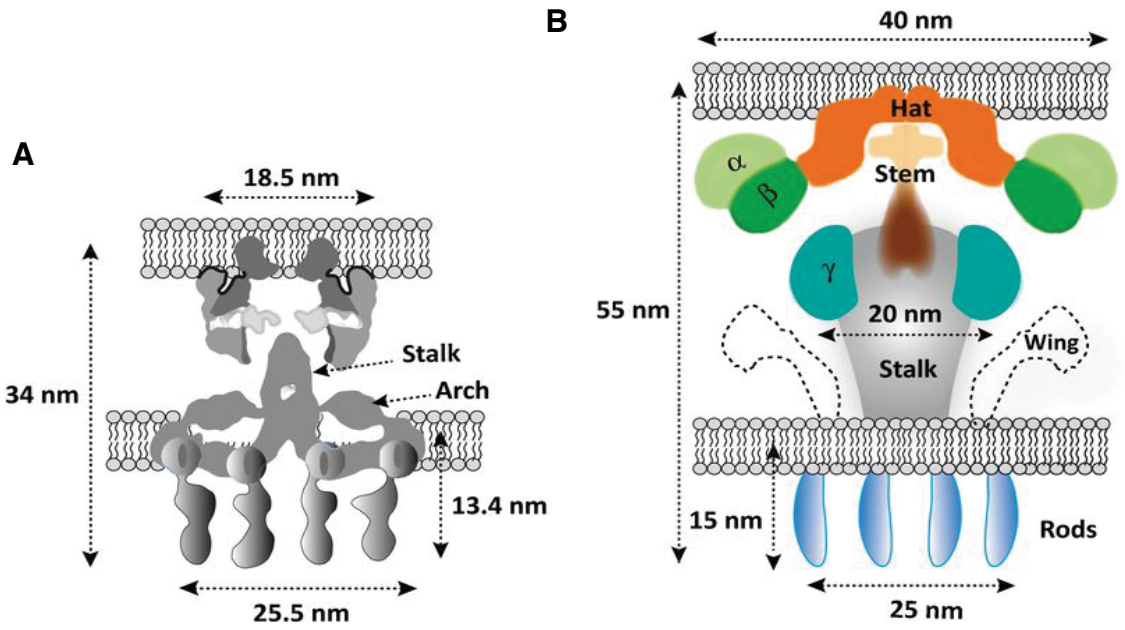


Figure 1.2 Structural comparison between T4ASS and T4BSS. (*A*) Negative-stain reconstruction of the *Escherichia coli* R388-encoded T4ASS. (*B*) Structure of the *Legionella pneumophila* T4BSS by electron cryotomography. Dimensions and structural features are identified. Adapted from (Ghosal et al., 2017).

1.4 Bacterial effectors

1.4.1 Effector identification. *Lpn* injects ~300 virulence factors, termed effectors, into the host cell. Effectors modulate host functions for the pathogen's benefit during infection. The specific function of most effectors remains to be solved and knocking out a single effector rarely causes a defect in intracellular growth. Oftentimes, the deletion of single substrates of the T4SS does not produce a defect (Luo and Isberg, 2004). This points to the functional redundancy of effectors and the possibility that certain effectors play a major role only in specific hosts.

Identification of effectors is the first step toward understanding the progression of *Legionella* infections. Various experimental and computational techniques have been used. The first effector, RalF, was identified by sequence homology to a eukaryotic Guanine Exchange Factor (GEF) domain. Since the human ADP ribosylation factor-1 (ARF1) was shown to localize with the LCV and the association of cytosolic ARF onto membranes requires guanosine triphosphate (GTP) activation, Nagai et al. searched the *Lpn* genome for proteins with homology to ARF-specific GEFs (Nagai et al., 2002). Since then, algorithms have been written that take into account features of validated effectors to help predict new effectors (Burstein et al., 2009). Seven main attributes were used in the classification (**Fig. 1.3A**).

One: genome arrangement. Mapping the genomic location of effectors has shown that they tend to cluster within four regions (**Fig. 1.3B**). This is likely due to horizontal gene transfer events between the eukaryotic host and the pathogen (de Felipe et al., 2005).

Two and three: G+C content and sequence similarity to host proteins. Many of these effectors have eukaryotic-like domains, such as WipA which has a coiled-coil domain, and AnkB, which has an ankyrin domain (Pinotsis and Waksman, 2017; Wong et al., 2017). Hence *Lpn* is well-equipped to interfere with a numerous of host processes. This raises the question of the origin of these effectors. Indeed, effector genes with eukaryotic-like domains have lower G+C contents (average 36.9%) compared with other *Lpn* genes (average 39.3%) (Burstein et al., 2009). This supports the horizontal gene transfer theory, in that most protozoan hosts have a low G+C content genome, such as 27% in *Tetrahymena thermophila* (Eisen et al., 2006).

Four, five and six: similarity to known effectors, secretory signals and abundance in metazoa/bacteria. Many effectors are functionally redundant, which translates to local sequence similarity between validated effectors and unidentified effectors. It has also been suggested that a

C-terminal secretory signal is present in certain effectors. However, the exact signal and whether this signal applies to all translocated effectors is unknown. For example, a hydrophobic residue or a proline residue at the -23 or -24 position from the C-terminus is important for translocation of RalF (Nagai et al., 2005). Additional analyses have narrowed down the enrichment of serine and threonine in -3 to -11 and hydrophobic amino acids in -1 to -3 (Burstein et al., 2009). Another feature of effectors is that they are unlikely to be housekeeping genes that would be present in other bacteria.

Seven: regulatory elements. Finally, the expression of certain effectors has been linked to regulatory two-component systems such as PmrA and PmrB. PmrB acts as a membrane-bound histidine kinase sensor that activates PmrA under specific conditions, such as low pH levels (Al-Khodor et al., 2009; Chen and Groisman, 2013). In turn, PmrA binds to specific DNA promoters present in at least 35 effectors, modulating their expression (Zusman et al., 2007).

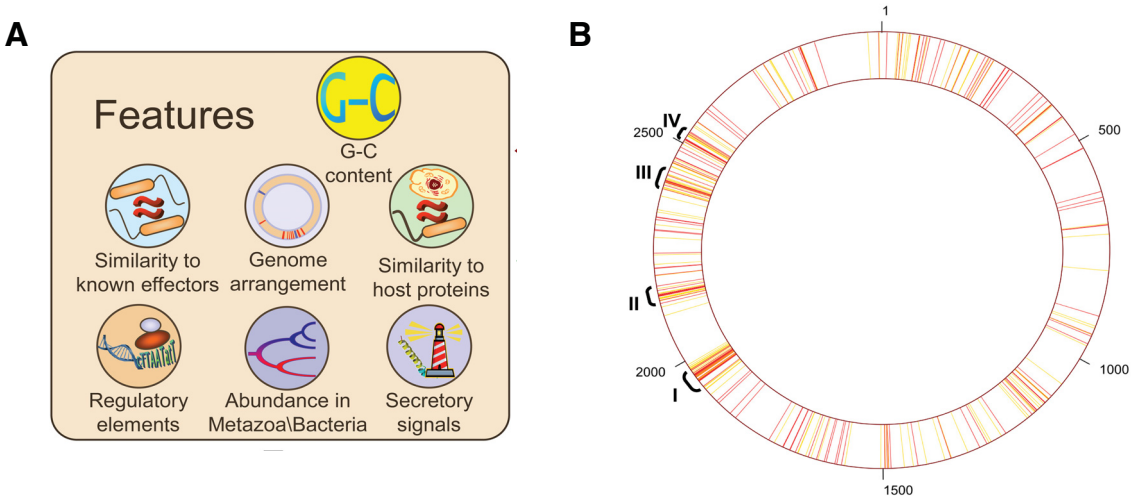


Figure 1.3 General features of *Legionella* effectors. (*A*) Seven attributes of previously validated effectors were used in a machine learning algorithm for prediction of new effectors. (*B*) Genomic distribution of validated (red) and putative (yellow) show clustering in four regions. Units are in ORF. Adapted from (Burstein et al., 2009).

1.4.2 Effector validation. The calmodulin-dependent adenylate cyclase (Cya) translocation assay is commonly used to validate newly identified effectors. In this system, candidate effectors are fused to the calmodulin-dependent adenylate cyclase domain. If the effector is translocated by the T4SS into a host cell, the adenylate cyclase is activated by calmodulin and converts adenosine triphosphate (ATP) into cyclic adenosine monophosphate (cAMP). *Lpn* lacks calmodulin. Uninfected host cells and infections using WT-Cya are used as negative controls. Therefore, an increase in cAMP concentration will be due to effector secretion into a host cell (Chakravarthy et al., 2017).

1.5 Canonical ubiquitination

1.5.1 General mechanism of protein ubiquitination. In eukaryotes, a major mechanism for regulation of a wide variety of cellular processes, is ubiquitination. It is a highly regulated process that covalently attaches ubiquitin (Ub) onto target proteins for different fates. Modification often leads to degradation by the proteasome. Proteins with Lys48-linked poly Ub chains are sent to the 26S proteasome (Thrower et al., 2000). Other consequences include alteration of function, DNA repair and trafficking, depending on the type of Ub linkage (Hubber et al., 2013).

This post-translational modification of proteins does not exist in *Lpn*, and *Lpn* has cleverly developed effectors to target this function. For example, AnkB functionally mimics a eukaryotic E3 ubiquitin ligase, SidE modifies Ub in a novel way, and ubiquitination of SidH results in its own degradation using the host proteasome (Kubori et al., 2010; Price et al., 2009; Qiu et al., 2016). It has also been shown that decoration of the LCV with Ub is required for protection of the pathogen containing vacuole from fusion with a lysosome (Price et al., 2009). Hence, manipulation of this

host pathway allows *Lpn* to temporally regulate host proteins and bacterial proteins for optimal intracellular proliferation.

Canonical protein ubiquitination involves a three-enzyme cascade consisting of E1, E2 and E3 proteins (**Fig. 1.4**). The organization of the ubiquitination cascade is pyramidal, with increasing number of enzymes and increasing specificity from E1 to E3. In the presence of ATP and Mg^{2+}, Ub is first activated by the E1 Ub-activating enzyme that covalently attaches itself via a thioester bond between a cysteine residue to the C-terminal Gly76 residue of Ub (Hershko et al., 1983). The activated Ub is transferred onto a cysteine residue of an E2 Ub-conjugating enzyme. The E2 then binds to an E3 Ub-ligase which confers substrate specificity through its protein-protein interacting domain (Hershko et al., 1986). The nitrogen atom on substrate lysines act as a nucleophile attacking the electron-deficient carbonyl carbon of the thioester linkage between Ub and the E2. Ultimately, Ub is attached via an isopeptide bond to a lysine of its target (Scheffner et al., 1995).

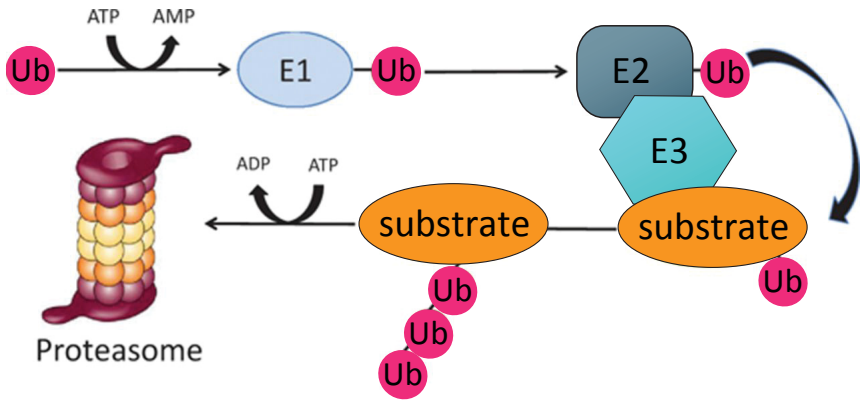


Figure 1.4 The ubiquitin proteasome system. Three enzymes are involved in tagging proteins with ubiquitin for proteasomal degradation. The E1 Ub-activating enzyme attaches itself to Ub and transfers it to an E2 Ub-conjugating enzyme. The E2 enzyme then transfers Ub onto a substrate with the help of an E3 Ub-ligase. Shown here is the RING-type E3 which acts as a scaffold. Adapted from (Pagan et al., 2013)

1.5.2 Types of E3 ubiquitin ligases. The E3 Ub-ligases is the largest group of enzymes, in the order of hundreds, involved in ubiquitination (Morreale and Walden, 2016). This results in thousands of potential different combinations between E2 and E3 enzymes. The E2 passes Ub to different E3 ligases in different ways. It may use the E3 as a scaffold, as in the Really Interesting New Gene (RING)-type E3 ligases that represent the largest family of E3s (**Fig. 1.5A**) (Pickart, 2001). Ub does not come into contact with the E3. This is exemplified by the Cullin protein of the SCF (Skp1-Cullin-F-box) E3 complex. At one end, Cullin binds a RING protein for interaction with a specific E2. The other end binds an adaptor Skp1 protein for interaction with a substrate receptor F-box protein (Sheikh et al., 2015). The main role of SCF E3 ligases is in the regulation of the cell cycle by the periodic ubiquitination of proteins such as Cyclin E and p27 (Nakayama and Nakayama, 2005).

An E2 enzyme may pass Ub onto a cysteine residue of a Homologous to the E6AP Carboxyl Terminus (HECT)-type E3 ligase, before Ub is transferred onto a target (**Fig. 1.5B**). In this case, transfer of ubiquitin to a substrate is a two-step reaction. The C-terminal HECT domain can be divided into an N-terminal lobe and a C-terminal lobe. The N-lobe attaches to the incoming E2 enzyme, while the C-lobe accepts the activated Ub through its catalytic cysteine (Huibregtse et al., 1995). Substrate specificity is achieved by a diverse range of N-terminal domains. NEDD4 is a highly-studied HECT-type E3. In addition to the C-terminal HECT domain, it contains a calcium/lipid-binding domain and multiple copies of WW domains for protein-protein interactions. NEDD4 is known to regulate intracellular Na^+ concentrations (Harvey and Kumar, 1999).

A hybrid of the two main E3 Ub ligases exists known as the Ring-in-Between-Ring (RBR) enzymes due to the presence of two RING domains (RING1 and RING2) separated by an In-

Between-RING domain (IBR) (**Fig. 1.5C**). Although RBR ligases have RING domains, the catalytic cysteine found in HECT E3 ligases is also present (Wenzel et al., 2011). Parkin is a notable RBR enzyme. Mutations in this gene is the leading cause of autosomal recessive Parkinson's disease, likely due to its role in mitophagy (Trempe et al., 2013). Recent studies have also identified Parkin as a tumor suppressor gene (Wahabi et al., 2018).

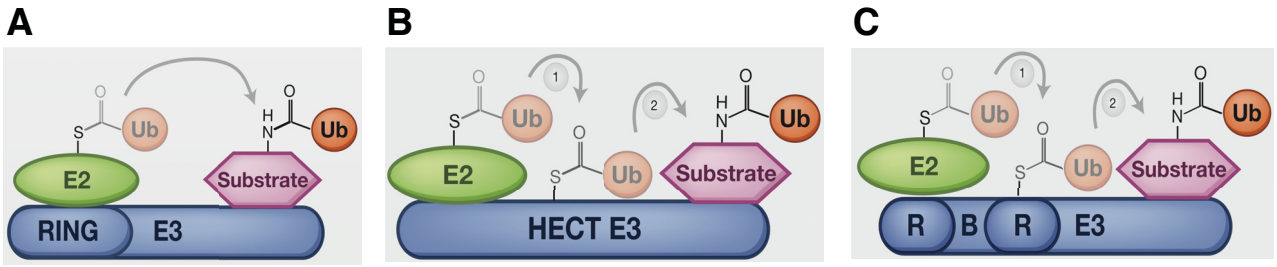


Figure 1.5 Types of E3 ubiquitin ligases. (*A*) The RING-type E3 acts as a scaffold between an E2 enzyme and the target protein. (*B*) The HECT-type E3 contains a catalytic cysteine to receive Ub directly from an E2 enzyme, before transferring it to a target. (*C*) The RBR-type E3s are a hybrid of the RING and HECT classes. They contain both RING domains and an active cysteine. Adapted from (Morreale and Walden, 2016).

1.5.3 Deubiquitinating enzymes. Ubiquitination is a reversible process. Deubiquitinating (DUB) enzymes specifically recognize different types of Ub linkages to cleave Ub from substrates and from poly-ubiquitin chains. Hence, like Ub, DUBs play a role in a wide array of processes. With regards to the degradation pathway, DUBs such as USP14, cleave Ub units off substrates upon entry into the proteasome, allowing the targeted protein to feed into the proteasome's catalytic core (Harrigan et al., 2018). DUBs also act as a fall back mechanism, by which labeled proteins may escape degradation if the removal of Ub is slower than initiation of degradation (Finley, 2009). To add another level of complexity, DUBs are themselves regulated. Phosphorylation of nuclear USP4 relocalizes it to the cytoplasm and membrane, where it induces pro-tumorigenic function of Transforming Growth Factor-β (TGF-β) signaling in breast cancer

cells (Zhang et al., 2012). Ubiquitination also mediates DUB activity. Ubiquitination of Ataxin-3, a DUB associated with Machado-Joseph disease, causes a conformational change to activate the enzyme (Faggiano et al., 2015).

1.6 Manipulation of host ubiquitination

1.6.1 Decoration of the *Legionella*-containing vacuole with ubiquitin. Following infection of host cells, the LCV becomes decorated with Ub in wild-type *Lpn* infections, but not in T4SS defective mutant infections (**Fig. 1.6**) (Dorer et al., 2006). This indicates recruitment of polyubiquitinated species to the LCV is mediated by effectors. By one-hour post infection of human bone marrow derived macrophages (BMDM), ~60% of LCVs were Ub-positive. Nearly all (>90%) stained positive by 7 hours. At 10 hours, staining decreased in large LCVs and LCVs containing single bacterium appeared (Ivanov and Roy, 2009). This is in agreement with the ~10-hour replication cycle *Lpn* has inside a host cell before lysis and reinfection (Xu and Luo, 2013).

To date, Lys11-linked, Lys48-linked, and Lys63-linked Ub chains have been traced to the surface of LCVs (Bruckert and Abu Kwaik, 2015b; Ivanov and Roy, 2009). Lys11 Ub linkages are not well understood, but have been shown to be involved in regulating cell division (Wickliffe et al., 2011). Lys48-linkage is commonly associated with proteasomal degradation, while Lys63-linkage has been linked to protein degradation, DNA repair and autophagy, among other pathways (Li and Ye, 2008). To that effect, proteasomal inhibition by MG132 was found to arrest intracellular replication (Dorer et al., 2006).

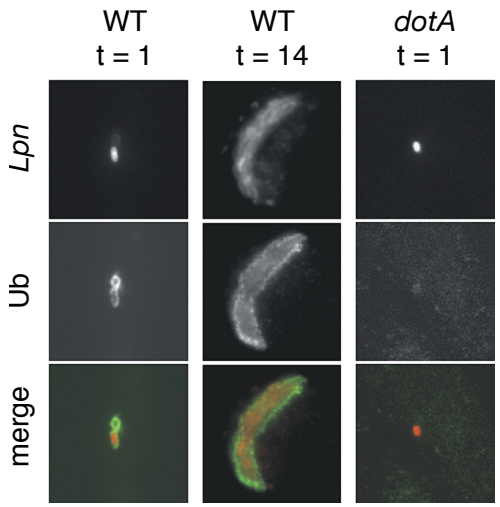


Figure 1.6 Accumulation of ubiquitin on the LCV during infection. Mouse BMDMs are challenged by WT and T4SS defective (*dotA*) *Lpn* as labeled. Cells were fixed at the indicated times and immunostained with anti-*Lpn* (red) and anti-Ub (green). Adapted from (Dorer et al., 2006).

1.6.2 AnkB. The *Lpn* Philadelphia-1 genome encodes for five effectors with an F-box domain: Ankyrin B (AnkB), LegU1, LicA, Lpg1975 and Lpg2525 (Ensminger and Isberg, 2010). Through interaction with Skp1, F-box proteins form part of the SCF E3 Ub ligase complex that tags a range of proteins for degradation (Skowyra et al., 1997). Substrate specificity is also determined by the F-box protein through differing N-terminal protein-protein interaction domains. Of these five F-box effectors, only AnkB and LicA are conserved across all sequenced *Legionella* genomes (Hubber et al., 2013). AnkB deletion mutants have been shown to cause severe defects in the *Lpn* life cycle in both *Acanthamoeba polyphaga*, human monocytes-derived macrophages (hMDMs) and U937 cells (Al-Khodor et al., 2008).

AnkB has three main features. First, it uses an N-terminal F-box to recruit the SCF Ub ligase machinery (Price et al., 2009). Second, it confers substrate specificity via three ankyrin repeats (Price et al., 2010a). This protein-protein interaction domain has never been reported in

eukaryotic F-box proteins. Thus, the acquisition of these repeats by AnkB is likely important for selective substrate recruitment. Finally, a C-terminal farnesylation motif anchors this effector to the membrane of a growing LCV (Price et al., 2010b).

These features allow AnkB to take part in two roles during *Lpn* growth. As it recruits substrates for ubiquitination, and can itself be ubiquitinated, association of AnkB with the LCV contributes to the LCV ubiquitination phenomenon (Bruckert and Abu Kwaik, 2015b). In addition, AnkB is involved in "nutritional virulence" (Abu Kwaik and Bumann, 2013). Part of the host defense mechanism against pathogen is the restriction of access to host metabolites (Eisenreich et al., 2013). As such, the levels of amino acids in the host cytosol are an insufficient source of carbon, nitrogen and energy for proliferation of *Lpn* (Abu Kwaik and Bumann, 2013). *Lpn* is auxotrophic for seven amino acids (Arg, Cys, Ile, Leu, Met, Thr, and Val), but utilizes amino acids as its main source of energy through the tricarboxylic acid (TCA) cycle (Pine et al., 1979). Dependent on host amino acids, *Lpn* uses AnkB to direct proteins for proteasomal degradation to free amino acids. This provides an increased pool of nutrients for continued proliferation (**Fig. 1.7**) (Price et al., 2011).

AnkB has been reported to modulate ubiquitination of host Parvin B (ParvB), a protein regulating focal adhesion, cell spreading and motility (Legate et al., 2006). Interestingly, the amount of endogenously ubiquitinated ParvB decreased in cells overexpressing AnkB (Lomma et al., 2010). However, the biological significance of decreased ParvB ubiquitination is unclear.

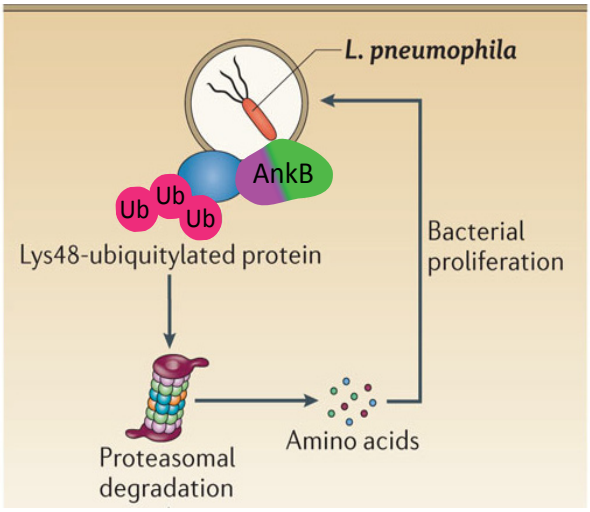


Figure 1.7 AnkB is involved in nutritional virulence. AnkB is anchored onto the LCV via its farnesylation motif. It recruits proteins to be Lys48-linked with Ub for proteasomal degradation. This generates a pool of free amino acids that feeds back into *Lpn* proliferation. Adapted from (Ashida et al., 2014)

1.6.3 LubX. *Legionella* U-box protein (LubX) is another effector that redirects the host ubiquitination machinery. U-box domains are RING-like domains that interact with E2 enzymes (Aravind and Koonin, 2000). LubX carries two U-box domains, one interacting with an E2, and the other functions as a substrate-binding site (Quaile et al., 2015). This effector was first reported to target host cell cycle CDC2-like kinase 1 (Clk1) for degradation, but the significance of this is still unknown (Kubori et al., 2008). More recently, LubX was proposed to be a metaeffector, an effector that regulates the function of another effector, in this case SidH, inside the host cell (Kubori et al., 2010). The intracellular level of LubX increases over time, resulting in a decrease of SidH protein as *Lpn* proliferates (**Fig. 1.8**). SidH could be detected by 15 minutes post infection and steadily declined to undetectable levels by 8 hours post infection (Kubori et al., 2010). This tight regulation of SidH is essential, as unregulated expression of SidH led to a growth arrest in yeast cells that could be alleviated by the re-introduction of LubX (Quaile et al., 2015). In line

with this result, infections of *Drosophila melanogaster* using a Δ*lubX* Δ*sidH* double mutant resulted in a lower lethality rate for the flies, and more viable *Legionella* per fly than infections with a Δ*lubX* mutant (Kubori et al., 2010).

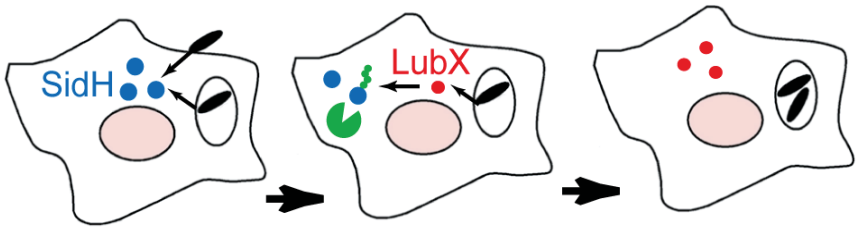


Figure 1.8 LubX functions as a metaeffector. LubX is a U-box containing protein capable of hijacking the host ubiquitin proteasome system for the degradation of SidH. Adapted from (Kubori et al., 2010)

1.7 Non-canonical ubiquitination

1.7.1. SidE-mediated ubiquitination. Non-canonical ubiquitination refers to ubiquitin linkages not formed between a substrate lysine and the ubiquitin molecule (**Fig. 1.9A**). Other nucleophilic groups in amino acids are capable of attacking the carbonyl carbons. Examples include serine, threonine and tyrosine that have electron-rich oxygen atoms which can form hydroxyester bonds with Ub (**Fig. 1.9B**) (Wang et al., 2007b). Cysteine residues can also form thioester bonds with ubiquitin, although the thioester linkage is much weaker than the amide bond (**Fig. 1.9C**) (Cadwell and Coscoy, 2005).

The SidE (substrate of Icm/Dot transporter) family of effectors was first identified in 2004 (Luo and Isberg, 2004). It includes four large proteins, SdeA, SdeB, SdeC and SidE, approximately 170 kDa, with the exception of an approximately 200 kDa SdeB which contains a repeated C-terminal domain (Bardill et al., 2005). *sdeA, sdeB,* and *sdeC* are located in proximity. *sidE* is located at a separate locus (**Fig. 1.10A**). A *Lpn* strain lacking all four members exhibited attenuated

growth in the environmental hosts, *A. castellanii* and *D. discoideum*, but not in mouse BMDMs. This defect could be rescued by expressing *sdeA* on a plasmid, pointing again to the functional redundancy of many bacterial effectors (Luo and Isberg, 2004; Sheedlo et al., 2015).

SidE proteins are capable of non-canonical ubiquitination of several ER-associated proteins on a cysteine residue (Kotewicz et al., 2017; Qiu et al., 2016). The mechanism of this reaction revealed many new cellular processes. Surprisingly, this ubiquitination occurs independently of the host ubiquitination machinery, ATP, and Mg^{2+} (Qiu et al., 2016). This novel mechanism requires only nicotinamide adenine dinucleotide (NAD) and Ub. Not only does the SidE family function as an E1, E2–independent enzyme, it generates a new type of Ub attachment. A substrate cysteine is attached by a phosphoribosyl moiety connected to an arginine residue of Ub, or a phosphoribosylated Ub (P-Rib-Ub) (Puvar et al., 2017; Qiu et al., 2016).

SidE family members have two domains involved in generating this P-Rib-Ub. First, the mono ADP-ribosyl transferase domain (mART) transfers an ADP-ribosyl group from NAD onto Arg42 of Ub, generating ADP-Rib Ub. Then, its phosphodiesterase domain (PDE) cleaves the pyrophosphate bond, releasing an AMP molecule. In this case, if the nucleophile is a serine residue, a ubiquitinated substrate is generated. However, if water is the nucleophile, hydrolysis generates P-Rib-Ub (**Fig. 1.10B**) (Bhogaraju et al., 2016; Qiu et al., 2016).

Ubiquitin chains of different linkage types (Lys11, Lys48, Lys63 and Met1) modified by the SidE family are protected from a variety of DUBs. From a panel of nine DUBs specific for Lys63-, Lys48- and Met1-linked ubiquitin chains, the attachment of ADP-Rib or P-Rib protected di-Ub from DUB hydrolysis except for two tested DUBs (Puvar et al., 2017). This is likely due to steric clashes of the additional groups on Arg42 of Ub. For example, from the crystal structure OTUlin in complex with Ub (PDB code 3ZNZ), a human DUB specific for Met1-linked ubiquitin

chains, Arg42 is located at the interface between Ub and the enzyme. In fact, a hydrogen bond is formed between Ub-Arg42 and Glu287 of OTUlin (**Fig. 1.11**) (Keusekotten et al., 2013).

Only a few substrates of SidE modification have been identified. Reticulon 4 (Rtn4) is an ER-associated protein known to control the shape of homeostasis of the ER (Zurek et al., 2011). Although LCVs avoid lysosomal fusion, they communicate extensively with the ER to create a mature LCV (Horwitz, 1983). SidE-mediated ubiquitination of Rtn4 causes considerable ER rearrangement. When challenged with WT *Lpn*, GFP-Rtn4 concentrates around the LCV and then nucleates outwards. This phenomenon is abolished when infecting with a SidE family deletion strain. In addition, rough ER sheets prematurely associated with the LCV (Kotewicz et al., 2017).

ER-associated Rab proteins, a family of small GTPases, are also targets of SidE ubiquitination. Rab1, Rab6A, Rab30 and Rab33b all showed a clear increase in molecular mass after treatment with SdeA (Qiu et al., 2016). Modification of Rab33b was the most intensive and has since become the model substrate for SidE-mediated ubiquitination. In particular Rab33b is involved in trafficking between the ER and the Golgi apparatus (Valsdottir et al., 2001). Its overexpression has been shown to decrease the number of LCVs containing more than 10 bacteria, although the reason is unclear (Qiu et al., 2016).

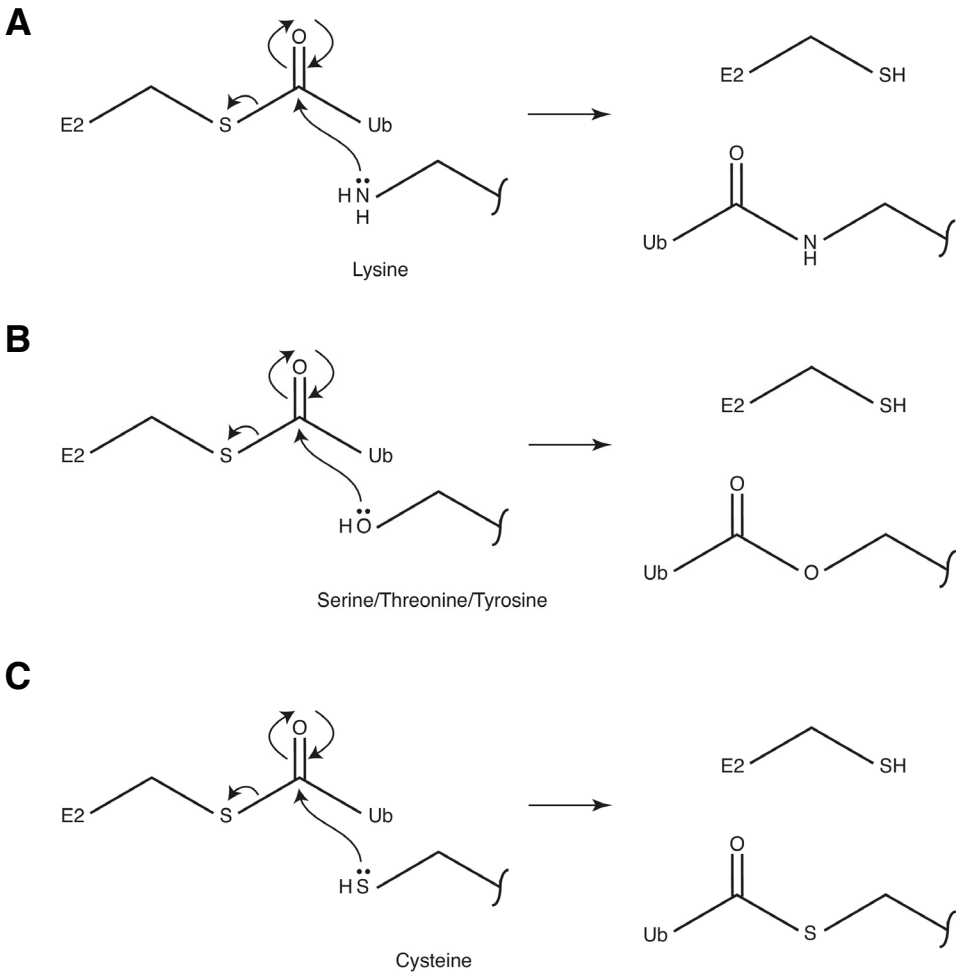


Figure 1.9 Different nucleophilic groups in ubiquitination. (*A*) In canonical ubiquitination, the amine group of lysine acts as the nucleophile, forming amide bonds. (*B*) The hydroxyl group of serine, threonine and tyrosine can form hydroxyester bonds. (*C*) The thiol group of cysteine can form a thioester linkage. Adapted from (McDowell and Philpott, 2013).

