A LEGIONELLA EFFECTOR SIDC – A NEW FAMILY OF E3 UBIQUITIN LIGASE WITH A SPECIFIC PI(4)P BINDING DOMAIN

A Dissertation

Presented to the Faculty of the Graduate School

Of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

By

Xi Luo

August 2015

A LEGIONELLA EFFECTOR SIDC – A NEW FAMILY OF E3 UBIQUITIN LIGASE WITH A SPECIFIC PI(4)P BINDING DOMAIN

Xi Luo, Ph. D.

Cornell University 2015

The opportunistic intracellular pathogen *Legionella pneumophila* is the causative agent of Legionnaires' disease. *L. pneumophila* delivers nearly 300 effector proteins into host cells for the establishment of a replication permissive compartment known as the *Legionella*-containing vacuole (LCV). Among these proteins, SidC anchors to the cytoplasmic surface of the LCV via binding to phosphatidylinositol-4-phosphate [PI(4)P] and is important for the recruitment of host endoplasmic reticulum (ER) proteins to this organelle. However, the biochemical function underlying this activity is unknown. We first determined the structure of the N-terminal domain of SidC, which has no structural homology to any protein. Sequence homology analysis revealed a potential canonical catalytic triad formed by Cys46, His444, and Asp446 on the surface of SidC. Unexpectedly, we found that SidC is an E3 ubiquitin ligase which utilizes the C-H-D triad to catalyze the formation of high molecular weight poly-ubiquitin chains through multiple ubiquitin lysine residues. A C46A mutation completely abolished the E3 ligase activity and the ability of the protein to recruit host ER proteins as well as poly-ubiquitin conjugates to the LCV. Thus, SidC represents a novel E3 ubiquitin ligase family important for phagosomal membrane remodeling by *L. pneumophila*. Next, we also reported the crystal structure of SidC (1-871). The structure revealed that SidC contains four domains that are packed into an arch-like shape. The P4C domain (PI(4)P binding of SidC) is comprised of a four α-helix bundle and covers the

ubiquitin ligase catalytic site of the SNL domain. Strikingly, a pocket with characteristic positive electrostatic potentials is formed at one end of this bundle. Liposome binding assays of the P4C domain further identified the determinants of phosphoinositide recognition and membrane interaction. Interestingly, we further found that binding with PI(4)P stimulates the E3 ligase activity, presumably due to a conformational switch induced by PI(4)P from a closed form to an open active form. Mutations of key residues involved in PI(4)P binding significantly reduced the association of SidC to the LCV and abolished its activity in the recruitment of ER proteins and ubiquitin signals, highlighting that PI(4)P-mediated targeting of SidC is critical for its function in the remodeling of bacterial phagosome membrane. Finally, a GFP-fusion with the P4C domain was demonstrated to be specifically localized to PI(4)P-enriched compartments in mammalian cells. This domain shows potential to be developed as a sensitive and accurate PI(4)P probe in living cells.

BIOGRAPHICAL SKETCH

Xi Luo was born and grew up in Beijing, China. In 2003, Xi graduated from Beijing No.5 High School and started his undergraduate education at Nankai University, majoring in Biotechnology. While at Nankai University, he worked on the transformation of Lotus corniculatus with double salt-tolerant genes in the laboratory of Dr. Yong Wang. Xi was elected to join the Student Summer Training Program at National Institute of Biological Sciences (NIBS), Beijing, China in the summer of 2006. And he continued to perform his undergraduate thesis research at NIBS in Dr. Jijie Chai's lab studying the structural basis for activation of plant immunity by bacterial effector protein AvrPto. His work is part of the project, which published on *Nature* in 2007. Xi was accepted into the Biochemistry & Molecular Biology (BMB) graduate program at University of Rochester in 2008. He worked in the laboratory of Dr. Regis O'Keefe and Dr. Jennifer Jonason since 2009, focusing on investigating the roles of BMP-2/Smad and Ihh signaling in β–catenin-mediated endochondral bone formation. After passed his qualified exam and received a Master of Science degree at University of Rochester, Xi decided to transfer to the Biophysics graduate program at Cornell University in 2012. He started pursuing his Ph.D under the supervision of Dr. Yuxin Mao, with the thesis project focusing on the structural and functional studies of a *Legionella* effector SidC. Xi's thesis research were published in *Proc Natl Acad Sci U S A* (2014) and *PLoS Pathogens* (2015) as the first author. He demonstrated that the *