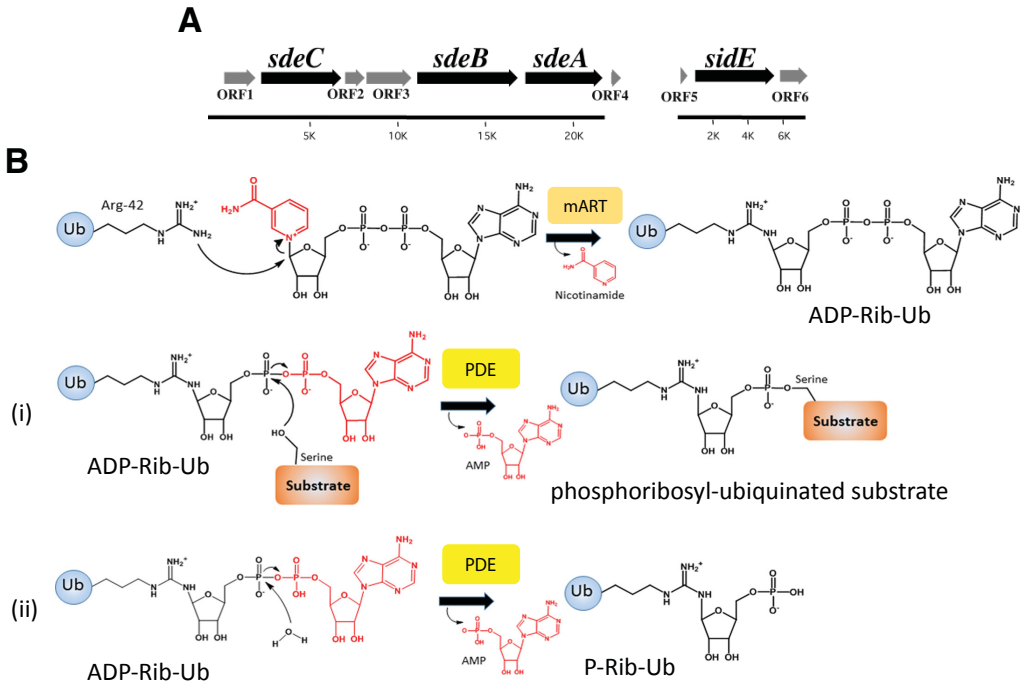


Figure 1.10 SidE-mediated ubiquitination. (*A*) *sdeA, sdeB, and sdeC* are located in proximity. *sidE* exists at a separate locus. Adapted from (Bardill et al., 2005). (*B*) SidE-mediated ubiquitination occurs in two steps. First, the mART domain cleaves off nicotinamide from NAD to generate ADP-Rib-Ub. Second, the PDE domain cleaves off AMP, to attach P-Rib onto either (i) a substrate serine or (ii) water. Adapted from (Puvar et al., 2017).

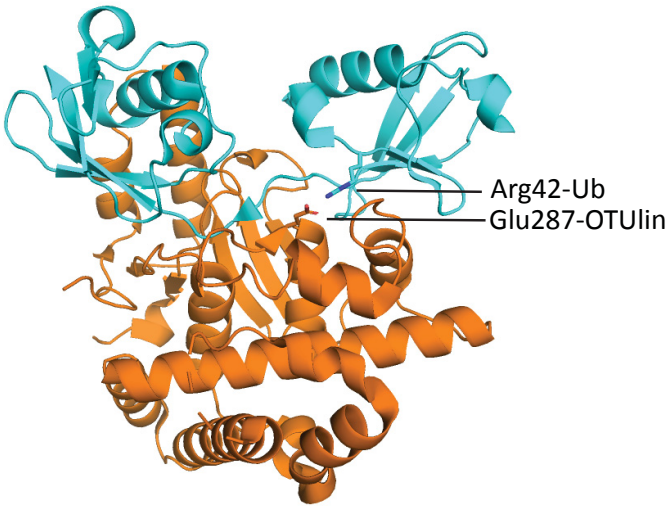


Figure 1.11 Protection of SidE family modified ubiquitin chains from host DUBs. Crystal structure of human OTUlin bound to Met1-linked di-Ub (PDB code 3ZNZ). Arg42 of Ub interacts with Glu287 of OTUlin. Addition of ADP-Rib or P-Rib onto Arg42-Ub would sterically hinder its interaction with the DUB. Orange = OTUlin. Cyan = Ub. Adapted from (Puvar et al., 2017).

1.7.2 SidE family impairment of the canonical ubiquitination cascade. If water acts as the acceptor in the final step of phosphoribosyl-ubiquitination mediated by the PDE domain, P-Rib-Ub is generated. Generation of P-Rib-Ub is one of the main functions of SdeA, as free Ub is modified regardless of the presence of a known substrate, Rab33b. This novel Ub molecule has implications in conventional E1/E2 mediated ubiquitination, impairing many downstream pathways including autophagy and proteasomal degradation (Bhogaraju et al., 2016; McDowell and Philpott, 2013).

Analysis of the P-Rib-Ub crystal structure confirmed that the ribosyl moiety indeed attaches to Arg42-Ub (PDB code 5M93) (Bhogaraju et al., 2016). However, electron density for the phosphate group is missing. The major difference between WT Ub and P-Rib-Ub is in the position of the side chains of Arg42 and Arg72 (**Fig. 1.12A**). As mentioned, Arg42 is the attachment site for both ADP-Rib and P-Rib. Arg72 is part of the C-terminal of Ub important for its recognition (Winget and Mayor, 2010). The ribose group forms hydrogen bonds with Arg72, shifting it from its WT position. Surprisingly, mutating Arg72 to alanine was able to completely abolish Ub modification by SdeA, pointing to its role not only in stabilization of the phosphoribosyl moiety, but in catalysis of the reaction (Bhogaraju et al., 2016).

Comparison of the P-Rib-Ub structure to that of Ub bound to a Uba1, a yeast E1 enzyme, reveals major steric clashes that would explain the incompatibility of P-Rib-Ub to be activated by E1 (**Fig. 1.12B**). The extra phosphoribosyl moiety attached to Arg42 and the Arg72 in its new position, clashes with the E1. Indeed, P-Rib-Ub was shown to inhibit ubiquitination by all three types of E3 Ub ligases, RING, HECT and RBR, by interfering with both the E1 and E2 steps of the reaction cascade (Bhogaraju et al., 2016). First, E1 cannot efficiently activate P-Rib-Ub using ATP. Second, transfer of P-Rib-Ub from E1 onto E2 is significantly diminished. This is depicted

by a fainter band of Ub-loaded E2 in a reaction mixture containing Ub, ATP, E1, E2, NAD and SdeA as opposed to in a reaction without SdeA. Finally, discharge of a P-Rib-Ub loaded E2 was also impaired (Bhogaraju et al., 2016).

SdeA directly affects multiple Ub-dependent pathways. For example, Parkin is an E3 Ub ligase that causes ubiquitination of damaged mitochondria, leading to mitophagy (Seirafi et al., 2015). Chemically induced depolarization by carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP) treatment effectively leads to ubiquitination of the mitochondria. However, in cells expressing SdeA, this phenomenon is significantly impaired (Bhogaraju et al., 2016).

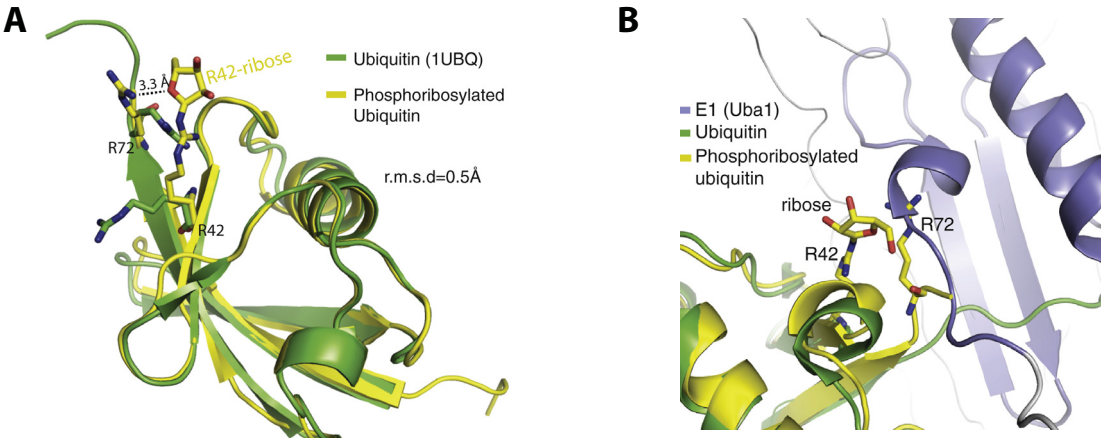


Figure 1.12 Structural analysis of phosphoribosylated ubiquitin. (*A*) Crystal structure of P-Rib-Ub in yellow (PDB code 5M93) overlaid with free Ub in green (PDB code 1UBQ). Arg42 and Arg72 interact with the ribosyl moiety. (*B*) Superimposition of P-Rib-Ub (PDB code 5M93) and Ub-bound E1 (Uba1) (PDB code 3CMM), highlighting steric clashes between modified Ub and E1. Adapted from (Bhogaraju et al., 2016).

1.7.3. SidE-mediated deubiquitination. Not only does the SidE family catalyze a unique form of ubiquitination, they also contain an N-terminal DUB domain which catalyzes the removal of Ub units. Although the DUB activity is dispensable for SidE family function during infection, it regulates the dynamics of LCV-associated ubiquitinated species (Sheedlo et al., 2015). As

mentioned, an infection of *Dictyostelium discoideum* with the SidE family deletion strain resulted in a ~100-fold defect in intracellular growth rate, which can be rescued by complementation with SdeA (Bardill et al., 2005). Interestingly, the catalytically DUB inactive Cys118-Ala mutant was still able to counteract the defect (Sheedlo et al., 2015). However, given the enrichment of the LCV membrane with ubiquitin, it is not surprising that SdeA-DUB plays a role in its regulation. Indeed, the Cys118-Ala mutant was unable to reduce LCV ubiquitination levels back to WT levels (Sheedlo et al., 2015).

The SdeA-DUB domain was crystallized in 2015 and is the first structural characterization of a prokaryotic DUB. Although it contains a classical Ub-like protease (Ulp) fold, its Ub interacting interface is quite different. It uses a Cys-His-Asp- (CHD) catalytic triad capable of recognizing the most common Ub linkages (Lys11, Lys48 and Lys63) with a clear preference for Lys63-linked chains (Sheedlo et al., 2015).

A few Ub-like proteins exist and can interact with DUBs. These include SUMO (small Ub-related modifier), NEDD8 (neural precursor cell expressed, developmentally downregulated 8), ISG15 (interferon-stimulated gene 15) and Atg8 proteins (Amerik and Hochstrasser, 2004). SdeA forms a complex with NEDD8 and ISG15, but not SUMO (Sheedlo et al., 2015). Again, this complex formation required the active Cys118. NEDD8 has a ~58% sequence identity with Ub is known to target cullin proteins (of the SCF complex) and the p53 tumor suppressor. Neddylation of cullin stimulates the Ub ligase activity of the SCF complex by accelerating E2-E3 binding (Kawakami et al., 2001). While neddylation of p53 inhibits its transcription activity (Xirodimas et al., 2004). ISG15 is the first identified Ub-like protein. Accumulation of ISG15 conjugates in the brain leads to reduced life expectancy in mice with necrosis of brain cells (Ritchie et al., 2002). The significance of broad specificity of SdeA-DUB remains to be determined.

Although a monomer in solution, two molecules in the asymmetric unit were crystallized (PDB code 5CRA). Two lobes (α and β) pack together connected by two loops (L2 and L10) (**Fig. 1.13**). The CHD catalytic triad exists at the interface between the two lobes and were caught in different orientations in three separate crystals pointing to plasticity capable of accommodating both Ub and Ub-like proteins. In the co-crystal structure, less interactions are observed between SdeA-DUB and Ub than in other DUB-Ub complex crystals possibly explaining the reactivity of SdeA-DUB with Ub, NEDD8 and ISG15. For example, the Ub-Ile44 hydrophobic patch does not participate in binding. This patch is essential for affinity in eukaryotic DUBs (Dikic et al., 2009). In addition, the β1–β2 loop of ubiquitin usually forms multiple bonds with a DUB (Phillips et al., 2013). This interaction is primarily replaced with weaker van der Waals contacts and only one hydrogen bond remains in SdeA (Sheedlo et al., 2015). Most of these interactions can still occur if Ub was replaced with ISG15 or NEDD8. However, SUMO is unable to bind due to incompatibility between its C-terminal tail and the DUB domain.

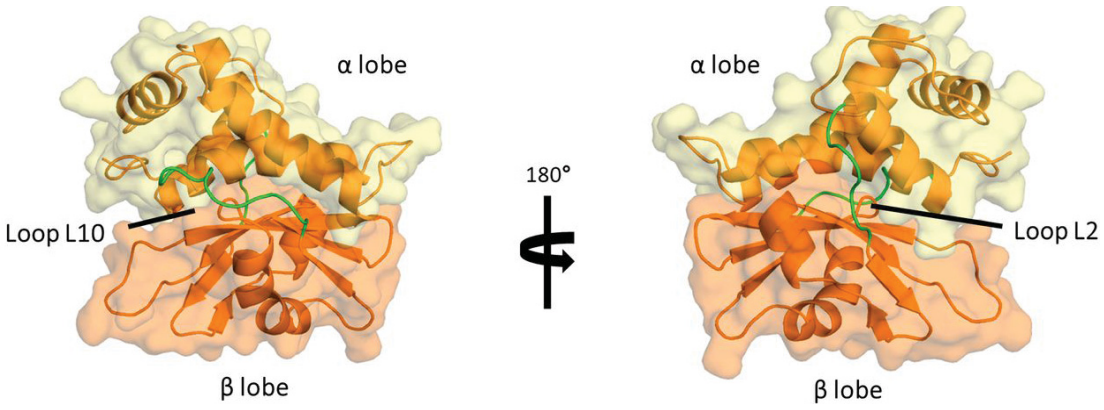


Figure 1.13 Structure of the SdeA deubiquitinase domain. The SdeA-DUB domain is formed by 2 lobes (α in yellow, and β in orange) connected by two loops (L2 and L10). Adapted from (Sheedlo et al., 2015).

1.7.4 SidJ regulation of the SidE family. SidJ is another example of a metaeffector by antagonizing the effect of SidE proteins. This effector is expressed from the same locus as *sdeA*, *sdeB*, and *sdeC* (ORF3 in **Fig. 1.11A**) and encodes a DUB capable of cleaving the phosphodiester bond between P-Rib-Ub and its substrates. Its activity is essential for efficiently reducing ubiquitinated Rab33b levels in infected cells in a timely manner (Qiu et al., 2017). Hence, SidJ counteracts the activity of the SidE family in a temporal manner. Secretion of SdeA into the host cell becomes constant two hours after infection. In contrast, the amount of SidJ continues to increase throughout infection, leading to an increasing ratio of SidJ to SdeA and the reduction of phosphoribosyl-ubiquitinated substrates at the later stages of infection (Qiu et al., 2017). It suppresses the yeast toxicity conferred by SidE proteins, as the overexpression of SdeA is only detrimental in the absence of SidJ (Havey and Roy, 2015; Jeong et al., 2015).

SidJ also acts as a canonical DUB. It is able to completely hydrolyze Lys11-, Lys48- and Lys63-linked di-Ub within two hours, and partially cleave Lys33-linked di-Ub. Of note, its efficiency was much lower than that of SdeA-DUB although it prefers Lys63-linked di-Ub like SdeA-DUB (Qiu et al., 2017). However, SidJ activity does not require catalytic cysteine residues as in other Cys-His-Asp (CHD) catalytic triad-containing DUBs. SidJ mutants lacking one of three cysteine residues still effectively counteracted SdeA mediated ubiquitination of Rab33b and hydrolyzed Lys63-linked di-Ub, albeit with lower efficiency (Qiu et al., 2017). In addition, SidJ does not form an adduct with HA-Ub-VME (HA-tagged Ub vinyl methyl ester), a suicide inhibitor of canonical DUB domains. This inhibitor has a thiol-reactive species at the C-terminus of Ub that irreversibly binds DUBs at their active site (de Jong et al., 2012). As a control, SdeA-DUB formed a Ub adduct causing a molecular weight shift (Qiu et al., 2017). Thus, SidJ uses a different mechanism than classical DUBs to antagonize the canonical ubiquitination system.

1.8 Rationale and objective of research

We are heading into a post-antibiotic era as reported by the World Health Organization in its first global report on antimicrobial resistance in 2014 (World Health Organization, 2014). Although it is a natural evolutionary step for pathogens to develop methods for protection against bactericidal agents, widespread misuse in humans and animals has favored the development of resistant bacteria (Zaman et al., 2017). If the current trend continues as is, antimicrobial resistance is predicted to cause 10 million deaths a year, making it more lethal than cancer by 2050 (Totsika, 2017). Antibiotic resistance has become a public health concern with each discovery of a new antibiotic followed by development of microbial resistance to it (**Fig. 1.14**) (Clatworthy et al., 2007). Clearly, a new strategy to combat bacterial infections is needed.

Traditional antibiotics aim to effectively kill or inhibit growth by acting on general bacterial functions such as DNA or protein synthesis (Kohanski et al., 2010). As such, antibiotics are broad spectrum, targeting not only the pathogen, but the host microbiome. An emerging approach to treatment is antivirulence therapy. By inhibiting bacterial virulence and disarming the pathogen, this form of treatment offers a more tailored and specific approach. A potential advantage, although still unproven, is that this form of treatment would impose weaker selective pressure for the development of resistance (Clatworthy et al., 2007).

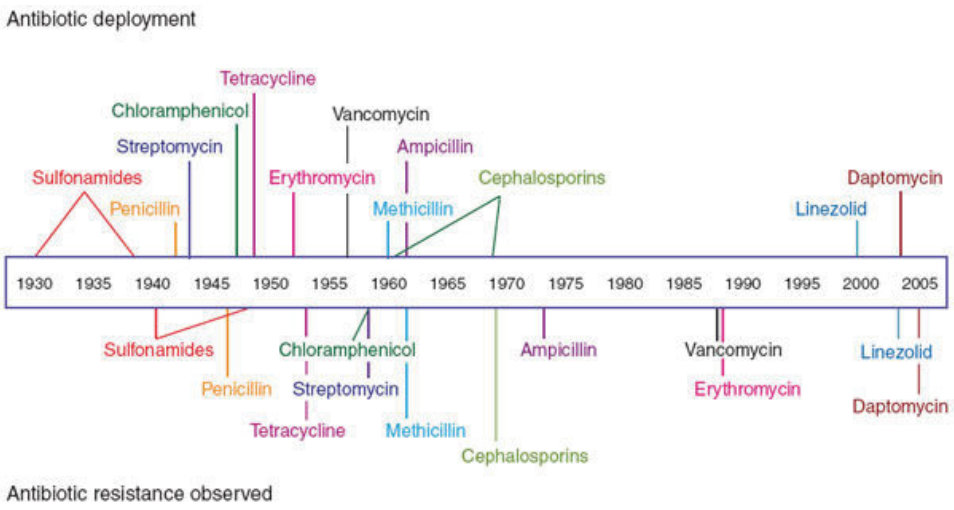


Figure 1.14 Rapid development of antibiotic resistance. Timeline of the discovery of new antibiotics versus emergence of antibiotic resistance heralding a post-antibiotic era. Adapted from (Clatworthy et al., 2007).

1.9 Organization of this thesis

This thesis focuses on the structural studies of *Lpn* virulence factors or effectors that modulate host ubiquitination. Chapter two centers on AnkB which interacts with the host SCF E3 Ub ligase to redirect substrates selected for ubiquitination (Wong et al., 2017). It takes part in nutritional virulence, generating a pool of free amino acids for *Lpn* replication within the LCV (Abu Kwaik and Bumann, 2013). Chapter three shifts the focus onto AnkC, an effector with the same substrate interacting domain as AnkB. Chapters four and five move into a novel mode of ubiquitination, phosphoribosyl-ubiquitination, achieved by the SidE family that shares a conserved domain with lpg1496. All these effectors are potential targets for antivirulence therapy to disarm *Lpn* as a cure for LD.

Chapter 2: Hijacking by an F-box effector

2.0 Connecting text

Host protein mimicry is a prevalent method used by pathogens to subvert their hosts. AnkB contains two eukaryotic-like domains. Its F-box domain interacts with a host E3 ubiquitin ligase complex to redirect ubiquitination towards substrates selected by its ankyrin repeats domain (Wong et al., 2017).

2.1 Summary

Ankyrin B (AnkB/LegAU13) is a translocated F-box effector essential for the intracellular replication of the pathogen, *Legionella pneumophila*. AnkB co-opts a host ubiquitin ligase to decorate the pathogen-containing vacuole with Lys48-linked polyubiquitinated proteins and degrade host proteins as a source of energy. Here, we report that AnkB commandeers the host ubiquitin proteasome system through mimicry of two eukaryotic protein domains. Using X-ray crystallography, we determined the 3D structure of AnkB in complex with Skp1, a component of the human SCF ubiquitination ligase. The structure confirms that AnkB contains an N-terminal F-box similar to Skp2 and a C-terminal substrate-binding domain similar to eukaryotic ankyrin repeats. We identified crucial amino acids in the substrate binding domain of AnkB and showed them to be essential for the function of AnkB in *L. pneumophila* intracellular proliferation. The study reveals how *Legionella* uses molecular mimicry to manipulate the host ubiquitination pathway and proliferate intracellularly.

2.2 Introduction

Legionnaire's disease is an atypical form of pneumonia with a fatality rate of up to 34% (Phin et al., 2014). The causative agent, *Legionella pneumophila,* is a Gram-negative bacterium found naturally in aquatic environments (McDade et al., 1977). It infects alveolar macrophages when contaminated aerosol is inhaled. *L. pneumophila* then uses the Dot/Icm type IV secretion system to translocate bacterial proteins, termed effectors, into the host cell where they manipulate eukaryotic processes to create a replicative niche termed the *Legionella*-containing vacuole (LCV) and avoid lysosomal fusion (Vogel et al., 1998). Approximately 300 effectors are injected, many of which are redundant (Burstein et al., 2009; Luo and Isberg, 2004). The *L. pneumophila* genome codes for a high number of eukaryotic-like proteins that interfere with the host through molecular mimicry. Ankyrin B (AnkB) is one of very few effectors essential for bacterial replication within human macrophages and amoeba (Al-Khodor et al., 2008) and is conserved across the sequenced *L. pneumophila* genomes (Burstein et al., 2016; Cazalet et al., 2004; Chien et al., 2004; Glockner et al., 2008). Bioinformatic analysis predicts that AnkB of strain AA100/130b (lpg2144) contains a N-terminal F-box domain, a two-repeat ankyrin domain (Cazalet et al., 2004; de Felipe et al., 2005) and a C-terminal CaaX farnesylation motif (Ivanov et al., 2010; Price et al., 2010b) (**Fig. 2.1**).

F-box containing proteins are part of Skp1-Cullin-F-box (SCF) E3 ubiquitin ligase complexes, which transfer ubiquitin from an E2 ubiquitin conjugating enzyme to a target protein (Lyapina et al., 1998; Skowyra et al., 1997). The F-box mediates interactions with Skp1, which in turn attaches to Cullin and the E2 enzyme. In eukaryotes, the F-box is typically paired with a protein-protein interaction domain that confers substrate specificity. These domains are typically either tryptophan-aspartate (WD) or leucine-rich (LRR) repeats (Price and Kwaik, 2010; Uro-

Coste et al., 1998; Zheng et al., 2002). F-box proteins without an extra interaction domain exist as the FBXO (F-box only) type (Kipreos and Pagano, 2000). In AnkB, the F-box is paired with ankyrin repeats – a ~33 residue helix-turn-helix repeat motif that mediates protein interactions (Li et al., 2006). The association of an F-box domain and ankyrin repeats is unusual and not found in proteins in metazoans.

The function of AnkB in cells and the reason it is required for *Legionella* intracellular growth are not clear. A null mutant of *ankB* exhibits severe intracellular defect in the protozoan host *Acanthamoeba polyphaga*, and human macrophages (Al-Khodor et al., 2008). Mutagenesis studies have shown that both the F-box and farnesylation motif are required for AnkB function of strain A100/130b (Al-Quadan et al., 2011; Ivanov et al., 2010; Price et al., 2009; Price et al., 2010b). Here, we present the crystal structure of AnkB in complex with human Skp1, revealing the specific host-pathogen interactions by which AnkB takes control of the host ubiquitin-proteasome system. We identify a protein-protein interaction site in the ankyrin domain for putative substrates and use mutagenesis and *in vivo* functional assays to show the ankyrin repeats are critical for poly-ubiquitination of the LCV and pathogen survival.

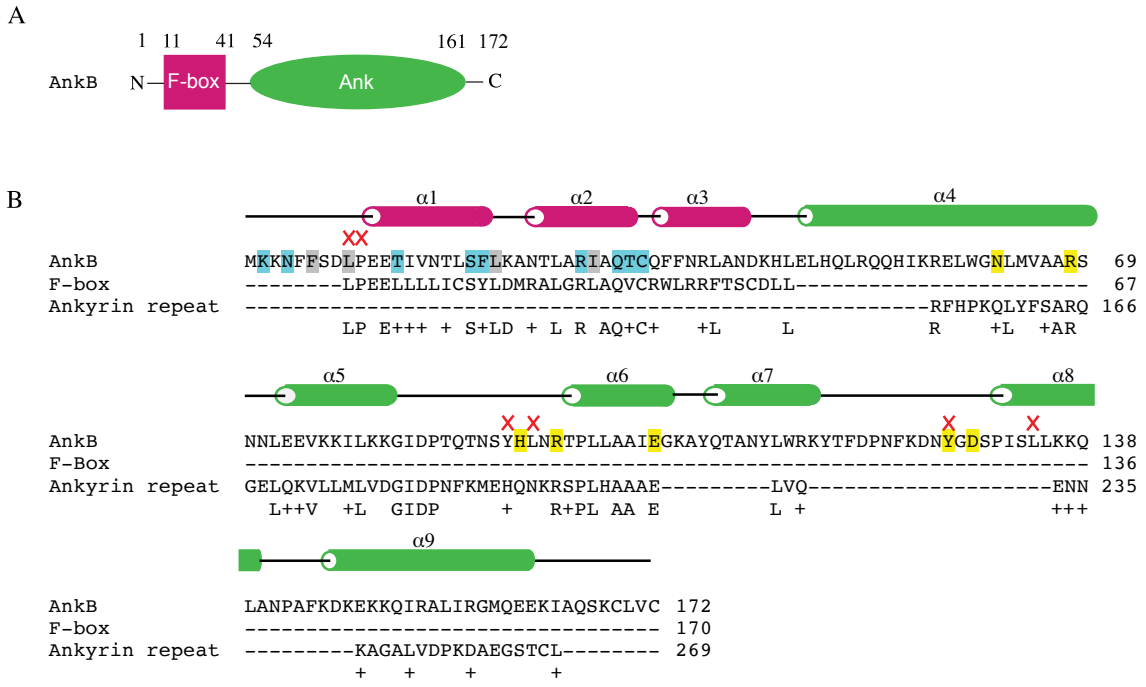


Fig. 2.1. Domain organization of AnkB. (*A*) AnkB is composed of two domains: an F-box (pink) and ankyrin repeats (green). (*B*) Sequence alignment of AnkB with human proteins containing an F-box [EAW49753.1] and ankyrin repeats [AAH11608.2]. The secondary structure elements and important residues in 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.

2.3 Results

2.3.1 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. 2.2A**). 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. 2.2B and Fig. 2.3**) (Schulman et al., 2000).

We also solved the structure of the isolated ankyrin domain (residues 54 to 168) to close to 1 Å resolution. The 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. 2.1B and 2.2A**) (Parra et al., 2015). The majority of ankyrin domains contain four to seven repeats, while up to 34 repeats have been reported (Li et al., 2006). The largest sequence differences generally occur in the loop regions and confer binding specificity. Refinement statistics for both structures are shown in **Table 2.1**.

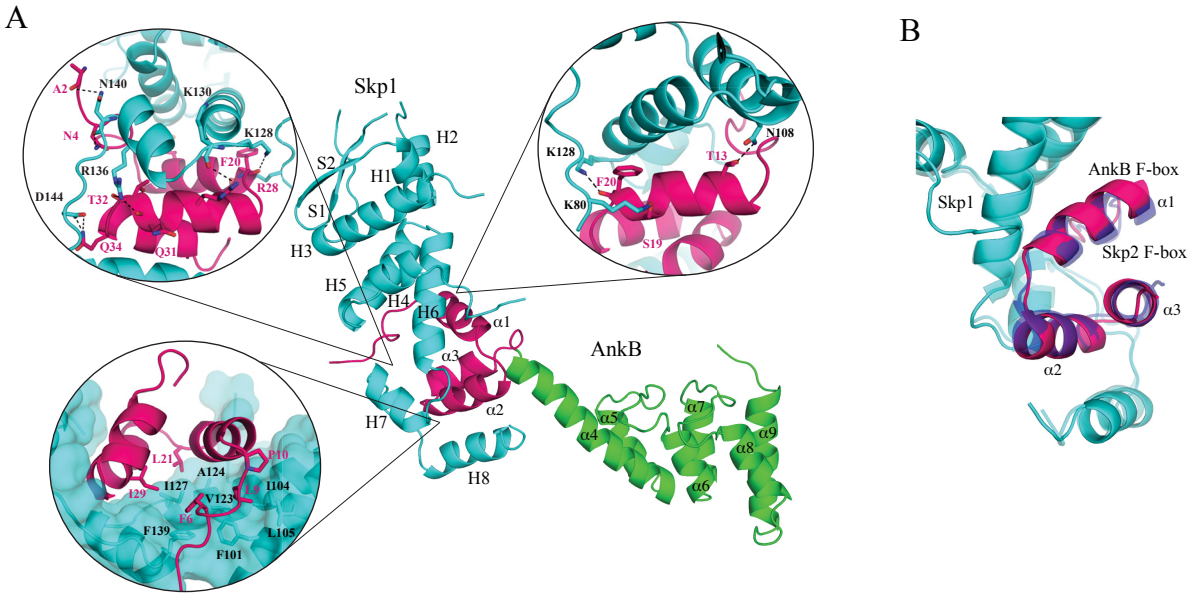


Fig. 2.2 Structure of AnkB as a component of the E3 ubiquitin ligase complex. (*A*) Co-crystal of AnkB (*pink,* F-box; *green,* ankyrin repeats) and Skp1 (*cyan*). Hydrogen bonds between AnkB and human Skp1 are highlighted. Residues involved in the hydrophobic interaction surface between Skp1 and F-box 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 (Price et al., 2009). 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.

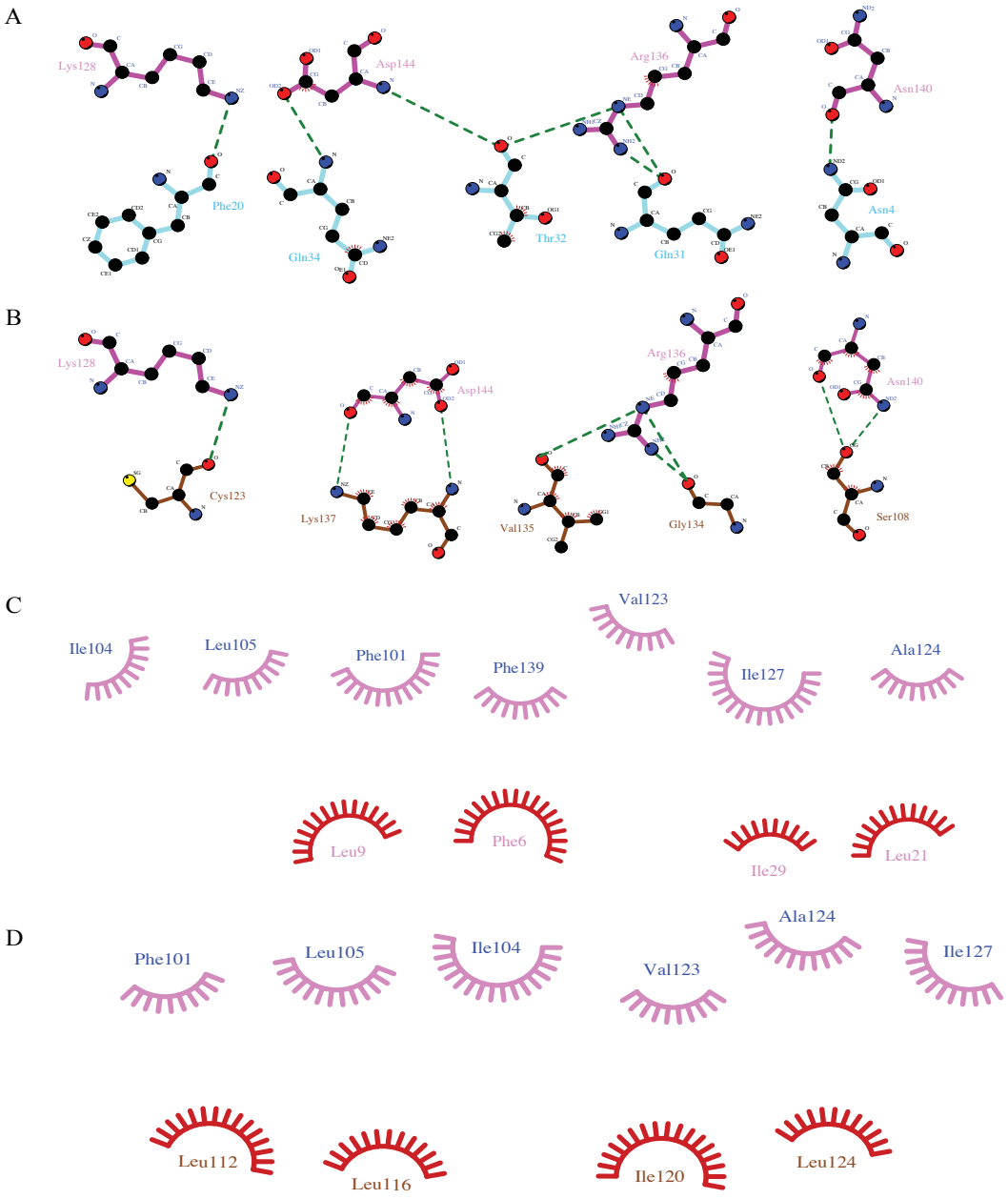


Fig. 2.3 Schematic diagrams of the interaction between Skp1 and AnkB or Skp2. (*A*) Hydrogen bonding network between Skp1 and AnkB. (*B*) Hydrogen bonding network between Skp1 and Skp2. (*C*) Hydrophobic interactions between Skp1 and AnkB. (*D*) Hydrophobic interactions between Skp1 and Skp2. Skp1-AnkB interactions are based on the complex structure presented in this paper. Skp1-Skp2 interactions are based on PDB code 1FQV. The interactions are similar, with the same Skp1 residues contacting the two F-boxes. Residues are labeled in blue for Skp1, pink for AnkB, and brown for Skp2. Hydrogen bonds are shown as green dashed lines, and hydrophobic interactions are indicated with starbursts.

Table 2.1 Data collection and refinement statistics for AnkB

	54-168 PDB code 5K34	1-168/Skp1(1-163) PDB code 5K35
Data collection		
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)
$CC_{1/2}$ in highest shell	0.842	0.920
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
Water	166	5
B-factors		
Protein	12.8	40.4
Water	24.5	48.7
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.

2.3.2 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. 2.2A**). 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 (Price et al., 2009). 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 (Price et al., 2009). 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. 2.2A**). Both residues are highly conserved among F-box domains (Kipreos and Pagano, 2000) and their mutation to alanine likely prevents proper folding of the AnkB F-box domain.

2.3.3 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. 2.4A**). The contacts also occurred in the crystals of the AnkB-Skp1 complex that adopt a different space group (**Fig. 2.5**). 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 (Xu et al., 2012), showed smaller shifts suggestive of weaker binding (**Fig. 2.4B**).

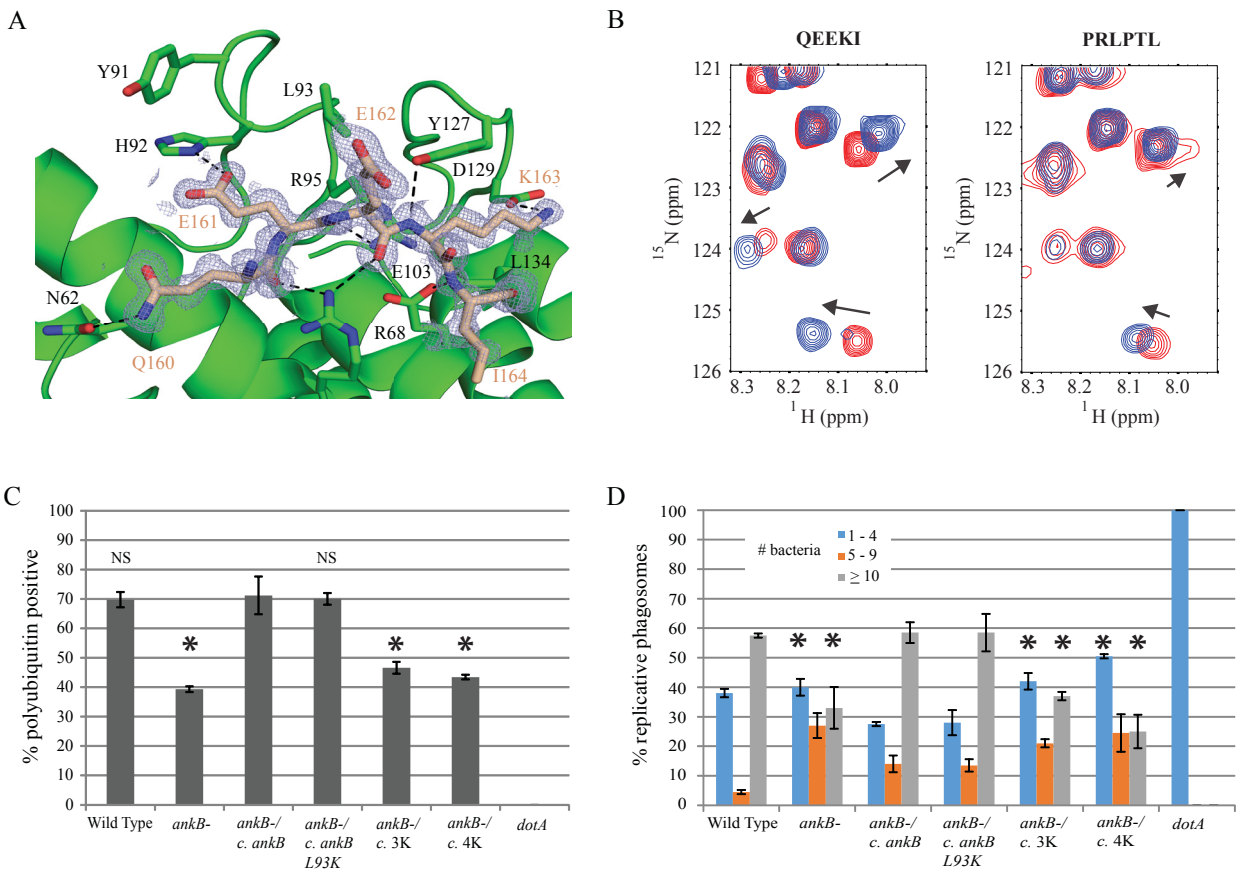


Fig. 2.4 Substrate binding by the ankyrin repeats. (*A*) Crystal contacts in the AnkB ankyrin repeats mimic substrate binding. The C-terminal tail (residues 160 to 164, QEEKI) of one molecule (*wheat*) binds to the ankyrin repeats of another (*green*). Hydrogen bonds are indicated by dashed black lines. 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. (*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). (*C*) 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. (*D*) 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*.

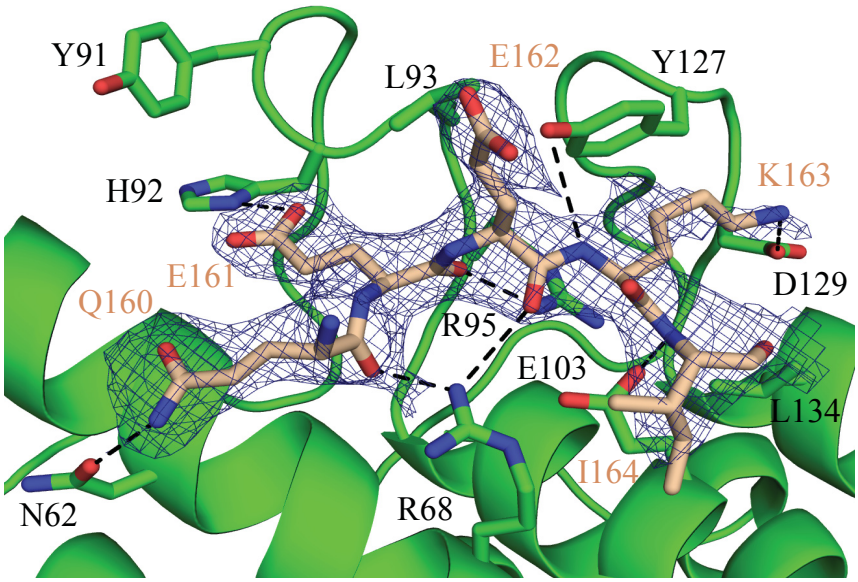


Figure 2.5. Crystal contacts from the Skp1-AnkB complex structure mimic substrate binding. The C-terminal tail (residues 160 to 164, QEEKI) of one AnkB molecule (wheat) binds to the ankyrin repeats of another (green). Hydrogen bonds are indicated by dashed black lines. The $2F_O$-F_C omit map (blue) is contoured at 1σ.

2.3.4 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 (**Fig. 2.6**). 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 70% of the LCVs of the wild-type strain-infected cells stained positively for polyubiquitin, whereas only 39% 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 70% of the LCVs staining positively (**Fig. 2.4C**). 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. In contrast, the single lysine mutants and the Y91K L93K double mutant were functionally similar to wild-type *ankB* in the intracellular growth kinetics in U937 macrophages (**Fig. 2.7**). This confirms the importance of the substrate binding site on the ankyrin domain for the recruitment of polyubiquitinated proteins to the LCV (Bruckert and Abu Kwaik, 2015a).

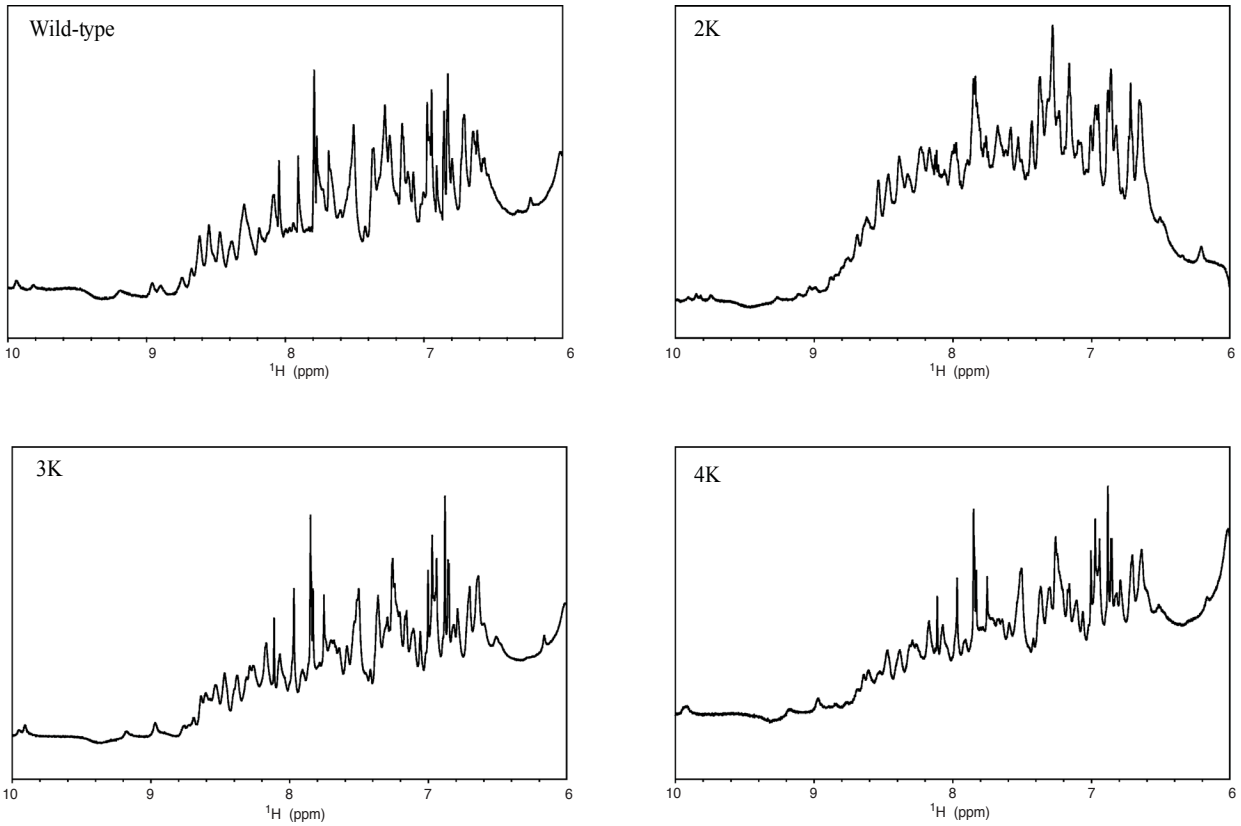


Figure 2.6 Stability of AnkB mutants. NMR spectra of the wild-type and the 2K (Y91K/L93K), 3K (Y91K/L93K/Y127K) and 4K (Y91K/L93K/Y127K/L134K) mutants. Spectra indicate proper folding of the mutants.

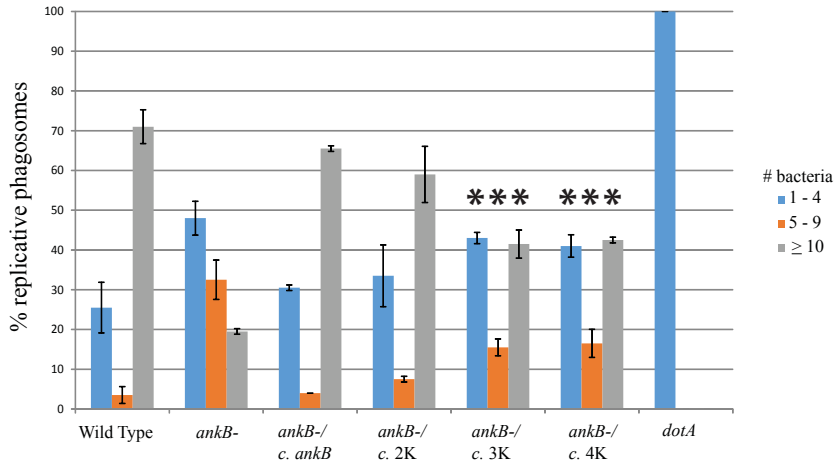


Figure 2.7 Substrate binding site required for survival in U937 macrophages. U937 macrophages were infected with wild type, *dotA* mutant, *ankB* mutant, or the *ankB* mutant complemented with either wild type *ankB* or *ankB* containing the indicated mutations at an MOI of 1 followed by 1-hour treatment 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. Error bars indicate +/- 1 SD. *p <0.05 compared to *ankB* mutant complemented with wild type *ankB*.

2.3.5 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. 2.4D**). 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 (**Fig. 2.7**). 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.

2.4 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 (Herbert et al., 2015). 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 (Price et al., 2009).

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 (Parra et al., 2015). 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 (Xu et al., 2012); we observed low affinity binding of AnkB to the peptide PRLPTL.

The ankyrin domain of AnkB shows broad specificity. This is typical of ankyrin domains that bind unfolded or extended peptide sequences (Parra et al., 2015) 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. 2.4B**). 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. 2.4C & D**).

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 high similarity, and a conservation of the substrate-binding site (**Fig. 2.8A**). 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

(Lomma et al., 2010). ParvB functions in regulating the actin cytoskeleton for cell adhesion and migration (Legate et al., 2006). Overexpression of AnkB competed with endogenous ubiquitin ligase for ParvB interaction and decreased ParvB ubiquitination (Lomma et al., 2010). 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 (Bruckert and Abu Kwaik, 2015b).

Studies have elucidated two roles of AnkB in *Legionella* virulence. Following phagocytosis, Legionella injects effector proteins into the host cell cytosol via the Dot/Icm secretion system (**Fig. 2.9**). 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. 2.4C & D**). 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 (Al-Quadan et al., 2011; Price et al., 2010b); however, there is strain specificity as AnkB from the *Legionella* strain Paris lacks the CaaX motif but retains function (Lomma et al., 2010). 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. 2.8B**) (Zheng et al., 2002; Zheng et al., 2000). 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. Lys48-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 *Legionella*-containing vacuole (LCV). As Lys48-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 (Bruckert and Abu Kwaik, 2015b).

AnkB also plays a role in enriching the cytosolic pool of free amino acids through triggering Lys^{48}-linked polyubiquitination and increased protein turnover (Price et al., 2011). The levels of amino acids in the infected host cell are insufficient sources of carbon, nitrogen and energy for *L. pneumophila* (Price et al., 2014). AnkB promotes intra-vacuolar proliferation by ubiquitinating host proteins for their degradation into free amino acids (Bruckert and Abu Kwaik, 2015a; Price et al., 2011). 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 (Bruckert et al., 2014; Price et al., 2011). 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.

A

```
AnkB            MKKNFFSDLPEETIVNTLSFLKANTLARIAQTCQFFNRLANDKHLELHQLRQQHIKRELWGNLMVAARS  69
lpc1593         -----------------MSFLKANTLGRIAQTCQFFNRLANDKHLELHQLRQQRIKRELWGNLMVAARS  52
lpp2082         MKKNFFSDLPEETIVNTLSFLKANTLGRIAQTCQFFNRLANDKHLELHQLRQQRIKRELWGNLMMAARS  69
lpl2072         MKKNFFSDLPEETIVNTLSFLKANTLGRIAQTCQFFNRLANDKHLELHQLRQQRIKRELWGNLMVAARS  69

AnkB            NNLEEVKKILKKGIDPTQTNSYHLNRTPLLAAIEGKAYQTANYLWRKYTFDPNFKDNYGDSPISLLKKQ 138
lpc1593         NNLEEVKKILKKGIDPAQTNSYHLNRTPLLAAIEGKAYQTANYLWRKYTFDPNFKDNYGDSPISLLKKQ 121
lpp2082         NNLEEVKKILKKGIDPAQTNSYHLNRTPLLAAIEGKAYQTANYLWRKYTFDPNFKDNYGDSPISLLKKQ 138
lpl2072         NNLEEVKNILKKRIDPAQTNSYHLNRTPLLAAIEGKAYQTANYLWRKYTFEPNFKDNYGDSPISLLKKQ 138

AnkB            LANPAFKDKEKKQIRALIRGMQEEKIAQSKCLVC                                    172
lpc1593         LANPAFTDKEKKQIRALIRGMQEEKIAQSKCLVC                                    155
lpp2082         LANPAFKDKEKNKYAP------------------                                    154
lpl2072         LANPAFKDKEKKQIRALIREMQEEKIAQSKCLVC                                    154
```

B

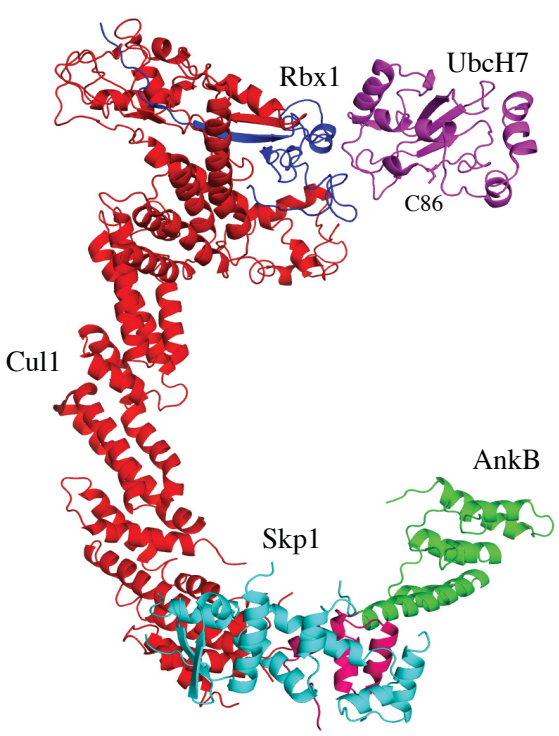


Figure 2.8 Model of AnkB as part of the SCF ubiquitin ligase complex. (*A*) Sequence alignment of AnkB in other strains. Differences between AnkB (*Pneumophila*), lpc1593 (*Corby*), lpp2082 (*Paris*) and lpl2072 (*Lens*) are highlighted in green. (*B*) Model of AnkB in context of ubiquitination. Model of AnkB as the F-box protein in the SCF E3 ubiquitin ligase complex (PDB 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).

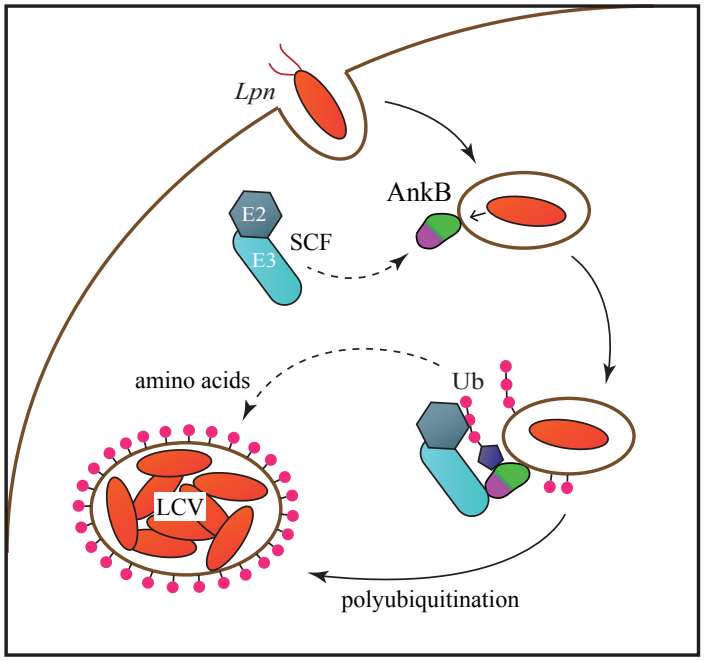


Figure 2.9 Role of AnkB in LCV maturation. After phagocytosis, *L. pneumophila* secretes AnkB which remains anchored to the vacuole via its farnesylation motif. AnkB then recruits the SCF E3 ubiquitin ligase complex through its F-box. Increased ubiquitination of host proteins leads to increased free amino acid production and the accumulation of ubiquitin on the mature LCV to prevent fusion with lysosomes.

2.5 Experimental procedures

2.5.1 Cloning, protein expression, and purification. 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.

2.5.2 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 (Otwinowski Z, 1997).

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 (Sheldrick, 2008). The initial model was built with ARP/wARP (Langer et al., 2008) and refined with Refmac5 (Murshudov et al., 2011). 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 (Zheng et al., 2002). The model was built by ARP/wARP (Langer et al., 2008), completed with Coot (Emsley and Cowtan, 2004) and improved by several cycles of refinement using Refmac5 (Murshudov et al., 2011). Water molecules were added in the last stage of refinement.

The refinement statistics are shown in **Table 2.1**. The final ankyrin domain and complex structures respectively have 0 and 1 outlier in the Ramachandran plot computed using MolProbity (Chen et al., 2010).

2.5.3 NMR spectroscopy. ^{15}N-^{1}H heteronuclear single quantum correlation spectroscopy and 1D experiments were performed at 25 ºC on a Bruker 600 MHz spectrometer. Samples of the AnkB (54-168) were prepared at 0.28 mM in 90% buffer B and 10% D_2O. Titrations were performed by the stepwise addition of QEEKI, PRLPTL, ^{Ac}QEEAI^{NH2}, ^{Ac}QAEKI^{NH2}, ^{Ac}AEEKI^{NH2}, ^{Ac}QEERI^{NH2}, ^{Ac}QEEYI^{NH2} peptides (Bio Basic Inc.). The highest peptide

concentration was 8 mM. NMR spectra were processed with NMRPipe (Delaglio et al., 1995) and analyzed with SPARKY (Goddard and Kneller).

2.5.4 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 ºC prior to infection as previously described (Al-Khodor et al., 2008). 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 LB and antibiotics were used at a concentration of 100 µg/mL for ampicillin and 40 µg/mL for chloramphenicol.

Purification and preparation of human monocyte-derived macrophages (hMDMs) was performed as previously described (Habyarimana et al., 2008). 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 ºC and 5% CO_2 in RPMI 1640 supplemented with 20% FBS. Monocytes were then counted and resuspended in RPMI 1640 supplemented with 10% FBS and plated on coverslips at a density of 3×10^5 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. Maintenance of U937 macrophages was performed as described previously (Habyarimana et al., 2008).

Infection of hMDMs or U937 cells was performed as previously described (Habyarimana et al., 2008). Bacteria were suspended in RPMI 1640 with 10% FBS and macrophages were infected in duplicate for 1 hour at a multiplicity of infection (MOI) of 50. 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 and then RPMI containing 10% FBS was added. At 10 hours post infection, cells were fixed in 100% cold methanol and processed for confocal microscopy.

2.5.5 Confocal microscopy. Processing of infected cells for confocal microscopy was performed as we described previously (Price et al., 2009). Rabbit polyclonal anti-*L. pneumophila* was used at a dilution of 1/1000 and detected by Alexa-Fluor 488-conjugated donkey anti-rabbit IgG (Invitrogen). Polyubiquitinated proteins were detected using mouse anti-polyubiquitin FK1 antibody at a dilution of 1/50 (Enzo Life Sciences), followed by Alexa-Fluor 647-conjugated goat anti-mouse IgM (Invitrogen). An Olympus FV1000 laser scanning confocal microscope was used to examine cells as we described previously (Price et al., 2009). One hundred cells were examined in duplicate for each strain for both ubiquitin recruitment and replication.

2.6 Acknowledgments

K.W. acknowledges a studentship from Canadian Institutes of Health Research (CIHR). This work was funded by a CIHR Genomics grant GSP-48370. Y.A.K is supported by Public Health Service Award 1R01AI120244 and 1R21AI116517 from NIAID and by the commonwealth of Kentucky Research Challenge Trust Fund. Crystallographic data were acquired at the Macromolecular Diffraction (MacCHESS) facility at the Cornell High Energy Synchrotron Source (CHESS) supported by NSF award DMR-0225180 and NIH/NCRR award RR-01646. The authors declare no conflict of interest.

Chapter 3: Ankyrin repeats as a dimerization module

3.0 Connecting text

A common feature of bacterial effectors, is the adoption of eukaryotic-like domains for host mimicry. AnkB contains two such domains, an F-box domain and an ankyrin domain, allowing interaction with host proteins. The ankyrin domain exists in at least ten other effectors in sequenced genomes. AnkC is one such virulence factor (Kozlov et al., 2018).

3.1 Summary

Legionella pneumophila is a pathogen, causing severe pneumonia in humans called Legionnaires' disease. Ankyrin C (AnkC/LegA12) is a poorly characterized 495-residue effector protein conserved in multiple Legionella species. Here, we report the crystal structure of a C-terminally truncated AnkC (2-384) at 3.2 Å resolution. The structure shows seven ankyrin repeats (ARs) with unique structural features. AnkC forms a dimer along the outer surface of loops between ARs. The dimer exists both in the crystal form and in solution, as shown by analytical ultracentrifugation. This is the first example of ARs as a dimerization module as opposed to solely a protein interaction domain. In addition, a novel α-helix insert between AR3-AR4 is positioned across the surface opposite the ankyrin groove. Sequence conservation suggests that the ankyrin groove of AnkC is a functional site that interacts with binding targets. This ankyrin domain structure is an important step towards a functional characterization of AnkC.

3.2 Introduction

Legionella pneumophila is the causative agent of Legionnaires' disease, a severe form of pneumonia. When engulfed by alveolar macrophages, *L. pneumophila* uses the Dot/Icm type IV

secretion system to translocate bacterial proteins, termed effectors, into the host cell where they manipulate eukaryotic processes to create a replicative niche termed the Legionella-containing vacuole (LCV) and avoid lysosomal fusion (Vogel et al., 1998). The *L. pneumophila* genome codes for a high number of eukaryotic-like proteins that interfere with the host through molecular mimicry. One such group of effectors contains at least eleven eukaryotic-like ankyrin (Ank) proteins shared by multiple Legionella genomes that have been sequenced (AnkB/LegAU13, AnkC/LegA12, AnkD/LegA15, AnkF/LegA14, AnkG/AnkZ/LegA7, AnkH/AnkW/LegA3, AnkI/LegAS4, AnkJ/LegA11, AnkK/LegA5, AnkQ/LegA10, AnkN/AnkX, LegA8) (Al-Khodor et al., 2010; Pan et al., 2008).

The expression of AnkC in HEK293T cells caused more than a 3-fold increase in nuclear factor NF-κB activity relative to an empty vector (Losick et al., 2010). AnkC also causes selective defects in carboxypeptidase Y trafficking in yeast-based assays (Heidtman et al., 2009). Evidence for AnkC translocation in *Legionella pneumophila* was originally obtained using the β-lactamase fusion assay (de Felipe et al., 2008; Habyarimana et al., 2008). While these results demonstrated the ability of AnkC to affect host cellular pathways, more studies are necessary to clarify its role in pathogenesis.

AnkC of *Legionella pneumophila* Philadelphia (lpg0483, NCBI reference YP_094527) is a 495-residue protein. The secondary structure prediction identifies an unstructured ~35-residue region that potentially divides the protein into a 390-residue N-terminal domain and a small 60-residue C-terminal domain. The N-terminal domain contains three predicted ankyrin repeats, while the C-terminal domain has a low sequence similarity to part of a bacterial glucosamine-6-phosphate deaminase, though the functional relevance of that is unclear.

Ankyrin repeat domains primarily mediate protein-protein interactions and participate in

intracellular signaling. The signature Ank repeat (AR) consists of approximately 33 residues forming a pair of α-helices connected by a β-hairpin. With the helical tandems stacking together, the hairpin loops project outwards at almost 90 degrees to the helical stacks, thus forming a surface (ankyrin groove) used for binding other proteins. Many ankyrin proteins contain three or more ARs (Li et al., 2006). In Legionella effectors, the Ank domain is often found in combination with other functional domains such as the F-box (AnkB) and Fic (filamentation-induced by c-AMP) domains (AnkX) among others. Crystal structures of LegA1 (lpg2416), AnkB (LegAU13, lpg2144) in complex with its host target Skp1 (Wong et al., 2017), AnkX (LegA8, lpg0695) (Campanacci et al., 2013), LegAS4 (Son et al., 2015) and the Ank domain of LegA11 (lpg0436) have been previously determined.

Here, we determined a crystal structure of the N-terminal domain of AnkC. The structure reveals seven ankyrin repeats capped by a bundle of α-helices on the C-terminal side. Despite overall similarity to previously characterized ankyrin repeats, AnkC displays distinct structural features, such as being a novel dimerization module, and warrants further studies.

3.3 Results and Discussion

3.3.1 Crystallization of AnkC. Initial attempts to produce full-length lpg0483 using *E. coli*-based expression were unsuccessful. Consequently, we made several truncated constructs, all of which included the predicted Ank domain (residues 146-246). The largest construct containing residues 2-386 expressed with high yields and produced small crystals. Extensive optimization combined with the use of additive screens resulted in bigger crystals diffracting to 3.5-4 Å. In order to solve the phase problem, the protein was labeled with selenomethionine, but the labeled protein failed to produce crystals in either crystallization condition working for the native protein

or using *ab initio* screening. To solve this problem, we designed shorter truncations 2-380, 2-382 and 2-384 for crystallization trials. All three constructs produced crystals for native proteins with the biggest crystals resulting from the 2-384 construct. Importantly, selenomethionine-labeled proteins also crystallized with the 2-382 construct yielding the biggest crystals.

The best selenomethionine-labeled crystals diffracted to approximately 3.8 Å and their phasing proved to be challenging. The use of a novel program PRASA (Skubak, in preparation) in combination with Crank2 (Skubak and Pannu, 2013) for obtaining the anomalously scattering selenium substructure, SAD phasing and model building was critical to obtain the initial model. Due to the weak anomalous signal and the relatively low data resolution, multiple iterations of the Crank2 combined model-building algorithm were needed to build the model. Within each iteration, several parallel jobs differing by substructure improvement parameters were started and the model providing the best R-free was passed to the next iteration. The best model, with a majority of the structure being built with an R-free value of 33% occurred after the sixth iteration. This model was subsequently used to phase a native dataset that was refined to 3.2 Å (**Fig. 3.1**).

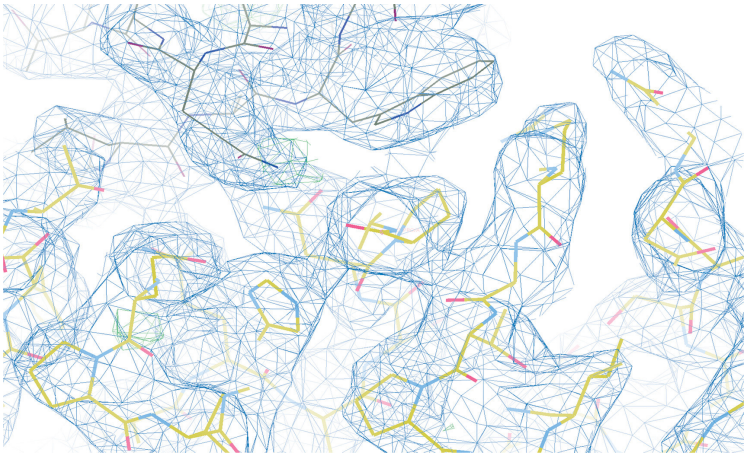


Figure 3.1 Sample of the electron density map at the AnkC dimerization interface. The $2F_o$-F_c map is contoured at 1σ.

3.3.2 AnkC structure. The structure contains one AnkC molecule in the asymmetric unit. We observed residues 3-366 in the electron density map. The N-terminal His-tag, Asp2, the C-terminal tail 367-384 and several internal regions are absent in the model due to disorder. The majority of the construct comprises seven ankyrin repeats (residues 3-296) (**Fig. 3.2**). Interestingly, only residues 146-246 accounting for ankyrin repeats 4-6 were annotated as such. Analysis of the structure shows the challenges for accurately predicting ankyrin repeats in low-homology bacterial proteins. The tandem helices of ankyrin repeats are typically connected by a short (~3 residues) β-turn, but AR2 has a disordered 12-residue insert (Ser56-Arg67) in the crystal. Furthermore, the β-hairpin loop is not formed between AR2 and AR3 because the sequence connecting them is too short. An even more unusual feature of the AnkC structure is an approximately 35-residue insert between AR3 and AR4. The middle part of that insert forms an α-helix, which binds across AR4-AR5 on the opposite side of the ankyrin groove. The helix-preceding and following segments are disordered in the crystal. Sequence alignments show that this insert is not conserved in more than half of Legionella species (data not shown). A C-terminal four-helical bundle then caps AR7. The fold of the fragment is completed by a helical turn Ser362-Leu366 binding to the C-capping sub-domain.


A
α1 AR1 α2 α3 AR2
pneumophila FVSEMNKILQMELEQFKEFIQKKLEEDKNYLEKLFWLPNGSQMTVLNYLIEQYSEPKL-- 60
micdadei SFRELEKAIAAGEEAFNAFILKEIENDEQFLRRRFLVRKNKQLTVLNYLIYKHQKGLGNP 62
longbeachae FITEINAAVKSGLQGFKLFIEQKLCVEPNYLEHPFWIENGEQVTVLNYLIQQHQEKEGKQ 62
dumoffii FIREINGAIKAGLQGFQEFIEKKLLVDKRYLEHPFWMPGGAQLTVLNYLIQIHQRNNDDS 62
parisiensis FITEINAAIKSGLEGFNLFIEQKLQIETDYLERPYWTEGGVQLTVLNYLIQQHEERNKNP 62
α4 α5 AR3 α6
pneumophila GIDDANRDLLAKIDFVLHEAKDVNVGEPLHQAIVAGKISLALHLLGVDENNFTPGESEKT 120
micdadei HMQKHIKMLLDF-------GAPFNSDCSIHLLLRLQKFELFSLF--------MPSSEEV- 106
longbeachae IHKGFSSDLTQFVDFILSKVKDKNTGEPLHQALALGKLKLALHLFDT------------ 109
dumoffii VEKAIPVDLTAFIDCVLNKIQNKNCGLPLHQAIAEGKVELALHLLDT------------ 109
parisiensis VHEENSSDLIGFIDVVLSKVRDKNIGSPLHQAIAANKIKLAFHLLEN------------ 109
α7 α8 AR4 α9
pneumophila DVLSILSKVRKKIEESFNHVKRCFFDVDKRDGYGRTLLSLALDAKRQELLIAILARNPIV 180
micdadei -----------------ASEIIQKINFDNTDIEGRTLLSRAIDTKHPDVIRRLLKNNFPI 149
longbeachae ----------------------GVFDVNRRDQEGRTLLSLVVETKDRRLLRRLLKSHPDV 147
dumoffii ----------------------NVFDVNHRDEQGHTLLSLALPTQNRQLLTKILSCRPDV 147
parisiensis ----------------------NVFDVNKRDEEARTLLSLVLPIRNRQLLLKILACKPDV 147
α10 AR5 α11 α12
pneumophila HATT----LRSSAYVPFQPIHQAVVLDYAEGITLLASMGAQLTNPLGSMRDTPVILAARL 236
micdadei DVNKPSRFDIETQFAEIQPLHQAVVNDFPHAIYYLLENGAILDSPCGELDETPILLSARL 209
longbeachae NATT----RMTNARVSFQPLHQAIALDFAFGVKSLANAGADLSNPLGAMKDTPVLLAARQ 203
dumoffii NAAT----LFKEHGVPFQPLHQAIVLDFISGVTNLVRAGADLSNPVGVKKETPLLLAARK 203
parisiensis NAAT----RLTEERIPFQPLHQAIALDFASGVINLVRSGADVANPVGELKDTPVLLAAQQ 203
AR6 α13 α14 AR7 α15
pneumophila GKINALAALLE-LP-TQSLSLESENNHLFEDKQTGHTAVEELCERMANENDKADALRGIA 294
micdadei GKIKSLEVLLAATAGKNHLDLEATNNQF-------DRPIDLLCKRLEEKKNPQEALCGIA 262
longbeachae GKIKALEALLE-FP-VKKLDLEAENNRLSADKMTADNAIESLCKLLANDKNNKDLIRGVA 261
dumoffii GRINVLEVLLDEAP-IENLKLEAENNAL--DKDSGDTAIEALCKRLAQDKNNKELIRGAA 260
parisiensis GKIKALEALLEESP-VEQLKLEARNKNFGEDKQTGENAIEALCKKLVNNRNDKDLLHGVA 262
α16 α17 α18 α19
pneumophila MLICRGAEPPRNEKMRNLLSSNRVAFLKAVSTYLADKPQLVDAFVERCHLRESALHNIVY 354
micdadei MLLCHGAKAPREEKFRDLLIKNRFALIDEVKKYKNKTGNSVAKFVRACHDKNNPLHDIMY 322
longbeachae MLLCRGAEPPRKESMCQLLADKRTDLLKEIDRYLNSRPELVDPFVNRCHLAESPLHKIIY 321
dumoffii MLLCRGAEPPRSEELCQLLADKRGALLKEIDNYMDTRPQLVDSFVNRCHLTDTALHRIIY 320
parisiensis MLLCRGAEPPQDEELCQLLAGKRGDLLKEIDDYMDTRPELVDPFVNRCHLANSALYNIIY 322
pneumophila ADHSWGSSIRHL 366
micdadei AN-TWFHALWRY 333
longbeachae VDNSWGNTLRQL 333
dumoffii VDHSWGSTLRQL 332
parisiensis VDHSWGSTLRQL 334
B
sp.40-6
rubrilucens
brunensis
massiliensis
micdadei
fallonii
moravica
shakespearei
waltersii
pneumophila
norrlandica
gratiana
longbeachae
cherrii
dumoffii
tucsonensis
parisiensis
C
α1
α2
α3
α4
α5
α6
α7
α8
α9
α10
α11
α12
α13
α14
α15
α16
α17
α18
α19

Figure 3.2 AnkC is conserved in different *Legionella* species. (*A*) Sequence alignment of Ank domains from *Legionella pneumophila* (NCBI reference YP_094527), *Tatlockia micdadei* (WP_045098608), *Legionella longbeachae* (WP_003635333), *Fluoribacter dumoffii* (WP_010654633) and *Legionella parisiensis* (WP_058517280). The positions of α-helices (α1-19) and ankyrin repeats (AR1-7) are shown according to the structure from *L. pneumophila*. AnkC from *Legionella pneumophila* contains an additional helix α7 that is absent in many other species. (*B*) Unrooted phylogenetic tree of AnkC from a subset of *Legionella* species. Sequences in panel A represent proteins from different branches of the phylogenetic tree. The figure was generated with Clustal Omega (Sievers et al., 2011) and Phylogeny.fr (Dereeper et al., 2008). (*C*) Cartoon representation of AnkC with the molecule colored in rainbow, from blue at the N-terminus to red at the C-terminus. Disordered segments are shown by a dashed line. This and subsequent figures were prepared with PyMol (www.pymol.org).

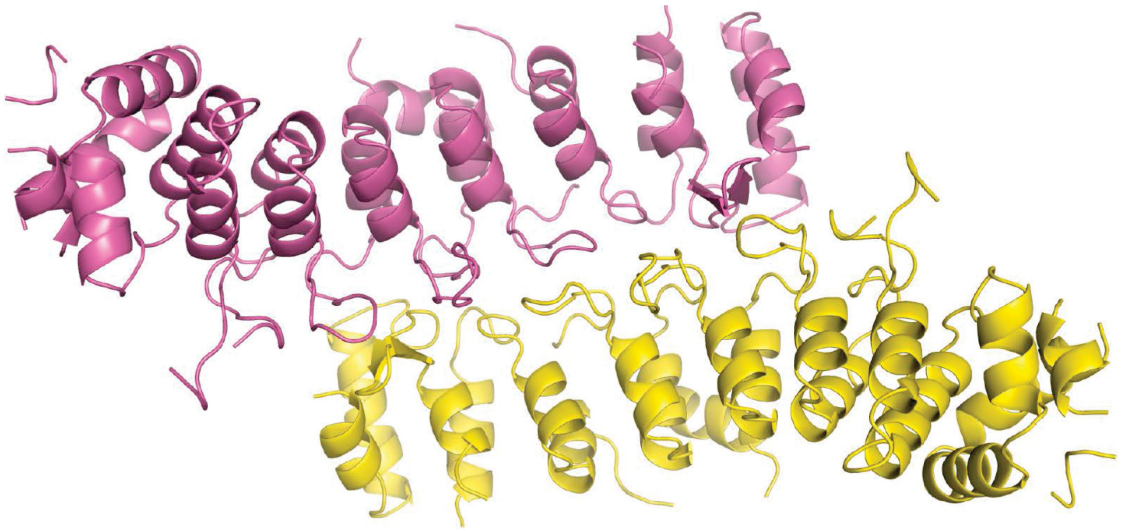
3.3.3 Ankyrin repeats mediate dimerization. Analysis of crystal contacts reveals two interesting features. AnkC forms a symmetric dimer in the crystal, where repeats AR2-AR6 create continuous surface contacts using the outer side of hairpin loops, keeping the ankyrin groove open for protein interaction (**Fig. 3.3A**). In order to test the relevance of this dimer, the oligomerization state of AnkC in solution was investigated using sedimentation equilibrium analytical ultracentrifugation (AUC). The predicted molecular weight of 93 ± 4 kDa is very similar to that expected for a dimer (**Fig. 3.3B & C and Fig. 3.4**). This result proves that the observed dimer is a natural dimer and not a crystallization artifact. In addition, the C-terminal helical turn (Ser362-Leu366) makes contacts with the N-terminal part of another AnkC molecule, suggesting a functional protein-binding site in the N-terminus (**Fig. 3.5**).

We analyzed sequence conservation of AnkC in *Legionella* species to identify functionally important sites of the protein (**Fig. 3.3D**). Some of the conserved residues form the hydrophobic core of the ankyrin repeats and clearly play a structural role in the overall fold. Interestingly, some of the most conserved residues are positioned on the dimer interface, highlighting the physiological

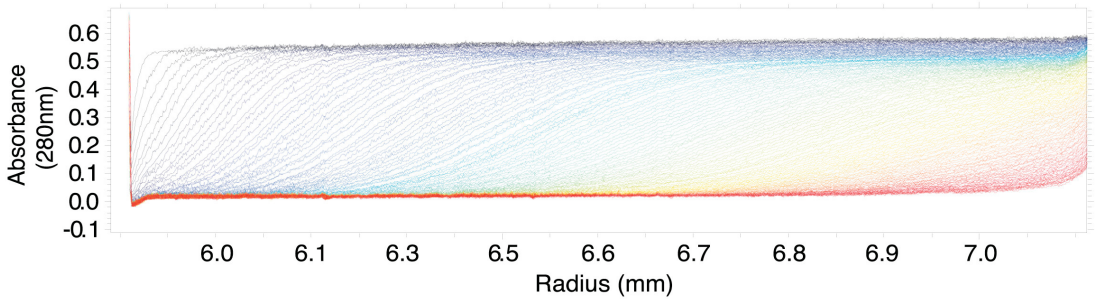
relevance of this ankyrin dimer. More importantly, the typical protein-binding surface of the ankyrin fold on the interface of hairpin loops and α-helices, or the ankyrin groove, contains a large number of conserved residues. This suggests the surface is functional for protein binding. Finally, the C-capping helices possess some highly conserved surface residues. Some of which could potentially bind to and stabilize the C-terminal domain of AnkC, though the functional relevance of that is currently unclear.

Analysis of surface charges shows that the ankyrin groove is largely positively charged (**Fig. 3.5**). This would suggest a preference for binding to negatively charged protein partners. In addition, there are distinct positively charged surfaces in the C-capping region and the inserted helix α7 on the opposite side of ankyrin groove. Coupled with high sequence conservation, the C-capping region may have a functional importance. On the other hand, the α7 helix is not conserved and may play a structural role in *Legionella pneumophila* and few other Legionella species.

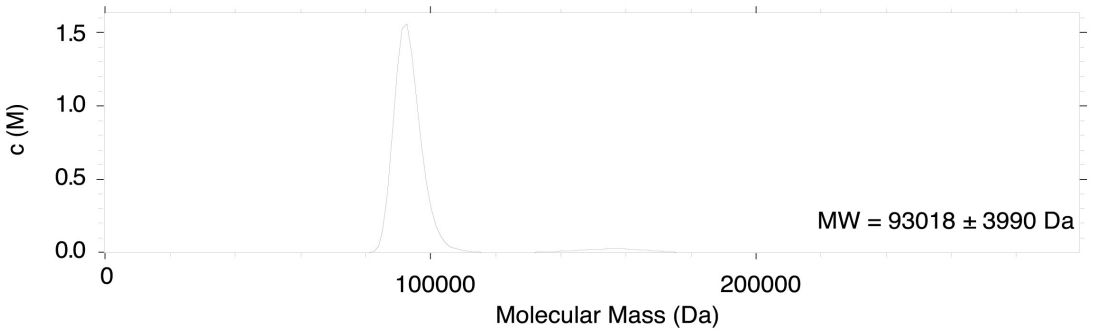
A



B


Absorbance (280nm)
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
6.0
6.1
6.3
6.5
6.6
6.7
6.8
6.9
7.0
Radius (mm)

C


c (M)
1.5
1.0
0.5
0.0
MW = 93018 ± 3990 Da
0
100000
200000
Molecular Mass (Da)

D

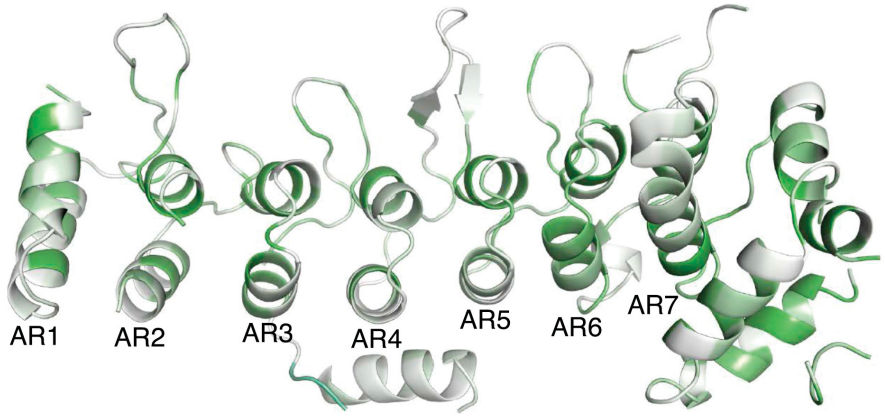

AR1
AR2
AR3
AR4
AR5
AR6
AR7

Figure 3.3 Structure of Ank domain from AnkC. (*A*) The Ank domain makes an apparent dimer in the crystal mediated by the outer surface of ankyrin loops. (*B*) A representative experimental sedimentation profile of 1.0 mg mL^{-1} purified AnkC at a rotor speed of 35,000 rpm. Shown is the absorbance of the sample at 280nm plotted against the radial position with the lines representing the boundaries of each scan. (*C*) The continuous molecular weight distribution, c(M), corresponding to the velocity experiment is plotted. The experimental molecular weight was calculated to be 93 ± 4 kDa with a frictional coefficient (f/fo) of 1.46, revealing the AnkC dimer. (*D*) Sequence conservation derived from AnkC proteins of ~50 *Legionella* species and mapped to the structure. Highly conserved residues are shown in green.

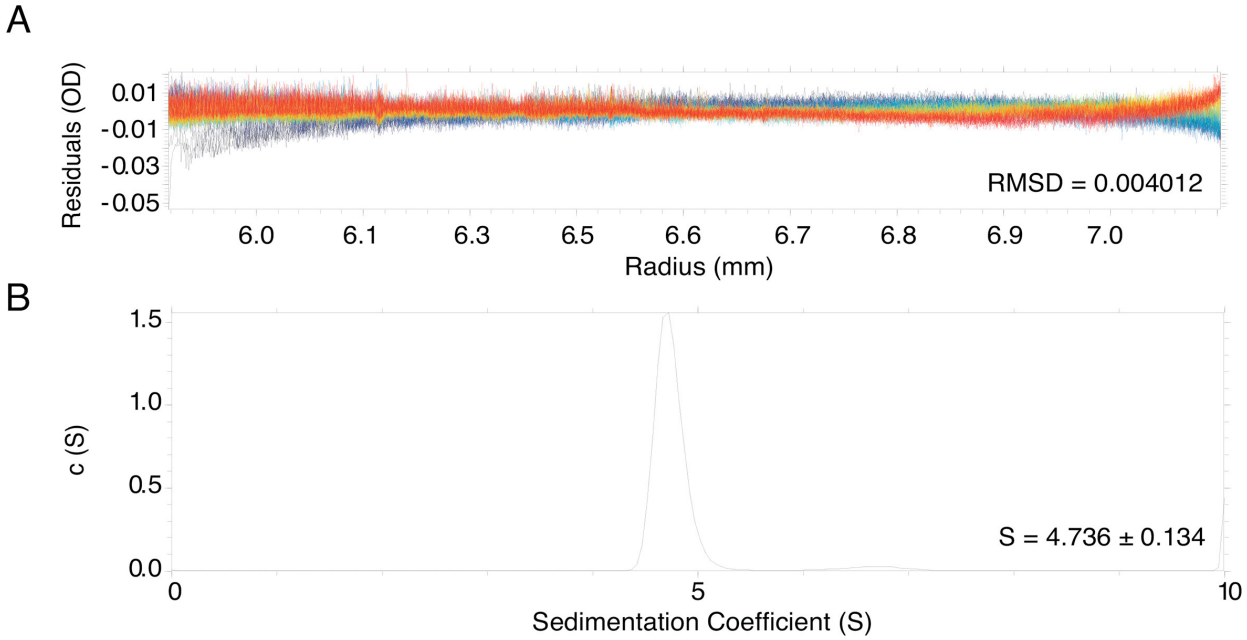


Figure 3.4 Sedimentation velocity analysis of AnkC. (*A*) The residuals of the model fitting for the data shown as an overlay representation (RMSD = 0.004012). (*B*) The continuous sedimentation distribution, c(S), corresponding to the velocity experiment was plotted. An experimental sedimentation coefficient of 4.736 ± 0.134 S was calculated using SEDFIT.

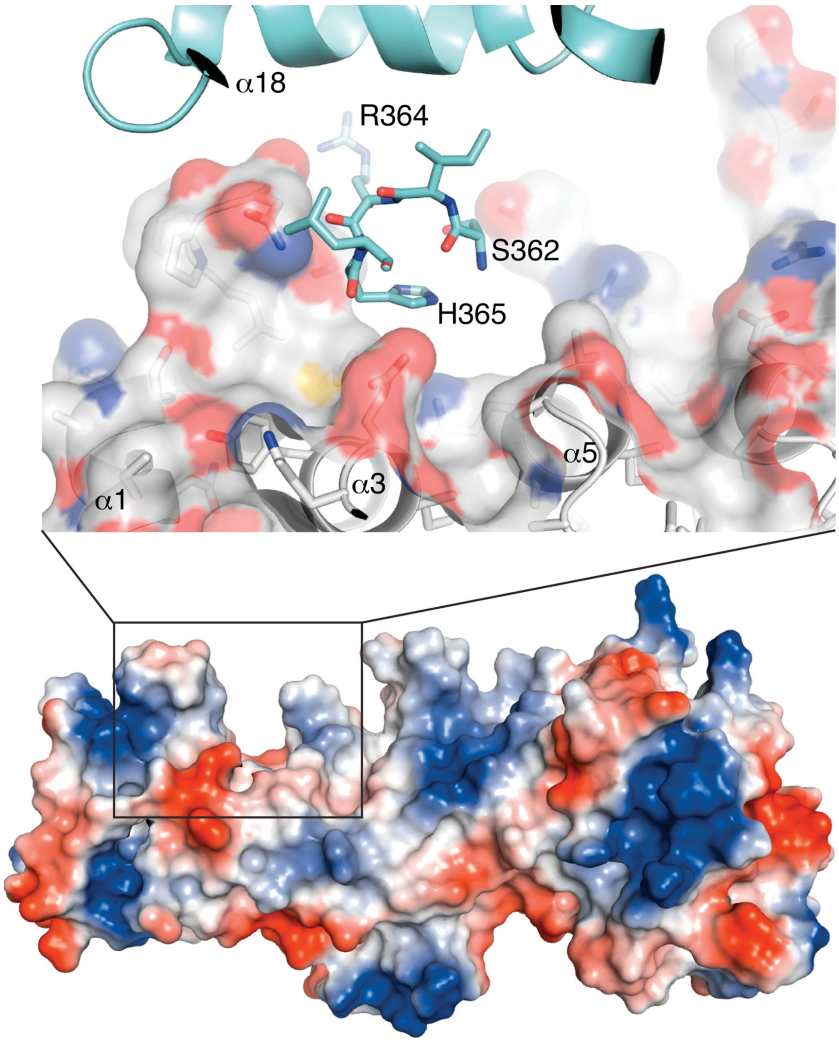


Figure 3.5 Ankyrin groove interactions of AnkC. Surface charge distribution of the Ank domain (blue = positive, red = negative) with an enlarged view of the C-terminal region of the crystallized fragment interacting with the ankyrin groove of another AnkC molecule in the crystal.

3.3.4 Structural comparison to other proteins. A structural similarity search using AnkC against the DALI database identified a number of Ank proteins (Holm and Rosenström, 2010). A common feature of these proteins is the presence of at least six ankyrin repeats (**Fig. 3.6**). The most structurally similar protein is B-cell lymphoma 3-encoded protein (BCL-3) (PDB code 1K1B, Z-score 17.5) with an RMSD of 2.7 Å for 215 Cα atoms. Among the other top hits are a number of NOTCH 1 structures with a Z-score of 17.3 (PDB code 2QC9), nuclear factor NF-κB P100

subunit (PDB code 4OT9, Z-score 17.1), and 26S proteasome regulatory subunit 10 (PDB code 3AJI, Z-score 16.7). AnkC has 17% sequence identity to BCL-3 and 20-22% identity to the other mentioned proteins. It is important to note that the revealed structural similarity is limited strictly to the ankyrin repeats. The C-capping helices did not produce significant hits in a structural similarity search. Within the Ank domain, AnkC shows distinct structural features that are not observed in other ankyrin domains. The most striking difference is the novel dimerization interface along the back of the ankyrin groove, allowing dimerization to occur without interfering with the ability of the ankyrin groove to bind potential partners. In addition, an extra helix α7 is inserted after AR3. This helix is positioned opposite to the ankyrin groove and should not interfere with the typical mode of protein binding to AnkC. It is plausible that this helix plays a structural role in stabilizing the protein fold. The AnkC structure is also missing the L-shaped hairpin loop in AR3. This interrupts the ankyrin groove and makes the fold less regular. The surface charge is also similar to BCL-3 and NOTCH 1 proteins (**Fig. 3.6**). It would be interesting to test if AnkC could interfere with cellular pathways involving BCL-3 or NOTCH 1.

AnkC is an ankyrin repeats-containing effector protein conserved in multiple *Legionella* species. The crystal structure unexpectedly displays seven ankyrin repeats that are generally similar to a typical ankyrin fold but contains unique structural features. AnkC is able to dimerize and sequence conservation analysis suggests that the ankyrin groove is a functional site that interacts with yet unidentified binding targets. Finding these binding partners and characterizing the C-terminal domain will clarify the function of AnkC and its role in the pathogenesis of Legionnaires' disease. The Ank domain structure is an important step towards that goal.

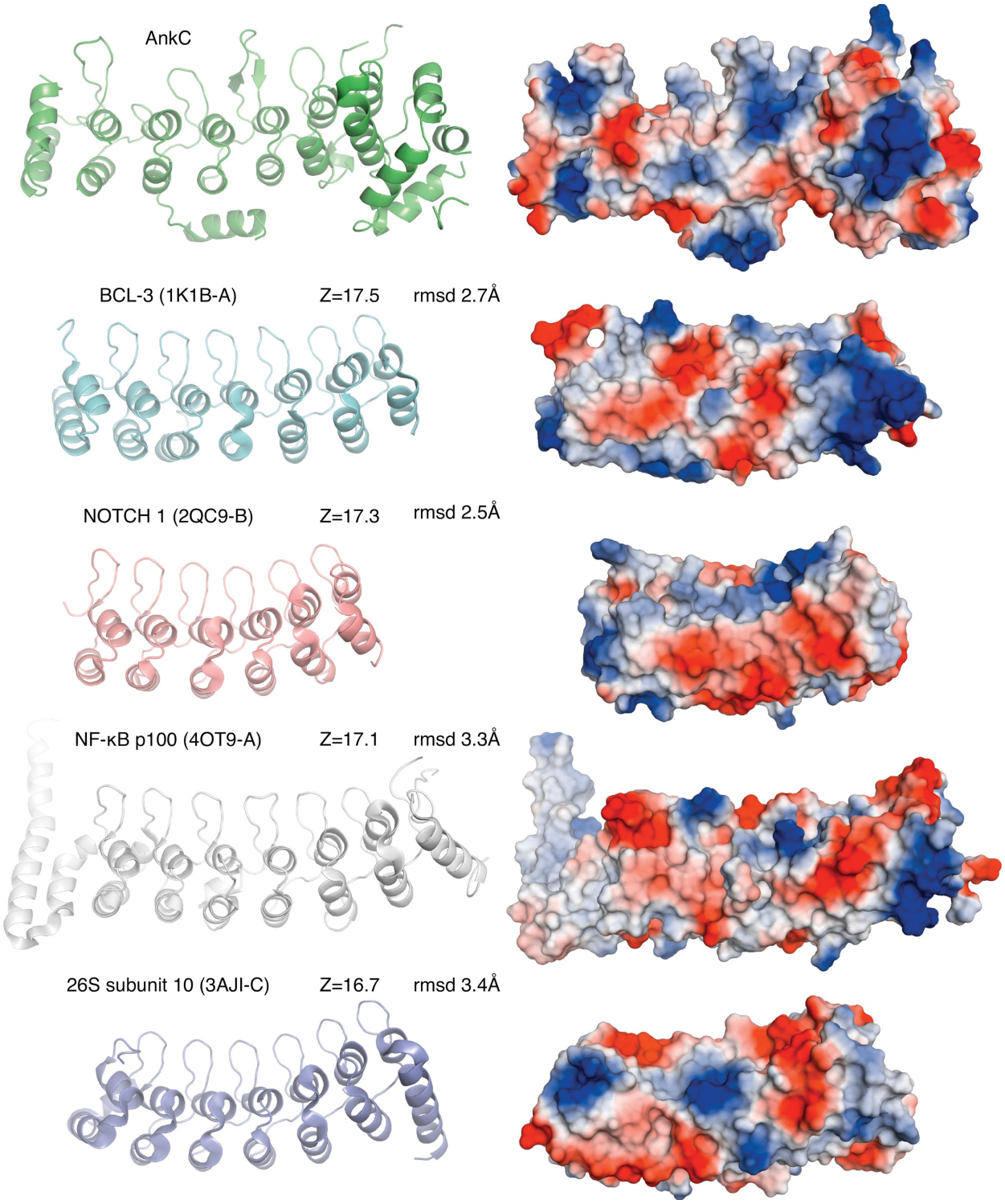


Figure 3.6 Structural similarity search of AnkC. Structural similarity of AnkC to Ank domains of BCL-3 (PDB code 1K1B, chain A), NOTCH 1 (PDB code 2QC9, chain B), nuclear factor NF-κB p100 subunit (PDB code 4OT9, chain A), and 26S proteasome non-ATPase regulatory subunit 10 (PDB code 3AJI, chain C).

3.4 Experimental procedures

3.4.1 Protein expression and purification. *Legionella pneumophila* AnkC (lpg0483, LegA12) fragment 2-384 was inserted into pMCSG7 vector as an N-terminal His-tagged fusion. The plasmid was transformed into BL21(DE3) and plated on LB-agar, with ampicillin (100 mg/L) for selection. A single colony was inoculated in 20 mL LB medium and incubated in a 37°C shaker overnight. The overnight culture was then inoculated into 1 L LB medium and grown at 37°C. When OD_{600} reached 0.8, the cell culture was induced with 1 mM IPTG and transferred into a 18 °C shaker for 18 h. For production of selenomethionine-labeled protein, the expression plasmid was transformed into the *E. coli* methionine auxotroph strain DL41 (DE3) and the protein was produced using LeMaster medium. After expression, the cell culture was spun down at 7000×g for 20 min and re-suspended in 50 mL lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol and 5 mM imidazole). Cells were lysed by sonication and centrifuged at 30,000×g for 45 min. The supernatant was loaded onto nickel-nitrilotriacetic acid (Ni-NTA) agarose resin (Qiagen, Valencia, CA, USA), which was pre-equilibrated with lysis buffer. The protein-resin mixture was incubated at 4 ºC for 30 min. After binding, the resin was washed with 5 column volumes of lysis buffer and then with 5 column volumes of lysis buffer containing 30 mM imidazole. Subsequently, the His-tagged protein was eluted with 500 mM imidazole, concentrated and loaded onto a size exclusion column (Superdex 200 10/300 GL column, GE), equilibrated with 10 mM HEPES pH 7.0, 100 mM NaCl, 2 mM DTT.

3.4.2 Crystallization. AnkC constructs were purified and concentrated to approximately 15-20 mg/mL. Crystallization screens were performed in 24-well plates in a hanging drop format using commercial Qiagen screens. Promising conditions were further explored by systematic modifications of the initial conditions within a narrow range combined with using an additive

screen. The best native crystals for AnkC (2-384) were obtained at 20 ºC by equilibrating a 0.8 µL drop of protein at 15 mg/mL in 10 mM HEPES pH 7.0, 0.1 M NaCl, 2 mM DTT, with 0.8 µL of reservoir solution containing 0.25 M magnesium formate, 5 mM EDTA, and 0.1 M sodium acetate (pH 5.0), suspended over 1 mL of reservoir solution. The best selenomethionine-labeled crystals for AnkC (2-382) were obtained at 20 ºC by equilibrating a 0.8 µL drop of protein concentrated to 10 mg/mL in 10 mM HEPES pH 7.0, 0.1 M NaCl, 2 mM DTT, with 0.8 µL of reservoir solution containing 0.22 M magnesium formate, 5 mM EDTA, and 0.1 M Bis-Tris (pH 4.5), suspended over 1 mL of reservoir solution. For cryoprotection, crystals were transferred into its crystallization condition containing 35% (w/v) glycerol. For data collection, crystals were picked up in a nylon loop and flash cooled in a N_2 cold stream (Oxford Cryosystem).

3.4.3 Structure determination and refinement. The selenium SAD dataset was collected on an Eiger detector at beamline 23ID-B at the Advanced Photon Source (APS). The native dataset from a AnkC crystal was collected using a single-wavelength (0.63 Å) regime on an ADSC Quantum-210 CCD detector (Area Detector Systems Corp.) at beamline A1 at the Cornell High-Energy Synchrotron Source (CHESS) (**Table 3.1**). Data processing and scaling were performed with HKL2000 (Otwinowski Z, 1997). The anomalously scattering selenium substructure was determined using the program PRASA (Skubak, in preparation). Crank2 (Skubak and Pannu, 2013) was used for the subsequent SAD phasing and model building. Crank2 used Refmac5 (Murshudov et al., 2011) for the reciprocal space phasing, phase combination and refinement, Parrot (Cowtan, 2010) for crystal space density modification and Buccaneer (Cowtan, 2006) for model building. The resultant model was used for phasing a native dataset. The model was further extended manually using the program Coot (Emsley et al., 2010) and was improved by multiple

cycles of refinement using the program PHENIX (Adams et al., 2011). Coordinates have been deposited in the RCSB Protein Data Bank with accession code 5VRQ.

Table 3.1. Data collection and refinement statistics for AnkC

	SeMet 2-382	Native 2-384 PDB code 5VRQ
Data collection		
Space group	$P4_32_12$	$P4_32_12$
a, *b*, *c* (Å)	78.29, 78.29, 160.68	78.71, 78.71, 156.32
Resolution (Å)	50-3.80 (3.87-3.80)[1]	50-3.20 (3.26-3.20)
R_{sym}	0.179 (0.735)	0.118 (0.558)
$I / \sigma I$	30.1 (5.3)	51.9 (7.1)
Completeness (%)	99.8 (100)	99.9 (99.9)
Redundancy	16.7 (17.2)	18.7 (19.4)
CC1/2	0.92	0.96
Refinement		
Resolution (Å)		13.7-3.20
No. reflections		8451
R_{work} / R_{free}		0.223/0.280
No. atoms		2350
Protein		2346
Water		4
B-factors		
Protein		91.3
Water		53.1
R.m.s deviations		
Bond lengths (Å)		0.002
Bond angles (°)		0.51
Ramachandran statistics (%)		
Most favored regions		93.8
Additional allowed regions		6.2

[1]Highest resolution shell is shown in parentheses.

3.4.4 Analytical ultracentrifugation. Sedimentation equilibrium analytical ultracentrifugation (AUC) experiments were performed at 20ºC using a Beckman Coulter XL-I analytical ultracentrifuge. The samples were spun with an angular velocity of 5000 rpm until concentration gradients reached equilibrium (~24 h). The protein concentration gradients were monitored by UV at 290 nm. Data was analyzed using Sedfit v1501b (Kipreos and Pagano, 2000)

with input parameters: Vbar = 0.742651; Buffer Density = 1.0031; Buffer Viscosity = 0.01002 (10mM HEPES, 100mM NaCl, 2mM TCEP, pH 7.0).

3.5 Acknowledgements

The authors declare no conflict of interests. We thank Dr. Miroslaw Cygler (University of Saskatchewan) for the AnkC plasmid. This work was funded by a CIHR Genomics grant GSP-48370 to Miroslaw Cygler and Kalle Gehring. Data were acquired at the Advanced Photon Source (APS) and at the Macromolecular Diffraction (MacCHESS) facility at the Cornell High Energy Synchrotron Source (CHESS). CHESS is supported by the NSF & NIH/NIGMS via NSF award DMR-0225180, and the MacCHESS resource is supported by NIH/NCRR award RR-01646. The authors declare no conflict of interest.

Chapter 4: An orphan SidE-related member

4.0 Connecting text

Not only has *Legionella* mimicked eukaryotic-like domains to trick the host into using bacterial proteins, as in AnkB, the bacterium has also developed novel proteins and domains capable of interacting with ubiquitin without a host intermediate. A novel ubiquitin ligase family (SidE family) capable of ubiquitinating Rab33b without ATP, Mg^{2+}, E1 or E2 was identified (Qiu et al., 2016). Instead, only NAD was required for its ligase activity. This effector family generates ADP-ribosylated ubiquitin (ADP-Rib-Ub) for cleavage by its novel phosphodiesterase (PDE) domain to attach phospho-ribsoylated ubiquitin (P-Rib-Ub) onto a substrate. Sequence similarity searches indicate that this SidE-PDE domain not only exists in the SidE family, but also in the uncharacterized effector, lpg1496 (Wong et al., 2015).

4.1 Summary

Pathogenic gram-negative bacteria use specialized secretion systems that translocate bacterial proteins, termed effectors, directly into host cells where they interact with host proteins and biochemical processes for the benefit of the pathogen. Lpg1496 is a previously uncharacterized effector of *Legionella pneumophila*, the causative agent of Legionnaires' disease. Here, we crystallized three nucleotide-binding domains from lpg1496. The C-terminal domain, which is conserved among the SidE family of effectors, is formed of two largely α-helical lobes with a nucleotide-binding cleft. A structural homology search has shown similarity to phosphodiesterases involved in cleavage of cyclic nucleotides. We have also crystallized a novel domain that occurs twice in the N-terminal half of the protein that we term KLAMP domains due to the presence of homologous domains in bacterial histidine kinase-like ATP-binding region-

containing proteins and S-adenosylmethionine-dependent methyltransferase proteins. Both KLAMP structures are very similar but selectively bind 3′,5′-cAMP and ADP. A co-crystal of the KLAMP1 domain with 3′,5′-cAMP reveals the contribution of Tyr61 and Tyr69 that produce π-stacking interactions with the adenine ring of the nucleotide. Our study provides the first structural insights into two novel nucleotide-binding domains.

4.2 Introduction

Legionella pneumophila is a gram-negative bacterium and is the causative agent of Legionnaires' disease, an acute form of pneumonia (Fraser et al., 1977). Pathogenic gram-negative bacteria use specialized secretion systems that translocate bacterial proteins, termed effectors, directly into host cells where they interact with host proteins and hijack eukaryotic biochemical processes for the benefit of the pathogen. These secretion machineries are highly conserved among different bacterial species. *L. pneumophila* uses a Dot/Icm type IV (T4SS) secretion system to inject effector proteins into the host cells (Vogel et al., 1998). The secreted effectors allow the bacterium to escape the host lysosomal pathway after phagocytosis.

Currently, about 300 Dot/Icm dependent effectors of *L. pneumophila* have been identified, using methods such as interaction with Dot/Icm components (Bardill et al., 2005; Ninio et al., 2005), the presence of a C-terminal secretion signal (Nagai et al., 2005), and a machine-learning approach where effectors were identified based on shared features (Burstein et al., 2009). However, the functions of most remain unknown. Lpg1496 is one such experimentally validated effector protein (Burstein et al., 2009).

From sequence alignment, lpg1496 was found to contain a conserved domain from the SidE family. An interbacterial transfer assay in 2004 led to the identification of this family, which

includes SdeA, SdeB, SdeC and SidE (Luo and Isberg, 2004). These four members were grouped based on their location on the chromosome, their interactions with IcmS, a putative chaperone for effectors, and sequence similarity (Bardill et al., 2005; Coers et al., 2000). SidE proteins are secreted during infection and localize to the poles of the bacterium, where they may interact with nearby Dot/Icm substrates such as LidA (Bardill et al., 2005; Conover et al., 2003). A member of this family, SdeA, is a paralog of LaiA, which contains homology to an integrin analogue gene of *S. cerevisiae* (Chang et al., 2005; Hostetter et al., 1995). LaiA has been shown to be required for adherence and entry into alveolar epithelial cells (Chang et al., 2005). In addition, SidE family secretion peaks 30 minutes post infection of mouse bone marrow macrophages (BMM), and SdeA transfected cells fragment the Golgi apparatus, suggesting a role in the early events of intracellular growth, such as modification of the *Legionella*-containing vacuole (LCV) (Bardill et al., 2005; Jeong et al., 2015). A deletion strain of this family results in an approximate 100-fold less growth than wild-type in *A. castellanii*, which can be complemented by the expression of SdeA on a plasmid (Bardill et al., 2005; Jeong et al., 2015). Recently, it has been reported that over-expression of SdeA in a SidJ mutant completely inhibits the growth of intracellular *Legionella*, to levels similar to a translocation deficient DotA mutant (Jeong et al., 2015).

No molecular characterization of lpg1496 has been done previously. Here, we identified three independently folded domains in the protein and determined their high-resolution crystal structures. The two N-terminal domains are a novel fold that we call KLAMP domains. The C-terminal domain is found in all SidE family proteins but had not been characterized structurally. We showed that all the domains bind to nucleotides and revealed molecular determinants of their binding specificity.

4.3 Results

4.3.1 The PDE domain. Sequence analysis of the C-terminus of lpg1496 reveals sequence similarity to the N-terminal region of the original members of the SidE family (SdeA, SdeB, SdeC and SidE) (**Fig. 4.1A & B**). This domain is also found in other strains of *Legionella pneumophila* such as *Shakespearei* and *Longbeachae* (**Fig. 4.1C**). Some proteins containing the PDE domain contained the Vip2 domain, which is present in an actin-ADP-ribosylating toxin family. These proteins confer virulence by modifying monomeric actin, resulting in depolymerization of the actin cytoskeleton, and eventually leading to cell death (Han et al., 1999).

The C-terminal domain of lpg1496 (293-580) was crystallized, yielding a high-resolution structure of the PDE domain. This construct diffracted to 1.6 Å resolution with one PDE domain in the asymmetric unit from Asp298 to Lys580. A longer 293-598 construct was also crystallized and refined to 2.34 Å (**Table 4.1**). Phasing was performed using the single-wavelength anomalous dispersion (SAD) method with seleno-methionine labeled crystals. The structure contains two PDE domains in the asymmetric unit with interpretable density for Asp298-Leu592 and Asp298-Ala589, respectively, with an RMSD of 0.24 Å over 264 Cα atoms. The structures of both constructs are very similar, yielding an RMSD of 0.58 Å over 245 Cα atoms.

The general fold of the PDE domain consists of two mostly α-helical lobes with a cleft in-between. The larger lobe comprises 11 α-helices, 1 α-helical turn and 2 antiparallel β-strands. The smaller lobe contains 3 α-helices (**Fig. 4.2A**).

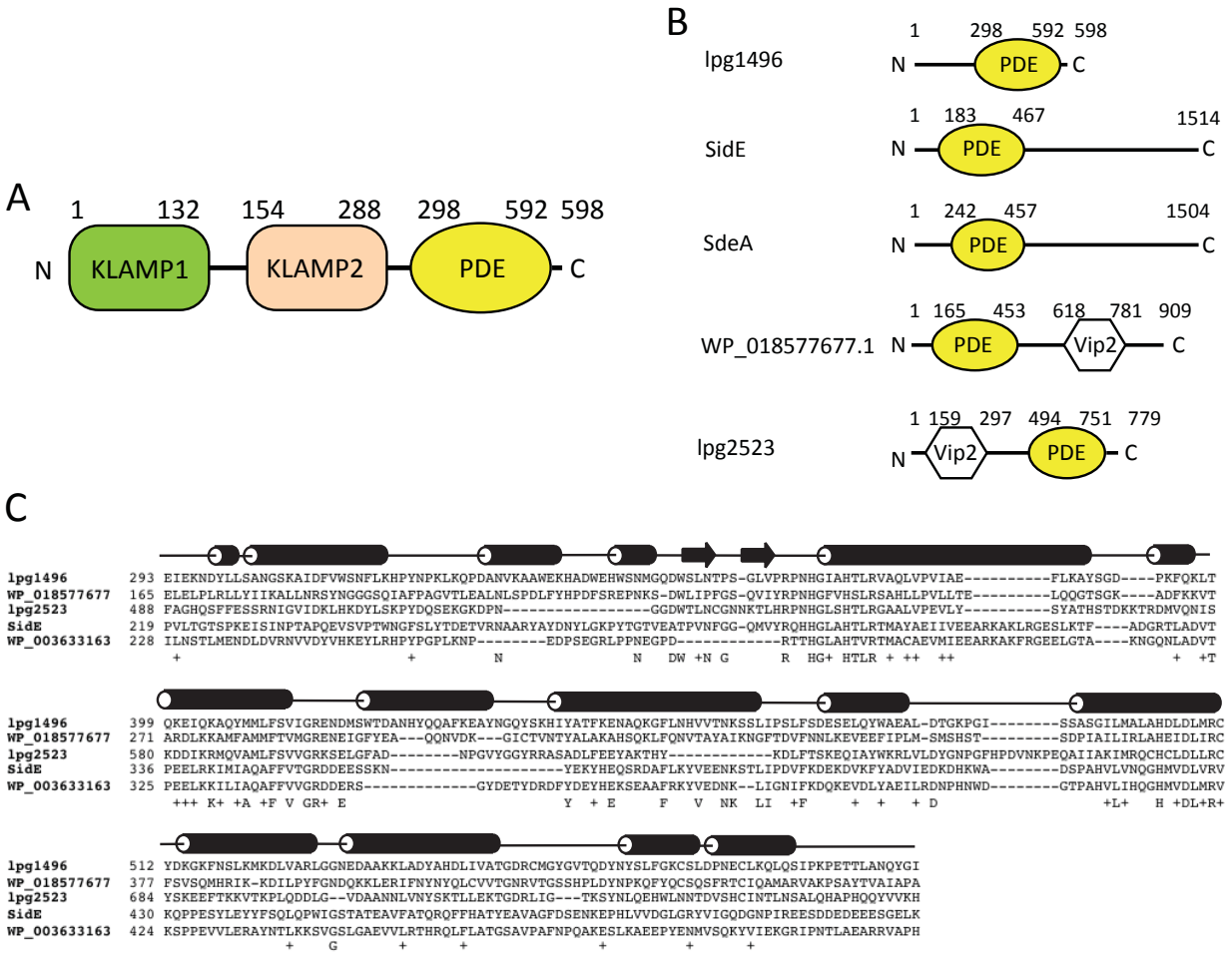


Figure 4.1 PDE sequence and structure. (*A*) Domain architecture of lpg1496 (KLAMP1 in green, KLAMP2 in wheat, PDE in yellow). (*B*) The occurrence of the PDE domain in *Legionella* proteins. (*C*) Sequence alignment of the PDE domain of lpg1496 with other bacterial proteins. The secondary structure of the lpg1496 PDE domain is overlaid on top (cylinder = α-helix; arrow = β-sheet).

Table 4.1 Data collection and refinement statistics for the PDE domain

	293-580 PDB code 5BU1	293-598 PDB code 5BU0	154-598/ADP PDB code 5BU2
Data collection			
Space group	$P2_12_12_1$	$P2_12_12_1$	P1
Cell dimensions□ □			
a, *b*, *c* (Å)	60.89, 71.47, 77.00	71.98, 77.23, 109.38	56.87, 69.71, 77.03
α, β, γ (°)	90, 90, 90	90, 90, 90	89.87, 72.30, 70.31
Resolution (Å)	50-1.60 (1.63-1.60)[1]	50-2.35 (2.39-2.35)	50-2.10 (2.14-2.10)
R_{sym}	0.093 (0.465)	0.169 (0.446)	0.098 (0.435)
$I / \sigma I$	34.2 (8.5)	42.5 (11.4)	11.3 (1.7)
Completeness (%)	99.9 (99.9)	99.8 (100)	95.0 (78.3)
Redundancy	14.4 (14.2)	13.0 (12.7)	2.6 (2.5)
Refinement			
Resolution (Å)	29.6-1.60	37.9-2.35	72.9-2.11
No. reflections	42595	25991	54209
R_{work} / R_{free}	0.161/0.191	0.239/0.293	0.229/0.273
No. atoms			
Protein	2306	4653	9242
Water	414	220	121
Nucleotide			141
B-factors			
Protein	15.32	27.23	37.57
Water	30.84	28.18	35.17
Nucleotide			44.66
R.m.s deviations			
Bond lengths (Å)	0.017	0.002	0.007
Bond angles (°)	2.019	0.651	1.212
Ramachandran statistics (%)			
Most favored regions	96.3	96.2	96.0
Additional allowed regions	3.7	3.8	4.0

[1]Highest resolution shell is shown in parentheses.

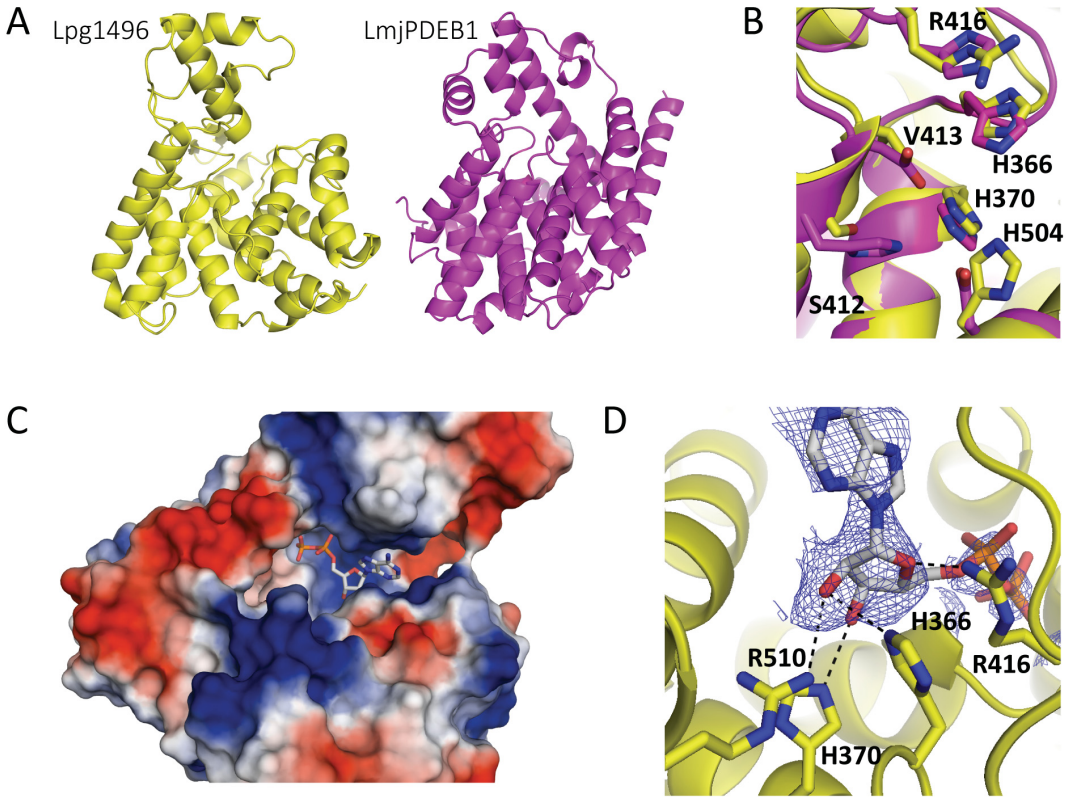


Figure 4.2 PDE ADP binding. (*A*) Structural similarity of the PDE domain of lpg1496 (yellow) and LmjPDEB1 (PDB code 2R8Q, purple). (*B*) Superposition of the active site of LmjPDEB1 (purple) and lpg1496 (yellow), with important catalytic residues labeled for lpg1496. (*C*) Surface charge representation of the PDE domain (blue = positive, red = negative) with the bound ADP (grey). (*D*) The PDE domain forms the most consistent polar contacts with the ribose ring of ADP. Two conserved histidines with LmjPDEB1 and human PDEs, are involved in hydrogen bonds (His370 and His366).

4.3.2 Putative substrate binding site. A structural homology search using the PDE domain of lpg1496 against the DALI database (Holm and Rosenström, 2010) led to the identification of LmjPDEB1, with a Z-score of 9.2 and RMSD of 3.3 Å. LmjPDEB1 is a cAMP specific cyclic nucleotide phosphodiesterase (PDE) found in Leishmania (PDB code 2R8Q) (**Fig. 4.2A**) (Wang et al., 2007a). It requires the binding of two divalent metal ions (zinc and magnesium) for activity. Half of the residues involved in metal- coordination are conserved between LmjPDEB1 (His685, His721, Asp722, Asp835) and lpg1496 (His370, Ser412, Val413, His504).

The presence of serine and valine instead of histidine and aspartate explains why there are no metal ions in the lpg1496 structure (**Fig. 4.2B**). Of the two histidines involved in catalysis in LmjPDEB1, one is conserved in lpg1496 (His366), while the other is substituted with arginine (Arg416). These two conserved histidines, His370 for metal coordination and His366 for catalysis, are also conserved among the human PDEs (Wang et al., 2007a).

Overall, the PDE domain is structurally a member of a superfamily of metal-dependent phosphohydrolases, named HD for the presence of a conserved histidine and aspartate involved in coordination of divalent cations. HD domains have demonstrated phosphatase, nucleotide triphosphatase, phosphodiesterase and ribonuclease activities (Aravind and Koonin, 1998). In the case of lpg1496, His504 and Asp505 are part of the potential active site that was identified by alignment with LmjPDEB1 (Asp835 and Val836). To test for putative phosphatase and phosphodiesterase activity, malachite green assays were performed against 3'-AMP, 5'-AMP, ADP, ADP ribose, 2',3'-cAMP, 3',5'-cAMP and 3',5'-cGMP. Lpg1496 did not show significant activity against any nucleotide (data not shown). This may be explained by the lack of metal binding, in addition to other differences in active sites between lpg1496 and known PDEs.

4.3.3 The PDE domain of lpg1496 binds nucleotides. In order to determine whether the C-terminal PDE domain nonetheless binds nucleotides, we set up crystallization screens for lpg1496 (154-598), in the presence of ADP or 3',5'-cAMP. Crystallization trials with ADP produced crystals in the P1 space group that diffracted to 2.1 Å with interpretable density for the conserved C-terminal PDE domain (**Table 4.1**). No density was observed for residues 154 to 297. There were four molecules in the asymmetric unit, all of which contained bound nucleotide. Three molecules showed the complete ADP nucleotide, while one only had interpretable density for the ribose moiety (**Fig. 4.2C**). Although lpg1496 does not have phosphatase activity, its PDE domain

is capable of binding ADP. However, the binding was not detected by isothermal titration calorimetry measurements.

An overlay of the four chains shows that this domain forms the most consistent polar contacts with the ribose ring. Arg510 and His366 form hydrogen bonds with O2', His370 with O3', and Arg416 with O4' (**Fig. 4.2D**). The phosphates of ADP are generally stabilized by interactions with the main chain amides. The ADP-bound C-terminal domain structure shows that the putative substrate binding site of lpg1496 is in fact shifted closer towards the metal binding sites of previously characterized PDEs, adding to the possibility that lpg1496 is involved in ADP-ribosylation rather than in the cleavage of phosphoester bonds.

4.3.4 Lpg1496 contains two homologous domains in the N-terminal half. To further understand lpg1496, we looked into the previously uncharacterized ~300-residue N-terminal half of the protein (**Fig. 4.1A**). Sequence similarity searches identified two repeats of approximately 120 residues (residues 8-116 and 162-277) displaying 32% sequence identity (**Fig. 4.3A**). Limited proteolysis experiments showed that these regions form independently folded domains. Trypsin digestion of the C-terminal fragment of lpg1496 containing residues 154 to 598 generated two stable fragments of approximately 16 kDa and 35 kDa (data not shown). The fragments could be separated by size-exclusion chromatography. The 35 kDa fragment corresponds to the PDE domain, while the 16 kDa fragment was later characterized as the middle domain of lpg1496.

A BLAST sequence similarity search using the N-terminal sequence of lpg1496 found a number of other proteins, notably a previously uncharacterized region of histidine kinase-like ATP-binding region-containing bacterial proteins and the C-terminal domain of bacterial S-adenosylmethionine-dependent methyltransferases (AdoMet MTases) (**Fig. 4.3A**). We propose to term the domain family as KLAMP domains for their presence in histidine kinase-like ATP-

binding region-containing proteins and AdoMet MTase proteins. Thus, lpg1496 contains two KLAMP domains (KLAMP1 and KLAMP2) and a C-terminal PDE domain (**Fig. 4.1A**).

Crystallization trials with the second KLAMP domain of lpg1496 produced crystals in the $P2_1$ space group that diffracted to 1.15 Å. The structure was solved using SAD (**Table 4.2**). The structure contains one KLAMP molecule with density from Asp154 to Thr288. Crystals were also obtained in the $P2_12_12_1$ space group with nearly identical structure (RMSD of 0.2 Å over 117 Cα atoms), confirming that the observed structures were not influenced by crystal contacts.

We also determined the structure of the first KLAMP domain. Lpg1496 (residues 1-138) yielded crystals that diffracted to 1.2 Å using synchrotron radiation (**Table 4.2**). The asymmetric unit contains two molecules consisting of residues Met1-Val132 and Met1-Pro131, respectively. The structures of KLAMP1 and KLAMP2 domains are very similar, displaying RMSD of 1.2 Å over 90 Cα atoms (**Fig. 4.3C**).

The structure of the KLAMP domain consists of two anti-parallel α-helices (α1-α2) flanked by a five-stranded anti-parallel β-sheet (β9-β1-β2-β4-β3) on one side and a four-stranded anti-parallel β-sheet on the other side (β5-β7-β6-β8) (**Fig. 4.3B**). This arrangement leaves the opposite sides of both β-sheets solvent-exposed. The biggest difference between the two KLAMP structures is the conformation of two loops between α1 and β3 and between β7 and β8, which are longer in KLAMP2. Three cis-proline residues are found in KLAMP2: Pro202 in the loop between α1 and β3, Pro208 in the loop between β3 and β4 and Pro219 in the loop between β4 and β5. This is an interesting feature of the structure, as cis-proline residues are not very common in protein structures. In KLAMP1, Pro62 and Pro79 are conserved and correspond to Pro202 and Pro219, respectively, but only Pro62 is in cis-conformation. The conformation of the β4-β5 loop is different between the two domains, which can be partly attributed to the trans-conformation of Pro79.

Structural similarity search with the program DALI (Holm and Rosenström, 2010) using the middle domain of lpg1496 shows that the KLAMP fold is relatively novel. The highest similarity hit was to a mixed α-β protein, pterin dehydratase-like protein (PDB code 4LOW) with a low Z-score of 4.7.

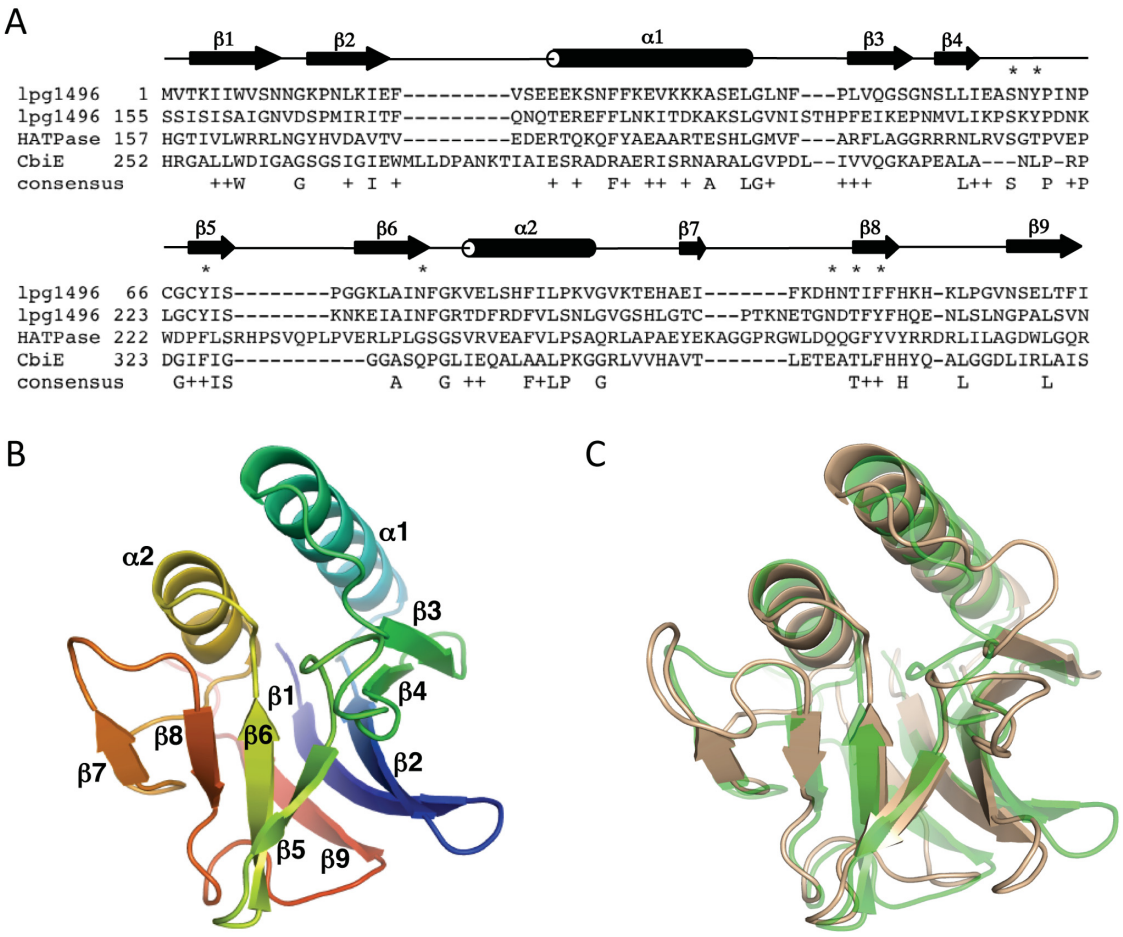


Figure 4.3 KLAMP sequence and structure. (*A*) Sequence alignment of lpg1496 domains with predicted ATPase from *Streptomyces pratensis* (WP_014153402) and precorrin-6y methyltransferase CbiE from *Beijerinckia indica* (WP_012386418). Secondary structure is shown according to lpg1496 domain structures. Residues involved in binding of 3′,5′-cAMP by KLAMP1 of lpg1496 are marked with asterisks. (*B*) Cartoon representation of the KLAMP1 domain structure, color-coded from the N-terminus (blue) to C-terminus (red). (*C*) Overlay of structures of KLAMP2 (wheat) and KLAMP1 (faint green) shows high similarity of secondary structure regions with differences in loop conformations.

Table 4.2 Data collection and refinement statistics for KLAMP1 and KLAMP2

	KLAMP1 PDB code 5BTW	KLAMP1/3'5'-cAMP PDB code 5BTX	KLAMP2 PDB code 5BTY	KLAMP2 PDB code 5BTZ
Data collection				
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_12_12_1$
Cell dimensions□ □				
a, *b*, *c* (Å)	32.41, 56.92, 60.18	32.62, 57.45, 63.78	41.80, 35.91, 43.66	36.14, 44.07, 77.53
α, β, γ (°)	90, 91.18, 90	90, 93.32, 90	90, 104.7, 90	90, 90, 90
Resolution (Å)	50-1.20 (1.22-1.20)[1]	50-1.75 (1.78-1.75)	50-1.15	50-1.60 (1.63-1.60)
R_{sym}	0.056 (0.567)	0.076 (0.497)	0.068 (0.437)	0.057 (0.406)
$I / \sigma I$	43.9 (2.8)	39.0 (5.3)	43.2 (3.1)	49.7 (4.9)
Completeness (%)	92.7 (84.3)	89.2 (92.3)	92.0 (57.9)	99.9 (100.0)
Redundancy	7.7 (6.0)	6.6 (6.4)	7.4 (5.3)	8.6 (8.7)
Refinement				
Resolution (Å)	60.17 - 1.20	63.67 – 2.10	40.5 - 1.15	38.8 - 1.60
No. reflections	60366	12544	39061	16100
R_{work} / R_{free}	0.158/0.180	0.238/0.292	0.169/0.192	0.199/0.231
No. atoms				
Protein	2138	2087	1071	1043
Water	207	107	200	130
Nucleotide		44		
B-factors				
Protein	14.0	21.3	11.5	13.0
Water	21.0	34.8	22.8	33.4
Nucleotide		38.6		
R.m.s deviations				
Bond lengths (Å)	0.012	0.011	0.005	0.006
Bond angles (°)	1.46	1.64	1.12	1.03
Ramachandran statistics (%)				
Most favored regions	95.1	94.4	99.2	97.6
Additional allowed regions	4.9	5.6	0.8	2.4

[1]Highest resolution shell is shown in parentheses.

4.3.5 KLAMP domains of lpg1496 bind nucleotides. The presence of KLAMP-like sequences in proteins predicted to bind ATP suggested that we test the KLAMP2 domain for binding to nucleotides. We used ^{15}N-labeled middle domain of lpg1496 and obtained its ^{1}H-^{15}N correlation spectrum. The spectrum showed well-dispersed signals for backbone amides,

characteristic of a well folded protein. NMR titrations were performed by a stepwise addition of potential nucleotide ligands monitored by HSQC experiments (**Fig. 4.4A**).

Titration of the KLAMP2 domain of lpg1496 with ATP resulted in chemical shift changes of roughly 20 backbone amides, indicating specific binding to the domain. The affinity of the interaction can be estimated by a fit of chemical shift changes of NMR signals versus ligand concentration. Using several signals with biggest chemical shift changes, the dissociation constant (K_d) of ATP binding was estimated to be 800±150 μM. Even larger chemical shift changes were observed upon addition of ADP with a K_d of 109±7 μM (**Fig. 4.4A & B**), while titration with AMP produced much smaller spectral changes with a K_d of 1500±200 μM. There is a preference for a diphosphate group in the nucleotide-binding site of KLAMP2. In order to test the base specificity, we titrated the ^{15}N-labeled middle domain of lpg1496 with GDP. Addition of GDP did not result in spectral changes (data not shown) indicating a clear preference of KLAMP2 for binding adenine nucleotides.

NMR binding studies can identify the ligand-binding site on a protein via a residue-specific assignment of NMR signals and mapping the binding-induced spectral changes on the three-dimensional structure. We prepared ^{13}C,^{15}N-labeled KLAMP2 and assigned the backbone amides using standard heteronuclear NMR experiments. The residues showing the biggest chemical shift changes upon ADP addition are Leu223 (0.80), Tyr218 (0.77), Lys217 (0.75), Gly238 (0.42), Gly224 (0.36), Tyr226 (0.31), Ile227 (0.30), and Ser216 (0.28). Mapping of the chemical shift changes on the structure identifies a pocket formed by the β-sheet β5-β7-β6-β8 and the surrounding loops (β4-β5, β5-β6 and α2-β8) (**Fig. 4C**). Some of the biggest changes come from the residues in the β4-β5 loop suggesting the loop is involved in nucleotide binding.

We hypothesized that the N-terminal KLAMP1 domain also binds nucleotides. NMR titrations using the same set of ligands (ATP, ADP, AMP and GDP) revealed a preference of KLAMP1 for ADP, albeit with significantly lower affinity. The K_d of ADP binding was estimated to be 800±250 μM, while the affinity towards other nucleotides was much lower and could not be reliably measured. Unexpectedly, titration of ^{15}N-labeled KLAMP1 with 3′,5′-cAMP resulted in large chemical shift changes for a number of signals, showing binding with a K_d estimated to be 280±32 μM (**Fig. 4.5A & B**). While sequence-specific signal assignments were not obtained for the KLAMP1 domain, the changes in the spectrum were similar, suggesting that KLAMP1 and KLAMP2 bind nucleotides in a similar fashion. The binding of the cyclic nucleotide is specific, as the NMR titration of KLAMP1 with 2′,3′-cAMP displayed no interaction. In addition, no binding was observed upon addition of 3′,5′-cAMP or 2′,3′-cAMP to KLAMP2. These experiments show that KLAMP1 and KLAMP2 are nucleotide-binding domains, but with differing specificities: KLAMP1 for 3′,5′-cAMP, and KLAMP2 for ADP.

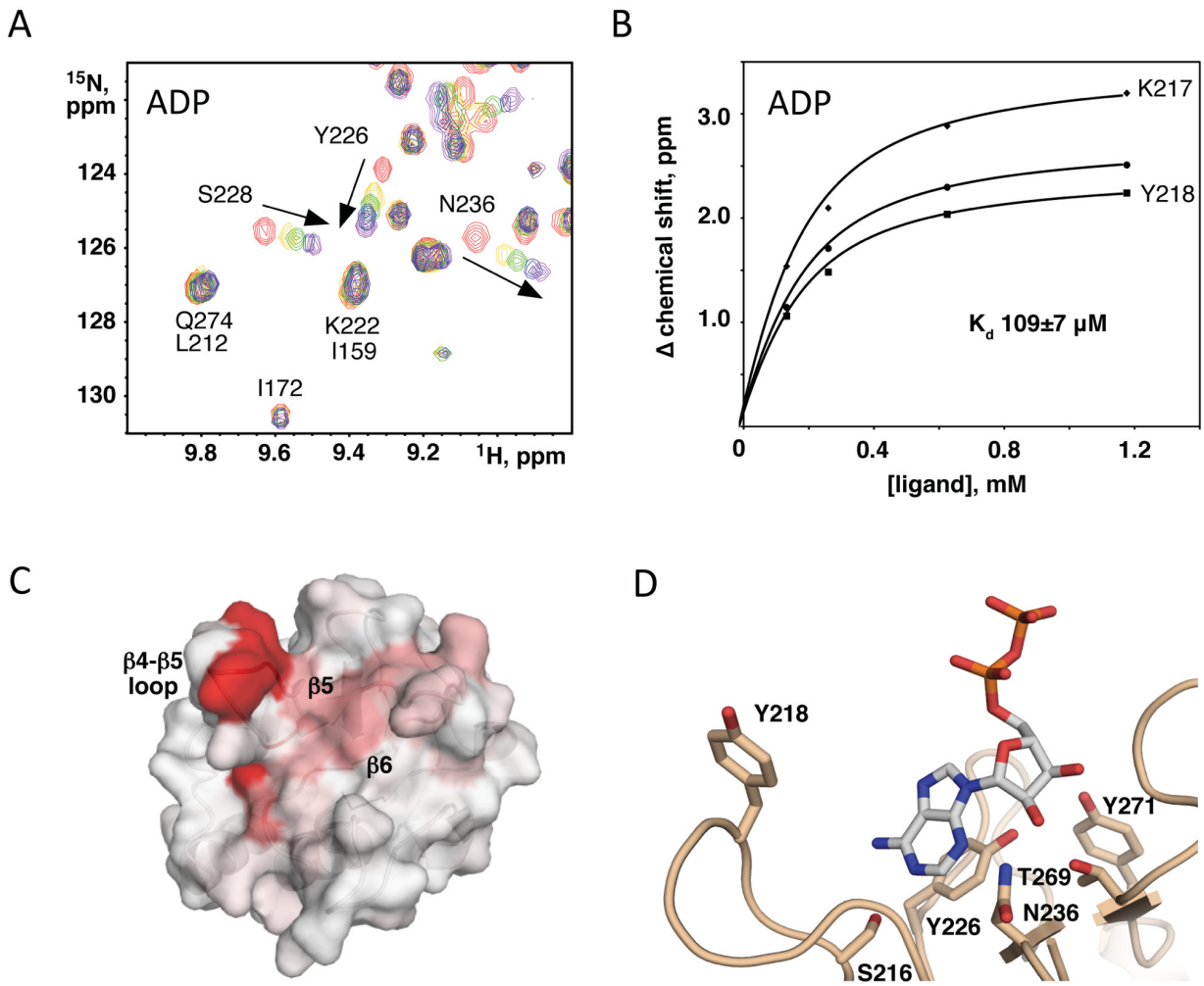


Figure 4.4 KLAMP2 ADP binding. (*A*) Downfield region of HSQC spectra of ^{15}N-labeled KLAMP2 domain titrated with increasing amounts of ADP at 0 mM (red), 0.13 mM (yellow), 0.26 mM (green), 0.63 mM (blue) and 1.18 mM (purple). The spectra show specific chemical shift changes for a number of signals. (*B*) K_d of the binding estimated from a fit of the ^{15}N chemical shift changes for two assigned and one unassigned signal that show large chemical shift changes. (*C*) Mapping of the chemical shifts measured onto the structure of the KLAMP2 domain. Red indicates largest chemical shift changes; white indicates no change detected. The most affected surface is centered on the β4-β5 loop and strands β5 and β6. (*D*) Model of an ADP bound KLAMP2 domain based on the 3',5'-cAMP/KLAMP1 structure, with residues important for binding shown.

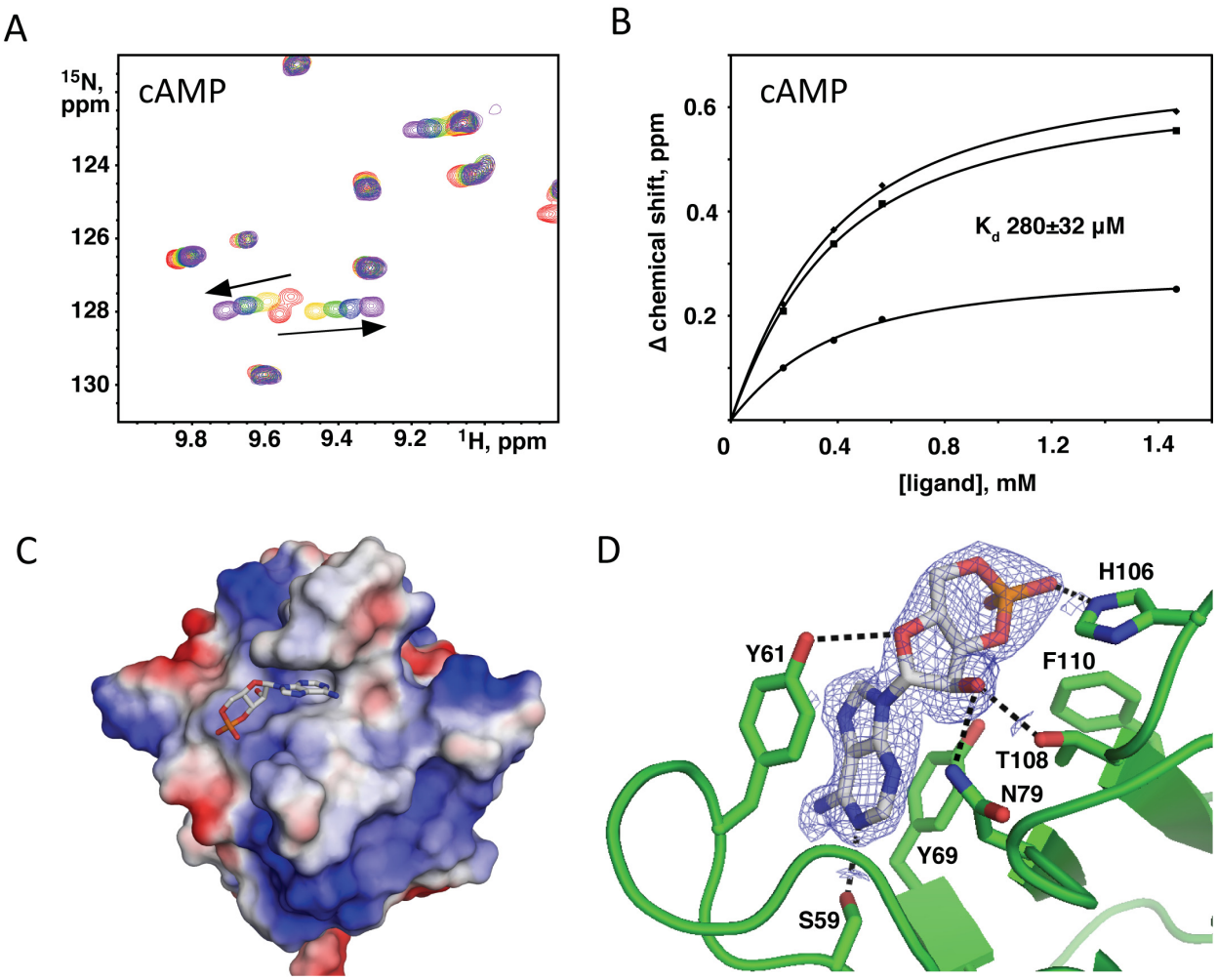


Figure 4.5 KLAMP1 3',5'-cAMP binding. (*A*) Downfield region of HSQC spectra of ^{15}N-labeled KLAMP1 domain show specific changes upon titration with increasing amounts of 3′,5′-cAMP at 0 mM (red), 0.20 mM (yellow), 0.39 mM (green), 0.57 mM (blue) and 1.67 mM (purple). (*B*) K_d of the binding estimated from a fit of the ^{15}N chemical shift changes for several signals with largest chemical shift changes. The NMR resonances of the KLAMP1 domain have not been assigned. (*C*) Surface charge representation of the KLAMP1 domain (blue = positive, red = negative) with the bound 3',5'-cAMP (grey). (*D*) Principal contribution to the binding comes from π-stacking of Tyr61 and Tyr69 with the adenine ring of cyclic nucleotide. Interaction is stabilized by hydrogen bonds (dash lines) between 3′,5′-cAMP and the side chains of Ser59, Tyr61, Asn79, His106 and Thr108.

4.3.6 Molecular determinants of 3′,5′-cAMP recognition by KLAMP1. In order to understand the molecular basis of nucleotide binding by the KLAMP domains, we soaked the crystals of KLAMP1 and KLAMP2 with 3',5'-cAMP and ADP, respectively. The electron density map of the KLAMP1 crystal soaked with 3′,5′-cAMP showed easily interpretable density for the nucleotide and the structure of the complex of KLAMP1 with 3′,5′-cAMP was refined to 2.1 Å (**Table 4.2**). The adenine ring fits snugly into a narrow ridge formed by the β4-β5 loop and the β-strand β5 (**Fig. 4.5C**). Three KLAMP1 residues are directly involved in the recognition of the adenine ring. Two tyrosine side chains (Tyr61 and Tyr69) provide π-stacking interactions with the adenine ring, while the N1 atom of adenine is hydrogen-bonded with the side chain of Ser59 (**Fig. 4.5D**). Comparison of the nucleotide-bound and unliganded KLAMP1 structures reveals movement of the β4-β5 loop, which closes on the nucleotide upon binding and presents Tyr61 as a major binding determinant. Three residues are engaged in polar contacts with the ribose ring of nucleotide. The side chain of Tyr61 hydrogen bonds with the oxygen atom of the ring, while the side chains of Asn79 and Thr108 both bind 2'OH. Finally, the O2 atom of the phosphate group forms a hydrogen bond with the side chain of His106.

We mutated residues Tyr61 and Tyr69 to test their roles in nucleotide binding. NMR spectra of the Y61A and Y69A KLAMP1 mutants are very similar to the wild-type protein, confirming that they are still well folded. However, NMR titrations of the ^{15}N-labeled mutants with 3′,5′-cAMP yielded no spectral changes confirming that these mutations abolished binding (data not shown).

The structure explains similarities and differences in the nucleotide-binding specificities of KLAMP domains. An overlay reveals the residues involved in adenosine recognition are strictly conserved, as Tyr218, Tyr226 and Ser216 of KLAMP2 correspond to Tyr61, Tyr69 and Ser59 of

KLAMP1 (**Fig. 4.3A**). Even the KLAMP1 residues involved in forming hydrogen bonds with ribose (Asn79, Thr108 and Tyr61) are conserved in KLAMP2 (Asn236, Thr269 and Tyr218) (**Fig. 4.4D**). Furthermore, the complex structure explains why the KLAMP domains will not recognize guanine and pyrimidine rings. Compared to adenine, guanine has an extra NH_2 group, which will push the ring out of the binding site. NMR titrations experimentally verified that GDP and 3′,5′-cGMP do not interact with the KLAMP domains. On the other hand, pyrimidines (cytosine and thymine) possess smaller rings, which will not reach the conserved serine for hydrogen bonding and would also result in insufficient stacking with the tyrosines.

The specificity of KLAMP1 and KLAMP2 for different adenine-containing ligands results from two structural differences. One of them is a rather subtle substitution of Phe110 in KLAMP1 for Tyr271 in KLAMP2. A tyrosine residue in this position would clash with the phosphate of 3'5'-cAMP, whereas a phenylalanine residue is in close contact with this phosphate (**Fig. 4.5D and 4.4D**). KLAMP2 also possesses a larger β7-β8 loop, which would also clash with the 3′,5′-cAMP phosphate group. The position of this phosphate group also explains the specificity of KLAMP1 for 3′,5′-cAMP as opposed to 2′,3′-cAMP, whose binding would be obstructed by the side chains of Phe110, Thr108 and Tyr69.

KLAMP2 is highly selective against binding cyclic nucleotides, but among the non-cyclic adenosine phosphates, there is increasing affinity for AMP, ATP and ADP. Analysis of the structure suggests that the length of two phosphates may allow ADP to reach the β7-β8 loop and form a hydrogen bond with Thr265. The third phosphate of ATP could be repelled by Glu264, while AMP is too short to make any contact with the loop.

4.4 Discussion

Here, we have identified a novel KLAMP domain, which is present in two copies in the N-terminal half of lpg1496, a *Legionella pneumophila* effector. Based on sequence similarity, a similar domain is also found in histidine kinase-like ATP-binding region-containing proteins (HATPase) and S-adenosylmethionine-dependent methyltransferase proteins (AdoMet MTase). The structures of the domains from lpg1496 are very similar to each other, but do not display significant structural similarity to other known protein structures. More significantly, we demonstrate using NMR that both domains bind nucleotides, albeit with different specificity.

Sequence alignment of KLAMP1/2 of lpg1496, HATPase, and AdoMet MTase highlight the importance of several conserved residues. Ser59 of lpg1496 that hydrogen bonds with the adenine ring through N1 is generally conserved. Aromaticity is also conserved at the Tyr69 position, with either tyrosine or phenylalanine, for π-stacking with adenine. Polar contacts with the 2'OH atom of the ribose ring is conserved in Thr108. This suggests that all domains of the identified family may interact with adenosine-containing molecules.

Specificity of the KLAMP1 domain of lpg1496 for 3′,5′-cAMP is intriguing. Lipopolysaccharides (LPS) found on the cell wall of gram-negative bacteria contribute to the activation of host inflammatory responses, but also serve to promote survival of the bacterium (Morrison and Ryan, 1987). For example, LPS induces arachidonic acid release, which in turn is metabolized to prostaglandins and leukotrienes. Increased release of prostaglandin E_2 (PGE_2) has been detected following activation of macrophages with LPS (Rosenstreich et al., 1977). PGE_2 suppresses microbicidal activity of macrophages through G_S-coupled receptors, increasing adenylyl cyclase activity and effectively increasing intracellular cAMP levels. cAMP functions as a secondary messenger influencing numerous cellular functions, acting through the downstream

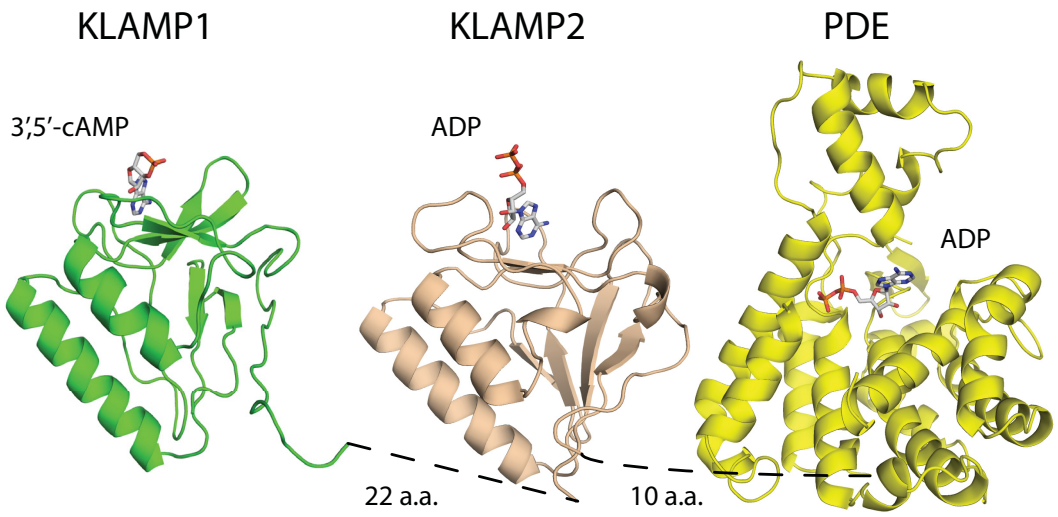
effectors, protein kinase A (PKA), Epac-1 and -2. Through cAMP, PGE_2 inhibits the microbicidal production of reactive oxygen intermediates (ROI) by NADPH oxidase (Serezani et al., 2007). In summary, elevated cAMP levels result in increased bacterial survival in macrophages.

Comparison with other protein structures in complex with 3′,5′-cAMP shows that in many cases, adenine recognition elements involve π-stacking with a tyrosine or phenylalanine residue, while KLAMP1 simultaneously uses two tyrosines for π-stacking with adenine. To our knowledge, there is only one other structure in Protein Data Bank, where the adenine ring is sandwiched between two aromatic residues (both tyrosines) providing π–stacking interactions (PDB code 1LPC) (Kurinov et al., 2004). However in KLAMP1, one of the tyrosines (Tyr61) additionally hydrogen bonds to the oxygen of the ribose ring, increasing its importance in ligand recognition.

We have also crystallized the conserved C-terminal PDE domain of lpg1496 that is found in the N-terminal region of members of the original SidE family. Bioinformatic analyses highlight a potential function for lpg1496, as this domain can be found in combination with the ADP-ribosylating domain, Vip2. In addition, we have crystallized the catalytic PDE domain in complex with ADP in a possible substrate-binding site. An overlay of this structure with a structurally similar phosphodiesterase shows that this binding occurs in a shifted catalytic pocket, explaining the inactivity of lpg1496 against cyclic nucleotides.

Taken together, all three domains of lpg1496 are capable of binding nucleotides pointing towards a connection with nucleotide metabolism (**Fig. 4.6**). The importance of nucleotide-binding for lpg1496 function can be tested in future cell-based assays using mutations of key residues identified here: Tyr61 and Tyr69 for 3′,5′-cAMP/KLAMP1, Tyr218 and Tyr226 for ADP/KLAMP2, and His366 and His370 for ADP/PDE. The discovery of KLAMP domains, a

novel nucleotide-binding fold, will have implications for understanding the function of other KLAMP-domain containing proteins.



Ligand	KLAMP1	KLAMP2	PDE
ATP	No binding	800±150 μM	Not determined
ADP	800±250 μM	109±7 μM	Co-crystal
AMP	No binding	1500±200 μM	Not determined
GDP	No binding	No binding	Not determined
3′,5′-cAMP	280±32 μM	No binding	Not determined
2′,3′-cAMP	No binding	No binding	Not determined

Figure 4.6 Lpg1496 structure and characterization. Schematic model of the arrangement of the KLAMP and PDE domains in lpg1496 with nucleotides bound. A model of ADP bound to KLAMP2 is shown. A summary of the binding affinities measured by NMR titration is presented for the KLAMP domains.

4.5 Experimental procedures

4.5.1 Cloning, protein expression, and purification. The gene lpg1496 from *Legionella pneumophila* strain Philadelphia was cloned into pLR652 as a N-terminal GST-tagged fusion protein and expressed in a BL21 Star E. coli strain. The cells were grown at 37ºC in Luria Broth to an optical density of 0.8, and expression was induced with 1 mM isopropyl β-D-1-

thiogalactopyranoside (IPTG) at 30°C for 4 hours or 16ºC overnight. After centrifuging the cells, the pellets were resuspended in phosphate buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4), containing 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.1 mg/ml lysozyme and 80U deoxyribonuclease (DNAse), and lysed by sonication. Cell debris was removed by centrifugation, and the GST-fusion protein was purified using Glutathione-Sepharose affinity columns (GE Healthcare). After eluting the protein in PBS containing 20 mM glutathione, the middle domain of lpg1496 was obtained by a 2 hour trypsin cleavage in PBS and further purified on a Superdex75 gel filtration column (GE Healthcare) in buffer A (10 mM HEPES pH 7.0, 100 mM NaCl) before crystallization trials.

The N-terminal domain of lpg1496 (residues 1-138) was cloned into pET29a, using the following primers: 5'-agatatacatatggttacgaaaataatttgggtttc-3' and 5'-ctagctcgagttttgttacgggaacaataacaggtg-3', as a C-terminal His-tagged fusion, and transformed into a BL21 *E. coli* strain. Mutagenesis for Y61A and Y69A was then performed using 5'-ctgatagaagcttcaaatgccccaattaatccttgtgg-3' and 5'-taatccttgtggttgtgctatatccccaggtggg-3', respectively. The expression conditions were the same as for the full-length protein. The fusion protein was bound to Ni–NTA Agarose (Qiagen) beads, washed with buffer B (50 mM HEPES pH 7.6, 0.5 M NaCl, 5% (v/v) glycerol) containing 30 mM imidazole and eluted with buffer B containing 250 mM imidazole. The protein was further purified by size-exclusion chromatography on a Superdex75 gel filtration column (GE Healthcare) in buffer A.

Three constructs of the PDE domain of lpg1496 were cloned. Lpg1496 (293-580) was cloned into pET29a using the following primers: 5'-agatatacatatggaaatagagaaaaatgattatctactatc-3' and 5'-ctagctcgagctttagacactcattgggatc-3', as a C-terminal His-tagged fusion protein. Lpg1496 (293-598) was cloned into pET15b using the following primers: 5'-

agatatacatatggaaatagagaaaaatgattatctactatc-3' and 5'-cgcggatccttaaataccatattgatttgccaag-3', as a N-terminal His-tagged fusion protein. Lpg1496 (154-598) was cloned into the pET29a vector using 5'-agatatacatatggattcttcgatttctattagtgc-3' and 5'-ctagctcgagaataccatattgatttgccaag-3'. Constructs were verified by DNA sequencing. *E.coli* BL21 star cells were transformed with plasmids encoding the three constructs. The expression and purification conditions were the same as for the N-terminal domain, except size-exclusion chromatography was performed using buffer C (10 mM HEPES pH 7.5, 100 mM NaCl). The His-tag in the pET15b construct was cleaved with thrombin before injecting the protein into a size-exclusion column.

For ^{15}N-labeling, the cells were grown in M9 minimal medium supplemented with ^{15}N-ammonium chloride as the sole source of nitrogen. For ^{15}N,^{13}C double-labeling, the cells were grown in M9 minimal medium supplemented with ^{15}N-ammonium chloride and $^{13}C_6$-D-glucose as the sole sources of nitrogen and carbon.

For selenomethionine labeling, the plasmid was transformed into a methionine-auxotroph DL41 (DE3) E.coli strain, and the cells were grown in LeMaster medium supplemented with selenomethionine. The expression and purification protocols were the same as for the native protein.

4.5.2 Crystallization, data collection, and processing. Crystallization was performed by the hanging drop vapor diffusion method at 293K using the Classics II and JCSG+ Suite commerical screens (Qiagen). Crystals of the middle domain of lpg1496 were obtained in approximately one week from a 1:1 mixture of the protein solution (16.6 mg/ml) and the reservoir solution (0.1 M HEPES pH 7.5, 25% (w/v) PEG3350). Crystals of the N-terminal domain were obtained by equilibrating a drop consisting of 0.6 μL lpg1496 (residues 1-138) (15 mg/mL) and 0.6 μL of 0.2 M sodium formate, and 20% (w/v) PEG3350. Crystals of lpg1496 (1-138) used for

soaking experiments were obtained in a 1:1 mixture of protein at 37.8 mg/mL and the mother liquor (0.2 M sodium chloride, 0.1 M Bis-Tris pH 6.5, 25% (w/v) PEG3350). Soaking experiments with 3´,5´-cAMP were performed by dipping the crystal into a solution of mother liquor containing 15 mM 3´,5´-cAMP for 20 minutes. Crystals of lpg1496 (293-580) were obtained overnight from a 1:1 mixture of the protein solution (11 mg/mL) and the reservoir solution (1.1 M sodium malonate, 0.1 M HEPES pH 7.0, 0.5% (v/v) Jeffamine ED-2001). Crystals of lpg1496 (293-598) at 7.1 mg/mL were obtained in condition #15 of the JCSG+ Suite (0.1 M Bicine pH 9.0, 20% (w/v) PEG 6000) in approximately one week. Crystals of the SeMet-labeled lpg1496 (293-598) were obtained in a week from a 1:1 mixture of protein at 10 mg/mL with the mother liquor (0.1 M Bis-Tris pH 6.5, 20% (w/v) PEG 5000 MME). Lpg1496 (154 – 598) was crystallized in 0.1 M HEPES pH 7.5, 25% (w/v) PEG3350 and 5 mM ADP.

The crystals were cryoprotected with 10 – 25% glycerol or ethylene glycol, and flash-cooled in a N_2 cold stream. X-ray diffraction data were collected using an ADSC Quantum 210 CCD detector (Area Detector Systems Corp.) on beamline A1 at the Cornell High-Energy Synchrotron Source (CHESS) at 0.9770 Å. Data processing and scaling were performed with HKL-2000 (**Table 4.1 and 4.2**) (Minor, 1997).

4.5.3 Structure determination and refinement. The diffraction data of the middle domain were phased using anomalous signal from selenium atoms by the single-wavelength anomalous dispersion (SAD) method, with the program SHELX (Sheldrick, 2008). The initial model, consisting of more than 90% of the residues, was built with ARP/wARP (Langer et al., 2008) and refined with Refmac5 (Murshudov et al., 2011).

The N-terminal domain structure (residues 1-138) was determined by molecular replacement (MR) using the middle domain of lpg1496 as the search model. The initial model was

built by ARP/wARP (Langer et al., 2008) using phases from the MR solution. Model building was completed with the help of the program Coot (Emsley and Cowtan, 2004) and was improved by several cycles of refinement using Refmac5 (Murshudov et al., 2011).

The N-terminal domain structure in complex with 3′,5′-cAMP was determined by MR using lpg1496 (1-138) as the search model. The model was built by PhaserMR (McCoy et al., 2007) and completed with Coot (Emsley and Cowtan, 2004). Refmac5 was used to improve density (Murshudov et al., 2011). The water molecules were added in the last stage of refinement. The refinement statistics for the N-terminal half are in **Table 4.2**.

Diffraction data of the 293-598 construct were phased using SAD, with the program PHENIX (Adams et al., 2010). The initial model was built using PHENIX (Adams et al., 2010), improved using Coot (Emsley and Cowtan, 2004) and further refined using PHENIX (Adams et al., 2010).

Data from the 293-580 and 154-598 constructs were phased by MR using the lpg1496 (293-598) structure as the search model. The initial model was built by ARP/wARP (Langer et al., 2008) and further built with Coot (Emsley and Cowtan, 2004). The final model was improved by several cycles of refinement using Refmac5 (Emsley and Cowtan, 2004). The refinement statistics are shown in **Table 4.1**.

The final models have no outliers in the Ramachandran plot computed using PROCHECK (Laskowski, 1993) and MolProbity (Chen et al., 2010). Coordinates of the middle domain in 2 space groups ($P2_1$ and $P2_12_12_1$) and the N-terminal domain without and with 3′,5′-cAMP have been deposited in the Protein Data Bank with the accession codes 5BTY, 5BTZ, 5BTW and 5BTX, respectively. Coordinates of the native PDE domain (293-580, 293-598) and its ADP bound

structure have been deposited in the Protein Data Bank with the accession codes 5BU1, 5BU0 and 5BU2, respectively.

4.5.4 NMR spectroscopy. NMR resonance assignments of the ^{15}N,^{13}C-labeled middle domain of lpg1496 were determined using HNCACB and HN(CO)CACB experiments. The 3D heteronuclear experiments were recorded at 298 K on a Bruker 800 MHz spectrometer. The samples were prepared as 400 μM in 90% buffer A and 10% D_2O. NMR spectra were processed with NMRPipe (Delaglio et al., 1995) and analyzed with SPARKY (Goddard and Kneller).

NMR titration experiments were performed at 303K on a Bruker 600 MHz spectrometer. For titrations with nucleotides, NMR samples of the ^{15}N-labeled lpg1496 (1-138) construct were prepared as 0.17-0.20 mM in 90% buffer A and 10% D_2O. Samples of the middle domain of ^{15}N-labeled lpg1496 were prepared as 135 μM in 90% buffer A and 10% D_2O. $^{15}N-^{1}H$ heteronuclear single quantum correlation spectroscopy (HSQC) titrations were performed by stepwise addition of 2′,3′-cAMP, 3′,5′-cAMP, 3′,5′-cGMP, ATP, ADP, and AMP to a final molar ratio of 1 to 10 of N-terminal lpg1496 to ligand. The N-terminal mutants were titrated with 3′,5′-cAMP only. $^{15}N-^{1}H$ HSQC titrations were performed by stepwise addition of ATP, ADP, AMP and GDP to the middle domain to a final molar ratio of 1 to 10 (50 for AMP). Minimal changes in volume and pH were ensured throughout the sample preparations. The NMR spectra were processed by NMRPipe (Delaglio et al., 1995) and analysed using SPARKY (Goddard and Kneller).

The affinities were measured by the calculations of the dissociation constant (K_d). Chemical shift changes were fitted to Equation 1,

$$C = C_{max}\frac{K_d+P_{tot}+L_{tot}-\sqrt{(K_d+P_{tot}+L_{tot})^2-4P_{tot}L_{tot}}}{2P_{tot}} \quad \text{(Eq. 1)}$$

where C is the chemical shift perturbation, C_{max} the chemical shift perturbation at saturation, K_d the dissociation constant, P_{tot} the total concentration of the labeled protein and L_{tot} the total ligand concentration.

4.5.5 Malachite green assay. Phosphodiesterase assays were performed in 96-well plates, in 10mM HEPES pH7.0, 100 μM nucleotide (3'-AMP, 5'-AMP, ADP, ADP ribose, 2',3'-cAMP, 3',5'-cAMP and 3',5'-cGMP) and 2.5 μM full length lpg1496, to a total reaction volume of 75μL. When necessary, 6U alkaline phosphatase was used to release the phosphate group for detection. The reaction was incubated at 37ºC for 15 minutes and stopped by addition of 43 μL 28 mM ammonium molybdate in 2.1 M H_2SO_4 and 32 μL 0.76 mM malachite green in 0.35% polyvinyl alcohol (MW ~ 16000 Da). Free phosphate was determined by measuring absorbance at 610nm.

4.6 Acknowledgements

We thank Dr. Miroslaw Cygler (University of Saskatchewan) for the initial construct of lpg1496, Dr. Irena Ekiel for helpful discussions, and the lab of Dr. Alexander Yakunin (University of Toronto) for screening potential substrates of lpg1496. KW acknowledges a studentship from Canadian Institutes of Health Research (CIHR). This work was funded by a CIHR Genomics grant GSP-48370. Data were acquired at the Macromolecular Diffraction (MacCHESS) facility at the Cornell High Energy Synchrotron Source (CHESS). CHESS is supported by the NSF & NIH/NIGMS via NSF award DMR-0225180, and the MacCHESS resource is supported by NIH/NCRR award RR-01646. The authors declare that they have no conflicts of interest.

Chapter 5: E1/E2 independent ubiquitination

5.0 Connecting text

Sequence analysis of the C-terminus PDE domain of lpg1496 reveals sequence similarity to the PDE domain of the original members of the SidE family (SdeA, SdeB, SdeC and SidE). The SidE proteins are involved in ADP-ribosylation of ubiquitin and subsequent phosphoribosyl ubiquitination of substrates. Some proteins containing the PDE domain also contained the Vip2 domain, which is present in an actin-ADP-ribosylating toxin family. This points to the potential ability of lpg1496 in replacing PDE function in the phosphoribosyl-ubiquitination pathway. Although this novel ubiquitination pathway has been discovered, its mechanism is not yet understood.

5.1 Summary

Ubiquitination is a eukaryotic post-translational process. *Legionella pneumophila* has developed strategies to control this process. One such strategy involves injecting bacterial virulence factors into the host cell. The SidE family of effectors processes ubiquitin in a novel way, first through ADP-ribosylation by its mono ADP-ribosyltransferase (mART) domain, and finally through cleaving off an AMP group by its phosphodiesterase (PDE) domain, leaving a phosphoribosyl-ubiquitinated substrate. We show here that the PDE domain found in a related protein, lpg1496, functions as a general phosphodiesterase, but is unable to remove AMP from ADP-ribosylated ubiquitin as by the SidE family. We demonstrate the ability of the sequential and separate activities of the mART and PDE domains to complement each other in trans. We have also crystallized a construct potentially containing both mART and PDE domains.

5.2 Introduction

Ubiquitination is a complex protein modification used by eukaryotes as a signal for various cellular processes. Depending on the type of ubiquitin linkage, these include autophagy, DNA repair, protein degradation, and immune responses, among others. The canonical ubiquitination pathway involves three enzymes. The E1 activating enzyme activates ubiquitin (Ub) in the presence of ATP and Mg^{2+}, covalently linking the Ub to itself via a thioester bond. The activated Ub is then transferred to an E2 conjugating enzyme, which attaches Ub to a substrate lysine with the help of an E3 ligase (Pickart and Eddins, 2004) (**Fig. 5.1A**). The discovery of a family of Ub ligases, SidE family, in the pathogen *Legionella pneumophila,* capable of functioning independently of the E1 and E2 enzymes has led to increasing research on this novel mechanism. Not only does the SidE family function as an E1, E2–independent enzyme, it generates a new type of Ub, phosphoribosylated Ub (P-Rib-Ub).

Since the Ub proteasome system only exists in eukaryotes, many bacterial pathogens have developed techniques to manipulate it for their own benefit. *Legionella pneumophila* is one such intracellular pathogen that has evolved eukaryotic-like E3s to interact with the host E1 and E2, allowing the bacteria to select its own substrates for ubiquitination. Well known examples include the U-box containing LubX (Quaile et al., 2015), and the F-box containing AnkB (Wong et al., 2017). Qiu et al. identified a novel Ub ligase family (SidE family) capable of ubiquitinating a substrate without ATP, Mg^{2+}, E1 or E2 (Qiu et al., 2016). Instead, only NAD was required. SdeA uses its mono-ADP ribosyl transferase (mART) domain and NAD to generate ADP-ribosylated Ub at Arg42 (ADP-Rib-Ub). The phosphodiesterase (PDE) domain then cleaves ADP-Rib-Ub to release AMP and attach P-Rib-Ub onto a substrate serine (Qiu et al., 2016) (**Fig. 5.1B and 5.2A**). However, the mechanism of how these two domains function remain unsolved.

The SidE family consists of four members, SdeA, SdeB, SdeC and SidE (Bardill et al., 2005). They are relatively large proteins, greater than 170 kDa and consist of three domains, an N-terminal deubiquitinase (DUB) domain, followed by a PDE domain and a mART domain (**Fig. 5.2B**). The DUB domain is proposed to aid in survival of the pathogen by disrupting the host cell's ability to recruit autophagy machinery, through deubiquitinating Lys63 chains on the LCV (Sheedlo et al., 2015).

To date, only three substrates of SdeA modification have been identified. ER-associated Rab GTPases are a common substrate of *Legionella* effectors, since ER vesicles are required for the formation of the LCV. Indeed, Rab33b overexpression has been shown to restrict bacterial growth (Qiu et al., 2016). Another ER-associated protein, Reticulon 4 (Rtn4), is also an identified substrate. Ubiquitination of this structural ER membrane protein enhances its oligomerization, resulting in appendages involved in ER remodeling and essential for formation of a replicative vacuole (Kotewicz et al., 2017). Finally, when water is the substrate, P-Rib-Ub itself is generated. This product has been shown to inhibit the canonical ubiquitination pathway by decreasing the efficiency of both loading of P-Rib-Ub onto an E2 enzyme, and discharging of a P-Rib-Ub loaded E2 (Bhogaraju et al., 2016).

Here, we demonstrate that the PDE domain found in a related effector, lpg1496, is unable to replace PDE function of a SidE family member. We show that the PDE and mART domains of an original SidE family member, can complement each other in trans.

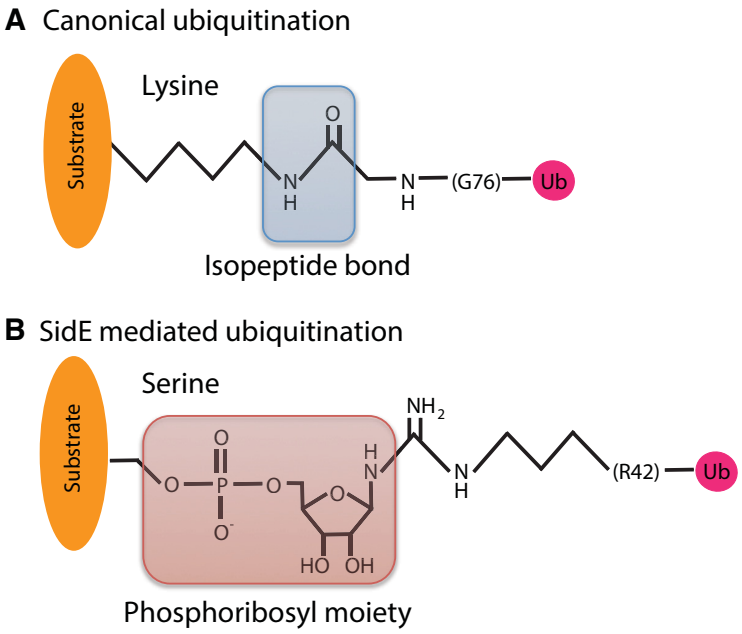


Figure 5.1 Products of ubiquitination. (*A*) Canonical ubiquitination generates an isopeptide bond between a substrate lysine to Gly76 of ubiquitin. (*B*) SidE family mediated ubiquitination creates a phosphoribosyl linkage between a substrate serine to Arg42 of ubiquitin.

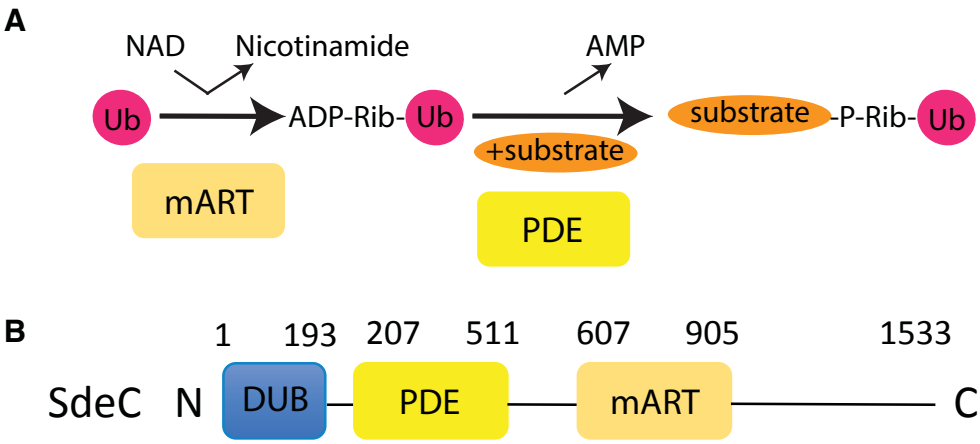


Figure 5.2 Phosphoribosyl ubiquitination by the SidE family. (*A*) Reaction mechanism mediated by two domains of SidE proteins. The mART domain mediates the first step where ubiquitin is first ADP-ribosylated using NAD. Then the PDE domain cleaves off AMP, attaching a phosphoribosylated ubiquitin onto a substrate. (*B*) Domain architecture of SdeC, a representative member of the SidE family. SidE proteins contain an N-terminal DUB (deubiquitinase) domain, followed by a PDE (phosphodiesterase) domain and a C-terminal mART (mono ADP-ribosyl transferase) domain.

5.3 Results and discussion

5.3.1 Lpg1496 is an orphan SidE-related member. The PDE domain found in lpg1496 contains a 25% sequence identity with that of the PDE domain in SdeC. To determine if lpg1496 could replace the PDE function in the phosphoribosyl ubiquitination mechanism and process the downstream product of SdeC-mART, a simple ubiquitination assay using HA-ubiquitin, Etheno-NAD, and SdeC, was performed where self-ubiquitination of SdeC was indication of a successful reaction, as in lane 1. Lpg1496 had no effect on the function of WT-SdeC. When a PDE inactive mutant of SdeC is used, the phosphoribosyl ubiquitination reaction stops after the first step. The active mART domain forms ADP-Rib-Ub, but the inactive PDE domain is unable to generate the final product, substrate-P-Rib-Ub (**Fig. 5.2A**). Addition of lpg1496 to this PDE inactive SdeC was unable to rescue the defect (**Fig. 5.3**). Indeed, lpg1496 was unable to ubiquitinate a known SidE family substrate (Kotewicz et al., 2017).

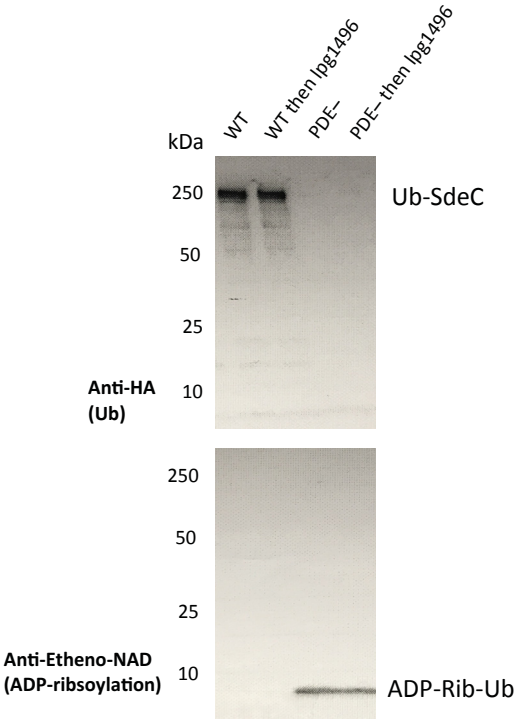


Figure 5.3 Lpg1496 is an orphan SidE-related member. Different combinations of SdeC and lpg1496 were incubated with HA-Ub and Etheno-NAD. Reactions were separated by SDS-PAGE and probed for anti-HA (Ub, top blot) and anti-ethenoadenosine (ADP-Rib-Ub, bottom blot). Lanes: WT, WT-SdeC; WT then lpg1496, 30 minutes of WT-SdeC followed by 30 minutes of lpg1496; PDE-, H416A PDE inactive SdeC mutant; PDE- then lpg1496, 30 minutes of PDE- followed by 30 minutes of lpg1496.

5.3.2 Lpg1496 is a general phosphodiesterase. Although PDE-lpg1496 does not function in the phosphoribosyl ubiquitination reaction scheme, it is still classified as a PDE based on a structural homology search against the DALI database (Holm and Rosenström, 2010). To test this hypothesis, a spectrophotometric assay involving cleavage of a general colorless phosphodiesterase substrate, bis(para-nitrophenyl) phosphate (bis-pNPP) to a yellow product, para-nitrophenol (pNP), was employed. Lpg1496 showed maximum activity at pH 8.5 (**Fig. 5.4A**). From there, metal dependence was tested. Although the predicted metal coordination site, based on structural homology to LmjPDEB1, is partially occluded by the binding of ADP, all PDEs require divalent metal ions for function (Francis et al., 2011). It is possible that lpg1496 uses a different site for metal binding. In line with this, lpg1496 showed preference for manganese, increasing the reaction rate by ~3X. However, addition of a second metal did not increase efficiency (**Fig. 5.4B &C**). Specificity for the general phosphodiesterase substrate, bis-pNPP, as opposed to the general phosphatase substrate, pNPP, was also tested. Lpg1496 indeed shows greater activity as a phosphodiesterase (**Fig. 5.4D**).

Using optimal conditions, at pH 8.5 with manganese, different combinations of its three domains were tested. Its phosphodiesterase activity requires the full-length protein (**Fig. 5.5A**). Finally, single amino acid substitutions in full-length lpg1496 was tested. Three histidines (H366, H370, H504) within its substrate binding pocket were mutated to alanines (**Fig. 4.2B**).

Surprisingly, the conserved histidines with LmjPDEB1, H366 and H370, were not essential for cleavage of bis-pNPP to pNP. Rather a non-conserved histidine, H504, was responsible for a ~2X decrease in activity (**Fig. 5.5B**).

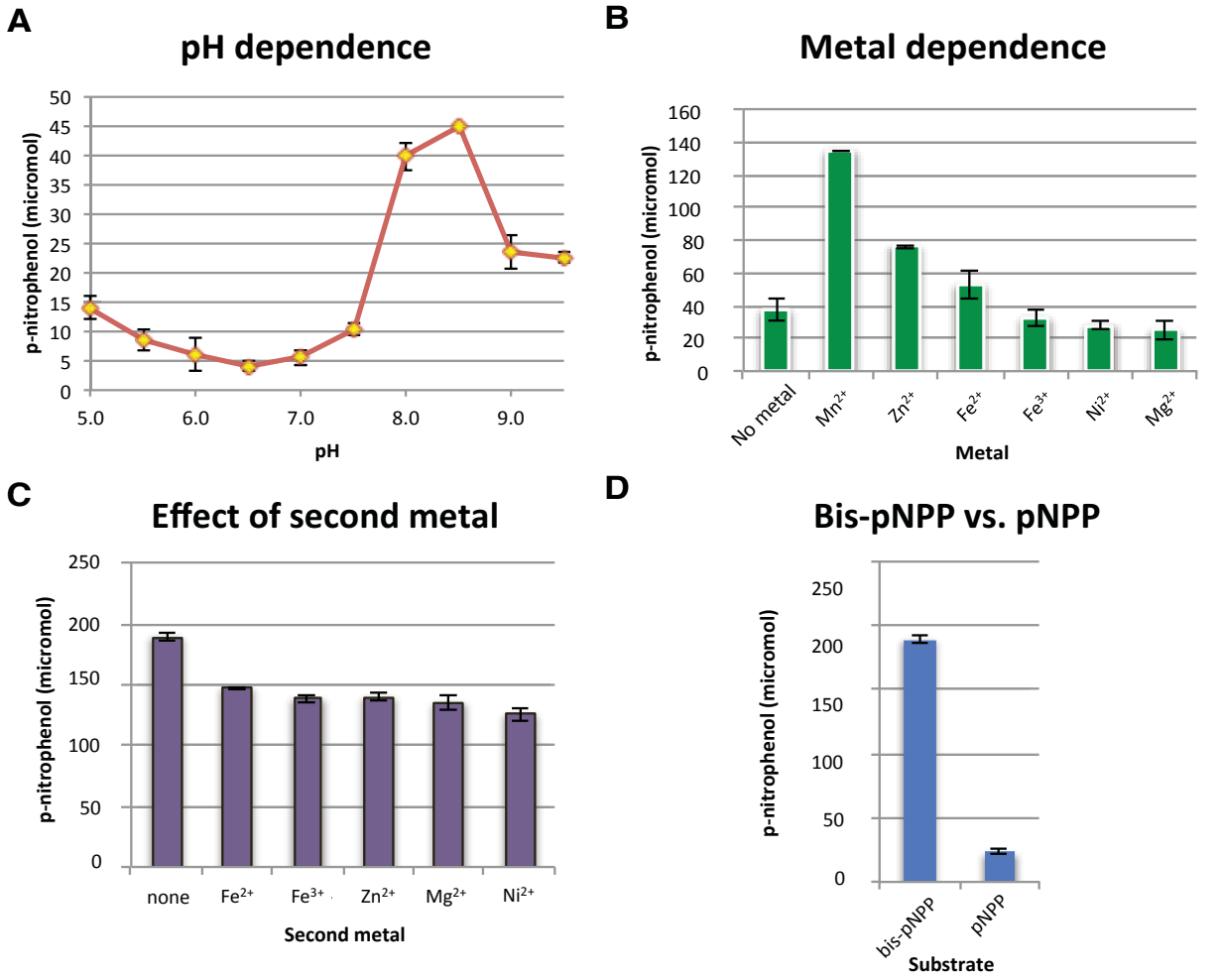


Figure 5.4 Lpg1496 functions as a general phosphodiesterase. 2.5 µM lpg1496 was incubated with 2 mM bis-pNPP or pNPP for 60 minutes at 37ºC. p-nitrophenol product formation was detected at 405 nm. Average values from triplicate readings are reported with error bars ± 1 standard deviation. (*A*) pH dependence of phosphodiesterase activity, from pH 5.0 to pH 9.5 was tested. (*B*) 5 µM of different metals was added to the reaction. (*C*) The effect of a second metal in addition to manganese was tested. (*D*) Lpg1496 has a preference for bis-pNPP as opposed to pNPP, confirming its function as a general phosphodiesterase.

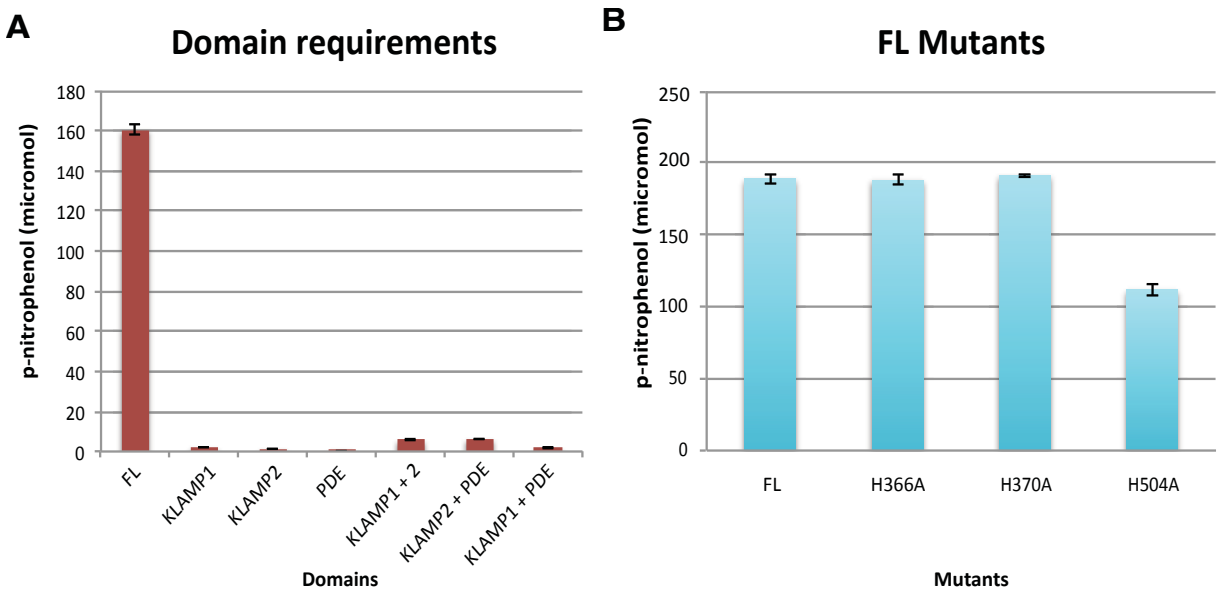


Figure 5.5 Lpg1496 activity requires the full-length protein. 2.5 μM lpg1496 was incubated with 2 mM bis-pNPP at pH 8.5 with 5 μM of Mn2+ for 60 minutes at 37ºC. p-nitrophenol product formation was detected at 405 nm. (*A*) Different combinations of the three domains, KLAMP1, KLAMP2 and PDE, are tested for activity against bis-pNPP. (*B*) Three histidines in full-length lpg1496, within the catalytic pocket of the PDE domain, are mutated to alanines and tested for importance in phosphodiesterase function.

5.3.3 Complementation of PDE and mART domains in trans. The PDE and mART domains of the SidE family are responsible for the two-step phosphoribosyl-ubiquitination of substrates (**Fig. 5.2A**) (Qiu et al., 2016). To determine if these two domains could complement each other in trans, inactive full-length mutants of each were generated (PDE- and mART-). A simple ubiquitination assay using HA-ubiquitin, Etheno-NAD, and SdeC, was performed where self-ubiquitination of SdeC was indication of a successful reaction, as in lane 7. Addition of the PDE- and mART- mutants together for 30 minutes or 60 minutes resulted in a WT reaction, indicating that the ADP-Rib-Ub product generated from a WT mART domain could be processed by the WT PDE domain of another protein. Next, we determined that the two domains did not need to be added to the reaction mixture at the same time. An active mART domain was able to

complement an inactive mART domain after 30 minutes and generate phosphoribosyl-ubiquitinated SdeC. The reverse, using the PDE domain, was also possible (**Fig. 5.6**).

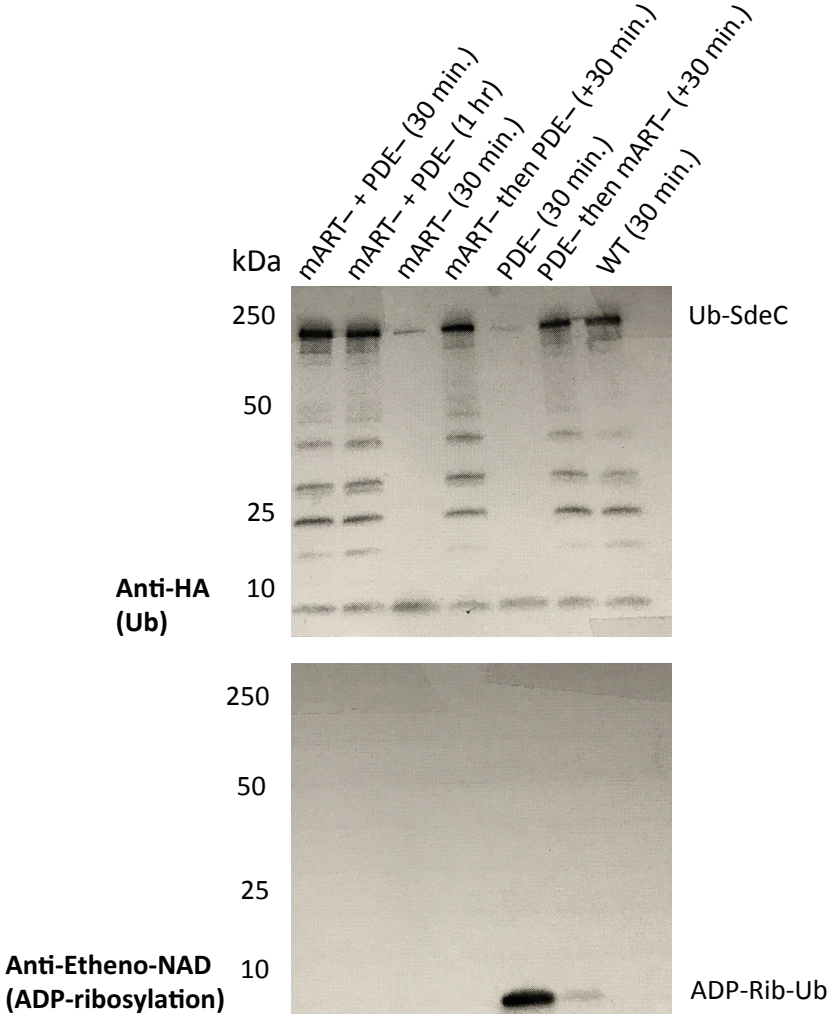


Figure 5.6 SidE family PDE and mART domains can function in trans to conjugate ubiquitin. SdeC was incubated with HA-Ub at 37ºC. Reactions were resolved by SDS-PAGE and probed for anti-HA (Ub, top blot) and anti-ethenoadenosine (ADP-Rib-Ub, bottom blot). Labels: mART-, E859A mART inactive SdeC mutant; PDE-, H416A PDE inactive SdeC mutant; WT, WT-SdeC.

5.4 Experimental procedures

5.4.1 Cloning, protein expression, and purification. Lpg1496 was cloned, expressed and purified as previously reported (Wong et al., 2015). Mutagenesis for H366A, H370A, and H504A

was performed using the QuikChange Multi Site-Directed Mutagenesis Kit (Agilent Technologies). Constructs were verified by DNA sequencing before transformation into BL21 *E. coli*. The expression conditions were the same as for the full-length protein.

The genes for SdeB and SdeC from *Legionella pneumophila* strain Philadelphia were cloned out of genomic DNA and into pRham as a C-terminal His-tagged fusion protein and expressed in a BL21 *E. coli* strain. The cells were grown at 37ºC in Luria Broth to an optical density of 0.8, and expression was induced with 2 grams of L-rhamnose (Sigma-Aldrich) per Liter of culture at 37ºC for 4 hours. 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 (PMSF) 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 anion exchange chromatography (BioSuite Q column, Waters) in 10 mM Tris pH 8.0 and eluted with a gradient of increasing NaCl concentration to 0.5 M. This was followed by size-exclusion chromatography on a Superdex 200 10/300 GL column (GE Healthcare) in 10 mM HEPES, 100 mM NaCl, pH 7.0 before crystallization trials.

5.4.2 Phosphoribosyl-ubiquitination assay. 10 μM HA-Ub (Boston Biochem), 100 μM etheno-NAD (Santa Cruz Biotechnology) and 20 nM SdeC were incubated in 20 mM Tris pH 7.4, 100 mM NaCl at 37ºC for the indicated times. The reaction was terminated by addition of reducing SDS-PAGE loading buffer. Samples were then separated by SDS-PAGE and analyzed by Western blotting. Anti-HA (1:500) and anti-ethenoadenosine (1:500) (Santa Cruz Biotechnology) were used to probe for ubiquitin and ADP-ribose respectively.

5.4.3 Phosphodiesterase assay. 2.5 μM lpg1496 was incubated with 2 mM bis(para-nitrophenyl) phosphate (bis-pNPP) or pNPP (Sigma Aldrich) in 10 mM buffer for 60 minutes at 37ºC. pH 5.0 was buffered using MES. pH 5.5 and pH 6.0 were made using Bis-Tris. pH 6.5 to pH 7.5 in increments of 0.5 were buffered with HEPES. pH 8.0 and pH 8.5 were made with Tris. pH 9.0 and pH 9.5 were buffered using Bicine. 5 μM of different metals were added to the reaction mixture to test for metal dependence as indicated. Para-nitrophenol (pNP) product formation was detected at 405 nm. Reactions were performed in a 96-well clear flat bottom plate and read using a SpectraMax M5e (Molecular Devices).

5.5 Acknowledgements

We thank Dr. Kristin Kotewicz (Isberg lab, Tufts University) for genomic DNA of *Legionella* and SdeC protein used in ubiquitination assays.

Chapter 6: Discussion

6.0 Summary of thesis

This thesis provides the first structural characterization of several *Legionella* effectors that interfere with the host through molecular mimicry and modulation of ubiquitination. Chapter one delves into the importance of ubiquitination for *Legionella* intracellular replication. Chapter two presents the structure of AnkB, an effector comprised of two eukaryotic-like domains. Its F-box domain allows association with human E3 Ub ligases, while its Ank domain redirects ubiquitination to pathogen-selected targets. Chapter three highlights the ingenuity of *Legionella*. The structure of AnkC not only shows molecular mimicry via its Ank domain, but also reveals the novel use of Ank domains as dimerization modules. Chapters four and five center on the phosphoribosyl ubiquitination reaction mediated by the SidE family of effectors. Chapter four presents the structures of lpg1496, an effector containing a conserved domain with other SidE family members, and reveals its propensity binding nucleotides. Chapter five focuses on the functional aspect of the phosphoribosyl ubiquitination reaction scheme. Lpg1496 is incapable of partaking of partaking in the processing of ADP-Rib-Ub to a PR-ubiquitinated substrate but can function as a general phosphodiesterase. In addition, the mART and PDE domains of an original SidE family member, can complement each other in trans. Chapter six discusses the implications of the crucial ubiquitination of the *Legionella*-containing vacuole for pathogen survival and links to selective autophagy.

6.1 Ubiquitin and selective autophagy

6.1.1 Xenophagy. One of the first questions that comes to mind regarding ubiquitination of the LCV as a milestone of maturation, is why it promotes bacterial replication but does not

signal for autophagy, in cases where *Legionella* successfully manipulates the host. Xenophagy, or selective autophagy of invading pathogens, is an important host defense mechanism (**Fig. 6.1**). In a healthy host, the response to *Lpn* infection has been linked to an increase in transcription of numerous autophagy related proteins (Farbrother et al., 2006). Stretches of double membrane, termed phagophores, fully engulf bacteria to form the autophagosome. Phagophores are dotted with microtubule-associated protein 1 Light Chain 3 (LC3) proteins, a member of the Atg8 Ub-like protein family, that associate with the membrane through conjugation to phosphatidylethanolamine (PE) (Nguyen et al., 2016). The autophagosome then fuses with the lysosome for degradation (Lamark et al., 2017). Substrates for xenophagy are commonly recognized by autophagy receptors such as p62, that recognize both the forming autophagosome and Ub modified targets. For example, p62 possesses an LIR (LC3-Interacting Region) and a Ub-interacting domain (Lin et al., 2013). The LIR contacts the emerging phagophore, while its Ub-interacting domain contacts the LCV, with a strong preference for Lys63-linked Ub chains (Seibenhener et al., 2004).

Another signal for recruitment of the phagophore is a damaged membrane. In this case, when bacterial division ultimately ruptures the LCV membrane, carbohydrates become exposed to the cytoplasm, recruiting sugar receptors such as Galectin-8 (GAL8). The autophagy receptor Nuclear Domain 10 Protein 52 (NDP52) then acts as a bridge between GAL8 and LC3, forming an autophagosome (Thurston et al., 2012).

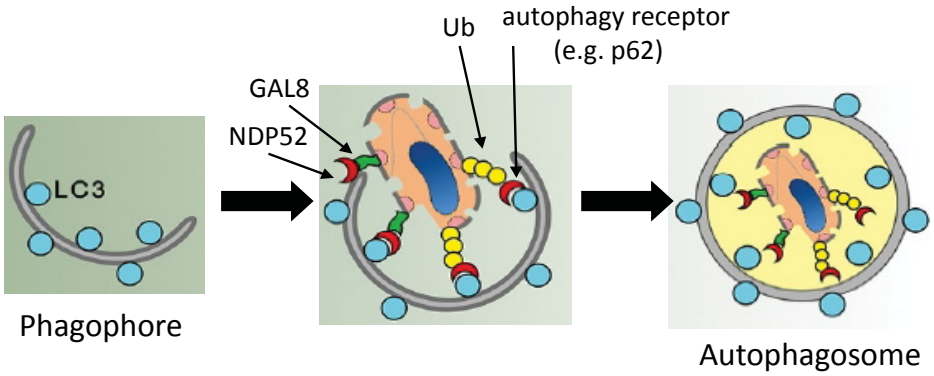


Figure 6.1 Overview of xenophagy. A phagophore decorated with LC3 is recruited the LCV by two signals. Ubiquitinated LCV is detected by autophagy receptors such as p62. Ruptured membranes with exposed carbohydrates are detected by GAL8 and NDP52. The autophagosome is formed by engulfment of the LCV. Adapted from (Kwon and Song, 2018).

6.1.2 Disruption of xenophagy by RavZ. It is not surprising that of the large arsenal of effectors *Lpn* employs, a subset should focus on targeting autophagy. RavZ (region allowing vacuole colocalization gene Z) is the first effector reported to directly interfere with autophagosome formation (Choy et al., 2012). It deconjugates LC3 from the PE on phagophores, removing a binding partner of autophagy receptors. RavZ hydrolyzes the peptide bond between the C-terminal glycine residue and the adjacent residue, resulting in an LC3 that cannot be reconjugated to PE by host proteins (Choy et al., 2012). Furthermore, RavZ contains a unique LIR. A tighter conformation relative to the canonical motif, suggests that RavZ is able to outcompete other proteins, such as autophagy receptors, for interaction with LC3 (Kwon et al., 2017). It is likely other effectors exist that target the same pathway, as a RavZ deletion strain retains the ability to prevent LC3 recruitment to LCVs (Choy et al., 2012).

6.1.3 Modulation of ubiquitin dynamics through AnkB. Although AnkB functions mainly in decorating the LCV with Lys48-linked ubiquitination, and sending ubiquitinated substrates for proteasomal degradation, it may play a role in the broader ubiquitin dynamics around

the LCV (Price et al., 2011; Wong et al., 2017). As mentioned, AnkB is anchored to the LCV through its farnesylation motif and can itself be modified with Lys11-linked ubiquitin by host TRIM21 Ub ligase with no effect on its stability (Bruckert and Abu Kwaik, 2015b). TRIM21 also catalyzes Lys48- and Lys63-linked ubiquitination (Pan et al., 2016). By bringing TRIM21, and potentially other Ub ligases, to the vicinity of the LCV, AnkB may increase the level of ubiquitination, promoting *Legionella* replication.

TRIM21 has also been linked to autophagy. Its overexpression induces the formation of cellular structures resembling autophagosomes, and it interacts with key components of autophagosomes, such as p62 (Pan et al., 2016; Rhodes et al., 2002). TRIM21 ubiquitinates p62 at Lys7 via a Lys63-linkage, which inhibits p62 dimerization (Pan et al., 2016). However, the consequence of this action in xenophagy is unknown.

6.1.4 Potential role of AnkC similar to BCL-3 regulation of NF-κB. The ankyrin domain of AnkC shows the highest structural similarity to BCL-3 with an RMSD of 2.7 Å over 215 Cα atoms (Kozlov et al., 2018). AnkC potentially plays a role in BCL-3 mediated pathways, which includes acting as a negative regulator of NF-κB activation and promoting autophagy in adult T cell leukemia (Wang et al., 2013). BCL-3 is an oncoprotein belonging to the IκB family which modulates the activity of the NF-κB family of transcription factors. NF-κB in turn is a master regulator in the cell, modulating expression of genes involved in apoptosis, proliferation and immune responses, among others (Hoesel and Schmid, 2013). Using a conserved ankyrin domain, the IκB family interacts with NF-κB leading to increased transcription from the κB promoter. BCL-3 causes p50 homodimers to dissociate from the κB site, allowing these inhibiting NF-κB proteins to be replaced by the trans-activating NF-κB proteins, such as p65 (Franzoso et al., 1992). At low dose infections of *Lpn*, the activation of NF-κB has been hypothesized to

contribute to bacterial proliferation by upregulating anti-apoptotic genes (Losick and Isberg, 2006). However, under certain conditions, BCL-3 has been shown to inhibit NF-κB activation and promote autophagy (Wang et al., 2013). Both activation and inhibition of NF-κB have led to both upregulation or downregulation of anti-apoptotic and pro-apoptotic genes (Trocoli and Djavaheri-Mergny, 2011). It is clear that the relationship between NF-κB and autophagy is complex and conflicting. It would be interesting to investigate the potential role of AnkC in BCL-3 pathways, NF-κB pathways, and autophagy.

6.1.5 The SidE family DUB domain antagonizes autophagy. An argument for a role of the SidE family in regulating autophagy is a result of its DUB domain. Although it is capable of cleaving the three most common Ub linkages (Lys11, Lys48, and Lys63), SdeA-DUB has a preference for removal of Lys63-linked chains (Sheedlo et al., 2015). In addition, this DUB domain removes Lys63-linked Ub from LCVs, which likely antagonizes recruitment of autophagy machinery (Sheedlo et al., 2015).

Another interesting point of research would be to test whether the SidE family DUB can recognize the Atg8 family of Ub-like proteins. Four Ub-like protein families exist, and between NEDD8, ISG15 and SUMO, SdeA-DUB forms complexes with NEDD8 and ISG15, but not SUMO (Sheedlo et al., 2015). Its association with the last Ub-like species, Atg8, remains to be tested. Since LC3 is a member of the Atg8 family, whether or not SdeA-DUB interacts and modifies it provides more insight into autophagosome formation.

6.2 Interplay of effectors with free amino acids

The effectors detailed in this thesis were chosen based on their involvement in manipulating host Ub. As mentioned above, these effectors may also play a role in xenophagy.

Furthermore, the function of a few of these effectors, AnkB and the SidE family, intersect at the free amino acids level, highlighting the complexity and diversity of effector functions. It has been established that AnkB provides nutrients for bacterial consumption by increasing the free amino acids pool (Price et al., 2014).

A recent study also places the SidE family in inhibiting the mechanistic target of rapamycin complex 1 (mTORC1), a regulator of host amino acid metabolism (De Leon et al., 2017). mTORC1 responds to the nutrient levels in cells to regulate cell growth. An active mTORC1 represses autophagy and enhances translation (Laplante and Sabatini, 2012). It seems counterintuitive for *Legionella* to both enhance and repress autophagy. However, the SidE proteins block translation without activating mTORC1, by inhibiting Rag small GTPases that are required for mTORC1 to detect levels of amino acids (De Leon et al., 2017). Otherwise the increase in unused free amino acids from both protein synthesis inhibition, and AnkB function, would activate mTORC1.

6.3 Antivirulence therapy

6.3.1 Choosing targets. Although it is tempting to target the secretion system in charge of translocating an army of effectors into a host cell, the T4SS is not only found in the pathogen *Lpn*. Inactivating, for example *dotA*, could also affect beneficial bacteria. Many proteins forming the T4SS also contain sequence and/or structural similarity to other proteins forming other secretion systems (Green and Mecsas, 2016). However, broad spectrum drugs have proved to be effective, as with antibiotics. Pharmaceutical companies are developing compounds that target secretion systems, and many have entered clinical trials (Hauser et al., 2016).

Focusing on effectors themselves still offers the advantage of pathogen specificity. Even though many effectors are redundant and therefore do not result in death of *Legionella* when deleted or inactivated, targeting them would attenuate intracellular replication and provide a better chance for the host for clearance. Other approaches are being tried to improve the efficacy of antivirulence compounds. For example, targeting multiple effectors simultaneously may increase the probability of success. A single human monoclonal antibody capable of neutralizing five virulence factors of *Staphylococcus aureus* has been shown to increase protection in murine models of pneumonia and sepsis (Rouha et al., 2015). Another approach could be combining antivirulence therapy with conventional antibiotics. For example, treatment of pneumonia in a mouse model using both MEDI3902, an agent targeting the type 3 secretion system, and tobramycin, an antibiotic that prevents protein translation, improved survival rates compared to use of either compound alone (DiGiandomenico et al., 2014).

Of the four effectors discussed in this study, targeting AnkB would provide the greatest outcome. It is only one of very few effectors that cause a substantial defect in *Lpn* proliferation when deleted, and is conserved across the sequenced *Lpn* genomes (Al-Khodor et al., 2008; Burstein et al., 2016). Finding a compound that inactivates its ANK domain would prevent AnkB from selecting substrates for degradation and/or coating the LCV with Ub. Targeting the F-box-Skp1 interaction surface would most likely generate side effects detrimental to the host, as its mode of binding is conserved with how host F-boxes interact with Skp1 (Wong et al., 2017).

6.3.2 Future challenges. Although a highly tailored and specific antivirulence therapy is attractive, many challenges remain in its development. Practically, the smaller sales markets of highly specific compounds make them a less attractive option for investment from pharmaceutical companies. In contrast, one of the highly marketed benefit of antivirulence therapy is its lower

selective pressure for resistance relative to antibiotics. However, there is currently no data to support this (Cegelski et al., 2008). To accurately predict this would require a complete understanding of the exact function of a specific virulence factor during infection and in different environments within or outside the host which requires much more research. The conditions of the assays would need to be tailored to each specific pathogen and each clinical strain, as different strains may use a specific effector at a different time during infection, in a different manner, or not express that effector at all. In addition, because this approach is specific to even the strain level, rather than broad spectrum, rapid diagnostic techniques would need to be developed to identify patients (Totsika, 2017).

6.4 Concluding remarks

This thesis is a step toward understanding how the intracellular pathogen, *Legionella pneumophila*, manipulates host ubiquitination. Having the structures of AnkB, AnkC, lpg1496 and eventually the SidE family, provides a platform for structure-based drug design and for further studies into other pathways these effectors may be involved in. The reductionist approach of characterizing the structure of individual domains within a protein, especially in the case of AnkC and lpg1496, was required to overcome crystallization difficulties. To translate these results to the full-length protein and its physiological function would first require determination of the full-length shape and organization of these domains within that space. This can be addressed in future experiments using small angle X-ray scattering, followed by infection assays using deletion strains. Fully understanding their modes of action in multiple pathways will contribute to the development of antivirulence agents for the treatment of Legionnaire's disease.

Reference list

Abu Kwaik, Y., and Bumann, D. (2013). Microbial quest for food in vivo: 'nutritional virulence' as an emerging paradigm. Cell Microbiol *15*, 882-890.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Echols, N., Headd, J.J., Hung, L.W., Jain, S., Kapral, G.J., Grosse Kunstleve, R.W.*, et al.* (2011). The Phenix software for automated determination of macromolecular structures. Methods (San Diego, Calif.) *55*, 94-106.

Al-Khodor, S., Kalachikov, S., Morozova, I., Price, C.T., and Abu Kwaik, Y. (2009). The PmrA/PmrB two-component system of Legionella pneumophila is a global regulator required for intracellular replication within macrophages and protozoa. Infect Immun *77*, 374-386.

Al-Khodor, S., Price, C.T., Habyarimana, F., Kalia, A., and Abu Kwaik, Y. (2008). A Dot/Icm-translocated ankyrin protein of Legionella pneumophila is required for intracellular proliferation within human macrophages and protozoa. Mol Microbiol *70*, 908-923.

Al-Khodor, S., Price, C.T., Kalia, A., and Abu Kwaik, Y. (2010). Functional diversity of ankyrin repeats in microbial proteins. Trends Microbiol *18*, 132-139.

Al-Quadan, T., Price, C.T., London, N., Schueler-Furman, O., and AbuKwaik, Y. (2011). Anchoring of bacterial effectors to host membranes through host-mediated lipidation by prenylation: a common paradigm. Trends Microbiol *19*, 573-579.

Allombert, J., Fuche, F., Michard, C., and Doublet, P. (2013). Molecular mimicry and original biochemical strategies for the biogenesis of a Legionella pneumophila replicative niche in phagocytic cells. Microbes Infect *15*, 981-988.

Amerik, A.Y., and Hochstrasser, M. (2004). Mechanism and function of deubiquitinating enzymes. Biochim Biophys Acta *1695*, 189-207.

Aravind, L., and Koonin, E.V. (1998). The HD domain defines a new superfamily of metal-dependent phosphohydrolases. Trends in biochemical sciences *23*, 469-472.

Aravind, L., and Koonin, E.V. (2000). The U box is a modified RING finger - a common domain in ubiquitination. Curr Biol *10*, R132-134.

Ashida, H., Kim, M., and Sasakawa, C. (2014). Exploitation of the host ubiquitin system by human bacterial pathogens. Nature reviews. Microbiology *12*, 399-413.

Bardill, J.P., Miller, J.L., and Vogel, J.P. (2005). IcmS-dependent translocation of SdeA into macrophages by the Legionella pneumophila type IV secretion system. Mol Microbiol *56*, 90-103.

Bhogaraju, S., Kalayil, S., Liu, Y., Bonn, F., Colby, T., Matic, I., and Dikic, I. (2016). Phosphoribosylation of Ubiquitin Promotes Serine Ubiquitination and Impairs Conventional Ubiquitination. Cell *167*, 1636-1649 e1613.

Boamah, D.K., Zhou, G., Ensminger, A.W., and O'Connor, T.J. (2017). From Many Hosts, One Accidental Pathogen: The Diverse Protozoan Hosts of Legionella. Front Cell Infect Microbiol *7*, 477.

Bruckert, W.M., and Abu Kwaik, Y. (2015a). Complete and ubiquitinated proteome of the Legionella-containing vacuole within human macrophages. J Proteome Res *14*, 236-248.

Bruckert, W.M., and Abu Kwaik, Y. (2015b). Lysine11-linked polyubiquitination of the AnkB F-Box effector of Legionella pneumophila. Infect Immun *84*, 99-107.

Bruckert, W.M., Price, C.T., and Abu Kwaik, Y. (2014). Rapid nutritional remodeling of the host cell upon attachment of Legionella pneumophila. Infect Immun *82*, 72-82.

Burstein, D., Amaro, F., Zusman, T., Lifshitz, Z., Cohen, O., Gilbert, J.A., Pupko, T., Shuman, H.A., and Segal, G. (2016). Genomic analysis of 38 Legionella species identifies large and diverse effector repertoires. Nat Genet *48*, 167-175.

Burstein, D., Zusman, T., Degtyar, E., Viner, R., Segal, G., and Pupko, T. (2009). Genome-scale identification of Legionella pneumophila effectors using a machine learning approach. PLoS Pathog *5*, e1000508.

Cadwell, K., and Coscoy, L. (2005). Ubiquitination on nonlysine residues by a viral E3 ubiquitin ligase. Science *309*, 127-130.

Campanacci, V., Mukherjee, S., Roy, C.R., and Cherfils, J. (2013). Structure of the Legionella effector AnkX reveals the mechanism of phosphocholine transfer by the FIC domain. EMBO J *32*, 1469-1477.

Cazalet, C., Rusniok, C., Bruggemann, H., Zidane, N., Magnier, A., Ma, L., Tichit, M., Jarraud, S., Bouchier, C., Vandenesch, F*., et al.* (2004). Evidence in the Legionella pneumophila genome for exploitation of host cell functions and high genome plasticity. Nat Genet *36*, 1165-1173.

Cegelski, L., Marshall, G.R., Eldridge, G.R., and Hultgren, S.J. (2008). The biology and future prospects of antivirulence therapies. Nature reviews. Microbiology *6*, 17-27.

Chakravarthy, S., Huot, B., and Kvitko, B.H. (2017). Effector Translocation: Cya Reporter Assay. Methods in molecular biology (Clifton, N.J.) *1615*, 473-487.

Chandran Darbari, V., and Waksman, G. (2015). Structural Biology of Bacterial Type IV Secretion Systems. Annu Rev Biochem *84*, 603-629.

Chang, B., Kura, F., Amemura-Maekawa, J., Koizumi, N., and Watanabe, H. (2005). Identification of a novel adhesion molecule involved in the virulence of Legionella pneumophila. Infect Immun *73*, 4272-4280.

Chen, H.D., and Groisman, E.A. (2013). The biology of the PmrA/PmrB two-component system: the major regulator of lipopolysaccharide modifications. Annu Rev Microbiol *67*, 83-112.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr *66*, 12-21.

Chien, M., Morozova, I., Shi, S., Sheng, H., Chen, J., Gomez, S.M., Asamani, G., Hill, K., Nuara, J., Feder, M.*, et al.* (2004). The genomic sequence of the accidental pathogen Legionella pneumophila. Science *305*, 1966-1968.

Choy, A., Dancourt, J., Mugo, B., O'Connor, T.J., Isberg, R.R., Melia, T.J., and Roy, C.R. (2012). The Legionella effector RavZ inhibits host autophagy through irreversible Atg8 deconjugation. Science *338*, 1072-1076.

Clatworthy, A.E., Pierson, E., and Hung, D.T. (2007). Targeting virulence: a new paradigm for antimicrobial therapy. Nat Chem Biol *3*, 541-548.

Coers, J., Kagan, J.C., Matthews, M., Nagai, H., Zuckman, D.M., and Roy, C.R. (2000). Identification of Icm protein complexes that play distinct roles in the biogenesis of an organelle permissive for Legionella pneumophila intracellular growth. Mol Microbiol *38*, 719-736.

Conover, G.M., Derre, I., Vogel, J.P., and Isberg, R.R. (2003). The Legionella pneumophila LidA protein: a translocated substrate of the Dot/Icm system associated with maintenance of bacterial integrity. Mol Microbiol *48*, 305-321.

Correia, A.M., Ferreira, J.S., Borges, V., Nunes, A., Gomes, B., Capucho, R., Goncalves, J., Antunes, D.M., Almeida, S., Mendes, A.*, et al.* (2016). Probable Person-to-Person Transmission of Legionnaires' Disease. N Engl J Med *374*, 497-498.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr *62*, 1002-1011.

Cowtan, K. (2010). Recent developments in classical density modification. Acta Crystallogr D Biol Crystallogr *66*, 470-478.

Cunha, C.B., and Cunha, B.A. (2017). Legionnaire's Disease Since Philadelphia: Lessons Learned and Continued Progress. Infect Dis Clin North Am *31*, 1-5.

de Felipe, K.S., Glover, R.T., Charpentier, X., Anderson, O.R., Reyes, M., Pericone, C.D., and Shuman, H.A. (2008). Legionella eukaryotic-like type IV substrates interfere with organelle trafficking. PLoS Pathog *4*, e1000117.

de Felipe, K.S., Pampou, S., Jovanovic, O.S., Pericone, C.D., Ye, S.F., Kalachikov, S., and Shuman, H.A. (2005). Evidence for acquisition of Legionella type IV secretion substrates via interdomain horizontal gene transfer. Journal of bacteriology *187*, 7716-7726.

de Jong, A., Merkx, R., Berlin, I., Rodenko, B., Wijdeven, R.H., El Atmioui, D., Yalcin, Z., Robson, C.N., Neefjes, J.J., and Ovaa, H. (2012). Ubiquitin-based probes prepared by total synthesis to profile the activity of deubiquitinating enzymes. Chembiochem *13*, 2251-2258.

De Leon, J.A., Qiu, J., Nicolai, C.J., Counihan, J.L., Barry, K.C., Xu, L., Lawrence, R.E., Castellano, B.M., Zoncu, R., Nomura, D.K.*, et al.* (2017). Positive and Negative Regulation of the Master Metabolic Regulator mTORC1 by Two Families of Legionella pneumophila Effectors. Cell Rep *21*, 2031-2038.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR *6*, 277-293.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M.*, et al.* (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res *36*, W465-469.

Descours, G., Ginevra, C., Jacotin, N., Forey, F., Chastang, J., Kay, E., Etienne, J., Lina, G., Doublet, P., and Jarraud, S. (2017). Ribosomal Mutations Conferring Macrolide Resistance in Legionella pneumophila. Antimicrob Agents Chemother *61*.

DiGiandomenico, A., Keller, A.E., Gao, C., Rainey, G.J., Warrener, P., Camara, M.M., Bonnell, J., Fleming, R., Bezabeh, B., Dimasi, N.*, et al.* (2014). A multifunctional bispecific antibody protects against Pseudomonas aeruginosa. Sci Transl Med *6*, 262ra155.

Dikic, I., Wakatsuki, S., and Walters, K.J. (2009). Ubiquitin-binding domains - from structures to functions. Nat Rev Mol Cell Biol *10*, 659-671.

Dorer, M.S., Kirton, D., Bader, J.S., and Isberg, R.R. (2006). RNA interference analysis of Legionella in Drosophila cells: exploitation of early secretory apparatus dynamics. PLoS Pathog *2*, e34.

Doyle, R.M., and Heuzenroeder, M.W. (2002). A mutation in an ompR-like gene on a Legionella longbeachae serogroup 1 plasmid attenuates virulence. Int J Med Microbiol *292*, 227-239.

Eisen, J.A., Coyne, R.S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J.R., Badger, J.H., Ren, Q., Amedeo, P., Jones, K.M.*, et al.* (2006). Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. PLoS Biol *4*, e286.

Eisenreich, W., Heesemann, J., Rudel, T., and Goebel, W. (2013). Metabolic host responses to infection by intracellular bacterial pathogens. Front Cell Infect Microbiol *3*, 24.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.

Ensminger, A.W., and Isberg, R.R. (2010). E3 ubiquitin ligase activity and targeting of BAT3 by multiple Legionella pneumophila translocated substrates. Infect Immun *78*, 3905-3919.

Faggiano, S., Menon, R.P., Kelly, G.P., Todi, S.V., Scaglione, K.M., Konarev, P.V., Svergun, D.I., Paulson, H.L., and Pastore, A. (2015). Allosteric regulation of deubiquitylase activity through ubiquitination. Front Mol Biosci *2*, 2.

Farbrother, P., Wagner, C., Na, J., Tunggal, B., Morio, T., Urushihara, H., Tanaka, Y., Schleicher, M., Steinert, M., and Eichinger, L. (2006). Dictyostelium transcriptional host cell response upon infection with Legionella. Cell Microbiol *8*, 438-456.

Fields, B.S., Benson, R.F., and Besser, R.E. (2002). Legionella and Legionnaires' disease: 25 years of investigation. Clin Microbiol Rev *15*, 506-526.

Finley, D. (2009). Recognition and processing of ubiquitin-protein conjugates by the proteasome. Annu Rev Biochem *78*, 477-513.

Francis, S.H., Houslay, M.D., and Conti, M. (2011). Phosphodiesterase inhibitors: factors that influence potency, selectivity, and action. Handb Exp Pharmacol, 47-84.

Franzoso, G., Bours, V., Park, S., Tomita-Yamaguchi, M., Kelly, K., and Siebenlist, U. (1992). The candidate oncoprotein Bcl-3 is an antagonist of p50/NF-kappa B-mediated inhibition. Nature *359*, 339-342.

Fraser, D.W., Tsai, T.R., Orenstein, W., Parkin, W.E., Beecham, H.J., Sharrar, R.G., Harris, J., Mallison, G.F., Martin, S.M., McDade, J.E.*, et al.* (1977). Legionnaires' disease: description of an epidemic of pneumonia. N Engl J Med *297*, 1189-1197.

Ghosal, D., Chang, Y.W., Jeong, K.C., Vogel, J.P., and Jensen, G.J. (2017). In situ structure of the Legionella Dot/Icm type IV secretion system by electron cryotomography. EMBO Rep *18*, 726-732.

Glockner, G., Albert-Weissenberger, C., Weinmann, E., Jacobi, S., Schunder, E., Steinert, M., Hacker, J., and Heuner, K. (2008). Identification and characterization of a new conjugation/type IVA secretion system (trb/tra) of Legionella pneumophila Corby localized on two mobile genomic islands. Int J Med Microbiol *298*, 411-428.

Goddard, T.D., and Kneller, D.G. SPARKY 3 (University of California, San Francisco).

Green, E.R., and Mecsas, J. (2016). Bacterial Secretion Systems: An Overview. Microbiol Spectr *4*.

Habyarimana, F., Al-Khodor, S., Kalia, A., Graham, J.E., Price, C.T., Garcia, M.T., and Kwaik, Y.A. (2008). Role for the Ankyrin eukaryotic-like genes of Legionella pneumophila in parasitism of protozoan hosts and human macrophages. Environ Microbiol *10*, 1460-1474.

Han, S., Craig, J.A., Putnam, C.D., Carozzi, N.B., and Tainer, J.A. (1999). Evolution and mechanism from structures of an ADP-ribosylating toxin and NAD complex. Nat Struct Biol *6*, 932-936.

Harrigan, J.A., Jacq, X., Martin, N.M., and Jackson, S.P. (2018). Deubiquitylating enzymes and drug discovery: emerging opportunities. Nat Rev Drug Discov *17*, 57-78.

Harvey, K.F., and Kumar, S. (1999). Nedd4-like proteins: an emerging family of ubiquitin-protein ligases implicated in diverse cellular functions. Trends Cell Biol *9*, 166-169.

Hauser, A.R., Mecsas, J., and Moir, D.T. (2016). Beyond Antibiotics: New Therapeutic Approaches for Bacterial Infections. Clin Infect Dis *63*, 89-95.

Havey, J.C., and Roy, C.R. (2015). Toxicity and SidJ-Mediated Suppression of Toxicity Require Distinct Regions in the SidE Family of Legionella pneumophila Effectors. Infect Immun *83*, 3506-3514.

Heidtman, M., Chen, E.J., Moy, M.Y., and Isberg, R.R. (2009). Large-scale identification of Legionella pneumophila Dot/Icm substrates that modulate host cell vesicle trafficking pathways. Cell Microbiol *11*, 230-248.

Herbert, M.H., Squire, C.J., and Mercer, A.A. (2015). Poxviral ankyrin proteins. Viruses *7*, 709-738.

Hershko, A., Heller, H., Elias, S., and Ciechanover, A. (1983). Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. J Biol Chem *258*, 8206-8214.

Hershko, A., Heller, H., Eytan, E., and Reiss, Y. (1986). The protein substrate binding site of the ubiquitin-protein ligase system. J Biol Chem *261*, 11992-11999.

Hoesel, B., and Schmid, J.A. (2013). The complexity of NF-kappaB signaling in inflammation and cancer. Mol Cancer *12*, 86.

Holm, L., and Rosenström, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res *38*, W545-549.

Horwitz, M.A. (1983). Formation of a novel phagosome by the Legionnaires' disease bacterium (Legionella pneumophila) in human monocytes. J Exp Med *158*, 1319-1331.

Horwitz, M.A., and Maxfield, F.R. (1984). Legionella pneumophila inhibits acidification of its phagosome in human monocytes. J Cell Biol *99*, 1936-1943.

Hostetter, M.K., Tao, N.J., Gale, C., Herman, D.J., McClellan, M., Sharp, R.L., and Kendrick, K.E. (1995). Antigenic and functional conservation of an integrin I-domain in Saccharomyces cerevisiae. Biochem Mol Med *55*, 122-130.

Hubber, A., Kubori, T., and Nagai, H. (2013). Modulation of the ubiquitination machinery by Legionella. Current topics in microbiology and immunology *376*, 227-247.

Huibregtse, J.M., Scheffner, M., Beaudenon, S., and Howley, P.M. (1995). A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. Proc Natl Acad Sci U S A *92*, 5249.

Ivanov, S.S., Charron, G., Hang, H.C., and Roy, C.R. (2010). Lipidation by the host prenyltransferase machinery facilitates membrane localization of Legionella pneumophila effector proteins. J Biol Chem *285*, 34686-34698.

Ivanov, S.S., and Roy, C.R. (2009). Modulation of ubiquitin dynamics and suppression of DALIS formation by the Legionella pneumophila Dot/Icm system. Cell Microbiol *11*, 261-278.

Jelic, D., and Antolovic, R. (2016). From Erythromycin to Azithromycin and New Potential Ribosome-Binding Antimicrobials. Antibiotics (Basel) *5*.

Jeong, K.C., Sexton, J.A., and Vogel, J.P. (2015). Spatiotemporal regulation of a Legionella pneumophila T4SS substrate by the metaeffector SidJ. PLoS Pathog *11*, e1004695.

Kawakami, T., Chiba, T., Suzuki, T., Iwai, K., Yamanaka, K., Minato, N., Suzuki, H., Shimbara, N., Hidaka, Y., Osaka, F.*, et al.* (2001). NEDD8 recruits E2-ubiquitin to SCF E3 ligase. EMBO J *20*, 4003-4012.

Keusekotten, K., Elliott, P.R., Glockner, L., Fiil, B.K., Damgaard, R.B., Kulathu, Y., Wauer, T., Hospenthal, M.K., Gyrd-Hansen, M., Krappmann, D.*, et al.* (2013). OTULIN antagonizes LUBAC signaling by specifically hydrolyzing Met1-linked polyubiquitin. Cell *153*, 1312-1326.

Kipreos, E.T., and Pagano, M. (2000). The F-box protein family. Genome Biol *1*, REVIEWS3002.

Kohanski, M.A., Dwyer, D.J., and Collins, J.J. (2010). How antibiotics kill bacteria: from targets to networks. Nature reviews. Microbiology *8*, 423-435.

Kotewicz, K.M., Ramabhadran, V., Sjoblom, N., Vogel, J.P., Haenssler, E., Zhang, M., Behringer, J., Scheck, R.A., and Isberg, R.R. (2017). A Single Legionella Effector Catalyzes a Multistep Ubiquitination Pathway to Rearrange Tubular Endoplasmic Reticulum for Replication. Cell Host Microbe *21*, 169-181.

Kozlov, G., Wong, K., Wang, W., Skubak, P., Munoz-Escobar, J., Liu, Y., Siddiqui, N., Pannu, N.S., and Gehring, K. (2018). Ankyrin repeats as a dimerization module. Biochem Biophys Res Commun *495*, 1002-1007.

Kubori, T., Hyakutake, A., and Nagai, H. (2008). Legionella translocates an E3 ubiquitin ligase that has multiple U-boxes with distinct functions. Mol Microbiol *67*, 1307-1319.

Kubori, T., Shinzawa, N., Kanuka, H., and Nagai, H. (2010). Legionella metaeffector exploits host proteasome to temporally regulate cognate effector. PLoS Pathog *6*, e1001216.

Kurinov, I.V., Rajamohan, F., and Uckun, F.M. (2004). High resolution X-ray structure and potent anti-HIV activity of recombinant dianthin antiviral protein. Arzneimittelforschung *54*, 692-702.

Kwon, D.H., Kim, L., Kim, B.W., Kim, J.H., Roh, K.H., Choi, E.J., and Song, H.K. (2017). A novel conformation of the LC3-interacting region motif revealed by the structure of a complex between LC3B and RavZ. Biochem Biophys Res Commun *490*, 1093-1099.

Kwon, D.H., and Song, H.K. (2018). A Structural View of Xenophagy, a Battle between Host and Microbes. Mol Cells 41*, 27-34*.

Lamark, T., Svenning, S., and Johansen, T. (2017). Regulation of selective autophagy: the p62/SQSTM1 paradigm. Essays Biochem *61*, 609-624.

Langer, G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc *3*, 1171-1179.

Laplante, M., and Sabatini, D.M. (2012). mTOR signaling in growth control and disease. Cell *149*, 274-293.

Laskowski, R.A.M., M.W.; Moss, D.S.; Thornton, J.M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. J. Appl. Cryst. *26*, 283-291.

Legate, K.R., Montanez, E., Kudlacek, O., and Fassler, R. (2006). ILK, PINCH and parvin: the tIPP of integrin signalling. Nat Rev Mol Cell Biol *7*, 20-31.

Li, J., Mahajan, A., and Tsai, M.D. (2006). Ankyrin repeat: a unique motif mediating protein-protein interactions. Biochemistry *45*, 15168-15178.

Li, W., and Ye, Y. (2008). Polyubiquitin chains: functions, structures, and mechanisms. Cell Mol Life Sci *65*, 2397-2406.

Lin, X., Li, S., Zhao, Y., Ma, X., Zhang, K., He, X., and Wang, Z. (2013). Interaction domains of p62: a bridge between p62 and selective autophagy. DNA Cell Biol *32*, 220-227.

Lomma, M., Dervins-Ravault, D., Rolando, M., Nora, T., Newton, H.J., Sansom, F.M., Sahr, T., Gomez-Valero, L., Jules, M., Hartland, E.L.*, et al.* (2010). The Legionella pneumophila F-box protein Lpp2082 (AnkB) modulates ubiquitination of the host protein parvin B and promotes intracellular replication. Cell Microbiol *12*, 1272-1291.

Losick, V.P., Haenssler, E., Moy, M.Y., and Isberg, R.R. (2010). LnaB: a Legionella pneumophila activator of NF-kappaB. Cell Microbiol *12*, 1083-1097.

Losick, V.P., and Isberg, R.R. (2006). NF-kappaB translocation prevents host cell death after low-dose challenge by Legionella pneumophila. J Exp Med *203*, 2177-2189.

Low, H.H., Gubellini, F., Rivera-Calzada, A., Braun, N., Connery, S., Dujeancourt, A., Lu, F., Redzej, A., Fronzes, R., Orlova, E.V.*, et al.* (2014). Structure of a type IV secretion system. Nature *508*, 550-553.

Luo, Z.Q., and Isberg, R.R. (2004). Multiple substrates of the Legionella pneumophila Dot/Icm system identified by interbacterial protein transfer. Proc Natl Acad Sci U S A *101*, 841-846.

Lyapina, S.A., Correll, C.C., Kipreos, E.T., and Deshaies, R.J. (1998). Human CUL1 forms an evolutionarily conserved ubiquitin ligase complex (SCF) with SKP1 and an F-box protein. Proc Natl Acad Sci U S A *95*, 7451-7456.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

McDade, J.E., Shepard, C.C., Fraser, D.W., Tsai, T.R., Redus, M.A., and Dowdle, W.R. (1977). Legionnaires' disease: isolation of a bacterium and demonstration of its role in other respiratory disease. N Engl J Med *297*, 1197-1203.

McDowell, G.S., and Philpott, A. (2013). Non-canonical ubiquitylation: mechanisms and consequences. Int J Biochem Cell Biol *45*, 1833-1842.

Minor, Z.O.a.W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology *276*, 307-326.

Morreale, F.E., and Walden, H. (2016). Types of Ubiquitin Ligases. Cell *165*, 248-248 e241.

Morrison, D.C., and Ryan, J.L. (1987). Endotoxins and disease mechanisms. Annu Rev Med *38*, 417-432.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr *67*, 355-367.

Nagai, H., Cambronne, E.D., Kagan, J.C., Amor, J.C., Kahn, R.A., and Roy, C.R. (2005). A C-terminal translocation signal required for Dot/Icm-dependent delivery of the Legionella RalF protein to host cells. Proc Natl Acad Sci U S A *102*, 826-831.

Nagai, H., Kagan, J.C., Zhu, X., Kahn, R.A., and Roy, C.R. (2002). A bacterial guanine nucleotide exchange factor activates ARF on Legionella phagosomes. Science *295*, 679-682.

Nagai, H., and Kubori, T. (2011). Type IVB Secretion Systems of Legionella and Other Gram-Negative Bacteria. Front Microbiol *2*, 136.

Nakayama, K.I., and Nakayama, K. (2005). Regulation of the cell cycle by SCF-type ubiquitin ligases. Semin Cell Dev Biol *16*, 323-333.

Nguyen, T.N., Padman, B.S., Usher, J., Oorschot, V., Ramm, G., and Lazarou, M. (2016). Atg8 family LC3/GABARAP proteins are crucial for autophagosome-lysosome fusion but not autophagosome formation during PINK1/Parkin mitophagy and starvation. J Cell Biol *215*, 857-874.

Ninio, S., Zuckman-Cholon, D.M., Cambronne, E.D., and Roy, C.R. (2005). The Legionella IcmS-IcmW protein complex is important for Dot/Icm-mediated protein translocation. Mol Microbiol *55*, 912-926.

Otwinowski Z, M.W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol *276*, 307–326.

Pagan, J., Seto, T., Pagano, M., and Cittadini, A. (2013). Role of the ubiquitin proteasome system in the heart. Circ Res *112*, 1046-1058.

Pan, J.A., Sun, Y., Jiang, Y.P., Bott, A.J., Jaber, N., Dou, Z., Yang, B., Chen, J.S., Catanzaro, J.M., Du, C.*, et al.* (2016). TRIM21 Ubiquitylates SQSTM1/p62 and Suppresses Protein Sequestration to Regulate Redox Homeostasis. Mol Cell *61*, 720-733.

Pan, X., Luhrmann, A., Satoh, A., Laskowski-Arce, M.A., and Roy, C.R. (2008). Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. Science *320*, 1651-1654.

Parra, R.G., Espada, R., Verstraete, N., and Ferreiro, D.U. (2015). Structural and Energetic Characterization of the Ankyrin Repeat Protein Family. PLoS Comput Biol *11*, e1004659.

Phillips, A.H., Zhang, Y., Cunningham, C.N., Zhou, L., Forrest, W.F., Liu, P.S., Steffek, M., Lee, J., Tam, C., Helgason, E.*, et al.* (2013). Conformational dynamics control ubiquitin-deubiquitinase interactions and influence in vivo signaling. Proc Natl Acad Sci U S A *110*, 11379-11384.

Phin, N., Parry-Ford, F., Harrison, T., Stagg, H.R., Zhang, N., Kumar, K., Lortholary, O., Zumla, A., and Abubakar, I. (2014). Epidemiology and clinical management of Legionnaires' disease. Lancet Infect Dis *14*, 1011-1021.

Pickart, C.M. (2001). Mechanisms underlying ubiquitination. Annu Rev Biochem *70*, 503-533.

Pickart, C.M., and Eddins, M.J. (2004). Ubiquitin: structures, functions, mechanisms. Biochim Biophys Acta *1695*, 55-72.

Pine, L., George, J.R., Reeves, M.W., and Harrell, W.K. (1979). Development of a chemically defined liquid medium for growth of Legionella pneumophila. J Clin Microbiol *9*, 615-626.

Pinotsis, N., and Waksman, G. (2017). Structure of the WipA protein reveals a novel tyrosine protein phosphatase effector from Legionella pneumophila. J Biol Chem *292*, 9240-9251.

Price, C.T., Al-Khodor, S., Al-Quadan, T., and Abu Kwaik, Y. (2010a). Indispensable role for the eukaryotic-like ankyrin domains of the ankyrin B effector of Legionella pneumophila within macrophages and amoebae. Infect Immun *78*, 2079-2088.

Price, C.T., Al-Khodor, S., Al-Quadan, T., Santic, M., Habyarimana, F., Kalia, A., and Kwaik, Y.A. (2009). Molecular mimicry by an F-box effector of Legionella pneumophila hijacks a conserved polyubiquitination machinery within macrophages and protozoa. PLoS Pathog *5*, e1000704.

Price, C.T., Al-Quadan, T., Santic, M., Jones, S.C., and Abu Kwaik, Y. (2010b). Exploitation of conserved eukaryotic host cell farnesylation machinery by an F-box effector of Legionella pneumophila. J Exp Med *207*, 1713-1726.

Price, C.T., Al-Quadan, T., Santic, M., Rosenshine, I., and Abu Kwaik, Y. (2011). Host proteasomal degradation generates amino acids essential for intracellular bacterial growth. Science *334*, 1553-1557.

Price, C.T., and Kwaik, Y.A. (2010). Exploitation of Host Polyubiquitination Machinery through Molecular Mimicry by Eukaryotic-Like Bacterial F-Box Effectors. Front Microbiol *1*, 122.

Price, C.T., Richards, A.M., and Abu Kwaik, Y. (2014). Nutrient generation and retrieval from the host cell cytosol by intra-vacuolar Legionella pneumophila. Front Cell Infect Microbiol *4*, 111.

Puvar, K., Zhou, Y., Qiu, J., Luo, Z.Q., Wirth, M.J., and Das, C. (2017). Ubiquitin Chains Modified by the Bacterial Ligase SdeA Are Protected from Deubiquitinase Hydrolysis. Biochemistry *56*, 4762-4766.

Qiu, J., Sheedlo, M.J., Yu, K., Tan, Y., Nakayasu, E.S., Das, C., Liu, X., and Luo, Z.Q. (2016). Ubiquitination independent of E1 and E2 enzymes by bacterial effectors. Nature *533*, 120-124.

Qiu, J., Yu, K., Fei, X., Liu, Y., Nakayasu, E.S., Piehowski, P.D., Shaw, J.B., Puvar, K., Das, C., Liu, X.*, et al.* (2017). A unique deubiquitinase that deconjugates phosphoribosyl-linked protein ubiquitination. Cell Res *27*, 865-881.

Quaile, A.T., Urbanus, M.L., Stogios, P.J., Nocek, B., Skarina, T., Ensminger, A.W., and Savchenko, A. (2015). Molecular Characterization of LubX: Functional Divergence of the U-Box Fold by Legionella pneumophila. Structure *23*, 1459-1469.

Reingold, A.L., Thomason, B.M., Brake, B.J., Thacker, L., Wilkinson, H.W., and Kuritsky, J.N. (1984). Legionella pneumonia in the United States: the distribution of serogroups and species causing human illness. J Infect Dis *149*, 819.

Rhodes, D.A., Ihrke, G., Reinicke, A.T., Malcherek, G., Towey, M., Isenberg, D.A., and Trowsdale, J. (2002). The 52 000 MW Ro/SS-A autoantigen in Sjogren's syndrome/systemic lupus erythematosus (Ro52) is an interferon-gamma inducible tripartite motif protein associated with membrane proximal structures. Immunology *106*, 246-256.

Ritchie, K.J., Malakhov, M.P., Hetherington, C.J., Zhou, L., Little, M.T., Malakhova, O.A., Sipe, J.C., Orkin, S.H., and Zhang, D.E. (2002). Dysregulation of protein modification by ISG15 results in brain cell injury. Genes Dev *16*, 2207-2212.

Robertson, P., Abdelhady, H., and Garduno, R.A. (2014). The many forms of a pleomorphic bacterial pathogen-the developmental network of Legionella pneumophila. Front Microbiol *5*, 670.

Roig, J., Sabria, M., and Pedro-Botet, M.L. (2003). Legionella spp.: community acquired and nosocomial infections. Curr Opin Infect Dis *16*, 145-151.

Rosenstreich, D.L., Glode, L.M., and Mergenhagen, S.E. (1977). Action of endotoxin on lymphoid cells. J Infect Dis *136 Suppl*, S239-245.

Rouha, H., Badarau, A., Visram, Z.C., Battles, M.B., Prinz, B., Magyarics, Z., Nagy, G., Mirkina, I., Stulik, L., Zerbs, M.*, et al.* (2015). Five birds, one stone: neutralization of alpha-hemolysin and 4 bi-component leukocidins of Staphylococcus aureus with a single human monoclonal antibody. MAbs *7*, 243-254.

Scheffner, M., Nuber, U., and Huibregtse, J.M. (1995). Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. Nature *373*, 81-83.

Schulman, B.A., Carrano, A.C., Jeffrey, P.D., Bowen, Z., Kinnucan, E.R., Finnin, M.S., Elledge, S.J., Harper, J.W., Pagano, M., and Pavletich, N.P. (2000). Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. Nature *408*, 381-386.

Scott, C.C., Botelho, R.J., and Grinstein, S. (2003). Phagosome maturation: a few bugs in the system. J Membr Biol *193*, 137-152.

Segal, G., and Shuman, H.A. (1999). Legionella pneumophila utilizes the same genes to multiply within Acanthamoeba castellanii and human macrophages. Infect Immun *67*, 2117-2124.

Seibenhener, M.L., Babu, J.R., Geetha, T., Wong, H.C., Krishna, N.R., and Wooten, M.W. (2004). Sequestosome 1/p62 is a polyubiquitin chain binding protein involved in ubiquitin proteasome degradation. Mol Cell Biol *24*, 8055-8068.

Seirafi, M., Kozlov, G., and Gehring, K. (2015). Parkin structure and function. FEBS J *282*, 2076-2088.

Serezani, C.H., Chung, J., Ballinger, M.N., Moore, B.B., Aronoff, D.M., and Peters-Golden, M. (2007). Prostaglandin E2 suppresses bacterial killing in alveolar macrophages by inhibiting NADPH oxidase. Am J Respir Cell Mol Biol *37*, 562-570.

Sheedlo, M.J., Qiu, J., Tan, Y., Paul, L.N., Luo, Z.Q., and Das, C. (2015). Structural basis of substrate recognition by a bacterial deubiquitinase important for dynamics of phagosome ubiquitination. Proc Natl Acad Sci U S A *112*, 15090-15095.

Sheikh, M.O., Xu, Y., van der Wel, H., Walden, P., Hartson, S.D., and West, C.M. (2015). Glycosylation of Skp1 promotes formation of Skp1-cullin-1-F-box protein complexes in dictyostelium. Mol Cell Proteomics *14*, 66-80.

Sheldrick, G.M. (2008). A short history of SHELX. Acta Crystallogr A *64*, 112-122.

Shin, S. (2012). Innate Immunity to Intracellular Pathogens: Lessons Learned from Legionella pneumophila. Adv Appl Microbiol *79*, 43-71.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J.*, et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol *7*, 539.

Skowyra, D., Craig, K.L., Tyers, M., Elledge, S.J., and Harper, J.W. (1997). F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. Cell *91*, 209-219.

Skubak, P., and Pannu, N.S. (2013). Automatic protein structure solution from weak X-ray data. Nature communications *4*, 2777.

Son, J., Jo, C.H., Murugan, R.N., Bang, J.K., Hwang, K.Y., and Lee, W.C. (2015). Crystal structure of Legionella pneumophila type IV secretion system effector LegAS4. Biochem Biophys Res Commun *465*, 817-824.

Swanson, M.S., and Isberg, R.R. (1995). Association of Legionella pneumophila with the macrophage endoplasmic reticulum. Infect Immun *63*, 3609-3620.

Thrower, J.S., Hoffman, L., Rechsteiner, M., and Pickart, C.M. (2000). Recognition of the polyubiquitin proteolytic signal. EMBO J *19*, 94-102.

Thurston, T.L., Wandel, M.P., von Muhlinen, N., Foeglein, A., and Randow, F. (2012). Galectin 8 targets damaged vesicles for autophagy to defend cells against bacterial invasion. Nature *482*, 414-418.

Totsika, M. (2017). Disarming pathogens: benefits and challenges of antimicrobials that target bacterial virulence instead of growth and viability. Future Med Chem *9*, 267-269.

Trempe, J.F., Sauve, V., Grenier, K., Seirafi, M., Tang, M.Y., Menade, M., Al-Abdul-Wahid, S., Krett, J., Wong, K., Kozlov, G.*, et al.* (2013). Structure of parkin reveals mechanisms for ubiquitin ligase activation. Science *340*, 1451-1455.

Trocoli, A., and Djavaheri-Mergny, M. (2011). The complex interplay between autophagy and NF-kappaB signaling pathways in cancer cells. Am J Cancer Res *1*, 629-649.

Uro-Coste, E., Perret, E., Fonta, C., Mathieu, M., Delisle, M.B., Caput, D., and Imbert, M. (1998). The cell cycle gene SKP1 is regulated by light in postnatal rat brain. Brain Res Mol Brain Res *56*, 192-199.

Valsdottir, R., Hashimoto, H., Ashman, K., Koda, T., Storrie, B., and Nilsson, T. (2001). Identification of rabaptin-5, rabex-5, and GM130 as putative effectors of rab33b, a regulator of retrograde traffic between the Golgi apparatus and ER. FEBS Lett *508*, 201-209.

Vogel, J.P., Andrews, H.L., Wong, S.K., and Isberg, R.R. (1998). Conjugative transfer by the virulence system of Legionella pneumophila. Science *279*, 873-876.

Wahabi, K., Perwez, A., and Rizvi, M.A. (2018). Parkin in Parkinson's Disease and Cancer: a Double-Edged Sword. Mol Neurobiol.

Wang, H., Yan, Z., Geng, J., Kunz, S., Seebeck, T., and Ke, H. (2007a). Crystal structure of the Leishmania major phosphodiesterase LmjPDEB1 and insight into the design of the parasite-selective inhibitors. Mol Microbiol *66*, 1029-1038.

Wang, J., Niu, Z., Shi, Y., Gao, C., Wang, X., Han, J., Li, J., Gao, Z., Zhu, X., Song, X.*, et al.* (2013). Bcl-3, induced by Tax and HTLV-1, inhibits NF-kappaB activation and promotes autophagy. Cell Signal *25*, 2797-2804.

Wang, X., Herr, R.A., Chua, W.J., Lybarger, L., Wiertz, E.J., and Hansen, T.H. (2007b). Ubiquitination of serine, threonine, or lysine residues on the cytoplasmic tail can induce ERAD of MHC-I by viral E3 ligase mK3. J Cell Biol *177*, 613-624.

Wehrli, W. (1983). Rifampin: mechanisms of action and resistance. Rev Infect Dis *5 Suppl 3*, S407-411.

Wenzel, D.M., Lissounov, A., Brzovic, P.S., and Klevit, R.E. (2011). UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids. Nature *474*, 105-108.

Wickliffe, K.E., Williamson, A., Meyer, H.J., Kelly, A., and Rape, M. (2011). K11-linked ubiquitin chains as novel regulators of cell division. Trends Cell Biol *21*, 656-663.

Winget, J.M., and Mayor, T. (2010). The diversity of ubiquitin recognition: hot spots and varied specificity. Mol Cell *38*, 627-635.

Wong, K., Kozlov, G., Zhang, Y., and Gehring, K. (2015). Structure of the Legionella Effector, lpg1496, Suggests a Role in Nucleotide Metabolism. J Biol Chem *290*, 24727-24737.

Wong, K., Perpich, J.D., Kozlov, G., Cygler, M., Abu Kwaik, Y., and Gehring, K. (2017). Structural Mimicry by a Bacterial F Box Effector Hijacks the Host Ubiquitin-Proteasome System. Structure *25*, 376-383.

World Health Organization (2014). Antimicrobial resistance : global report on surveillance (Geneva, Switzerland: World Health Organization).

Xirodimas, D.P., Saville, M.K., Bourdon, J.C., Hay, R.T., and Lane, D.P. (2004). Mdm2-mediated NEDD8 conjugation of p53 inhibits its transcriptional activity. Cell *118*, 83-97.

Xu, C., Jin, J., Bian, C., Lam, R., Tian, R., Weist, R., You, L., Nie, J., Bochkarev, A., Tempel, W., *et al.* (2012). Sequence-specific recognition of a PxLPxI/L motif by an ankyrin repeat tumbler lock. Science signaling *5*, ra39.

Xu, L., and Luo, Z.Q. (2013). Cell biology of infection by Legionella pneumophila. Microbes Infect *15*, 157-167.

Zaman, S.B., Hussain, M.A., Nye, R., Mehta, V., Mamun, K.T., and Hossain, N. (2017). A Review on Antibiotic Resistance: Alarm Bells are Ringing. Cureus *9*, e1403.

Zhang, L., Zhou, F., Drabsch, Y., Gao, R., Snaar-Jagalska, B.E., Mickanin, C., Huang, H., Sheppard, K.A., Porter, J.A., Lu, C.X., *et al.* (2012). USP4 is regulated by AKT phosphorylation and directly deubiquitylates TGF-beta type I receptor. Nat Cell Biol *14*, 717-726.

Zheng, N., Schulman, B.A., Song, L., Miller, J.J., Jeffrey, P.D., Wang, P., Chu, C., Koepp, D.M., Elledge, S.J., Pagano, M., *et al.* (2002). Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. Nature *416*, 703-709.

Zheng, N., Wang, P., Jeffrey, P.D., and Pavletich, N.P. (2000). Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. Cell *102*, 533-539.

Zurek, N., Sparks, L., and Voeltz, G. (2011). Reticulon short hairpin transmembrane domains are used to shape ER tubules. Traffic *12*, 28-41.

Zusman, T., Aloni, G., Halperin, E., Kotzer, H., Degtyar, E., Feldman, M., and Segal, G. (2007). The response regulator PmrA is a major regulator of the icm/dot type IV secretion system in Legionella pneumophila and Coxiella burnetii. Mol Microbiol *63*, 1508-1523.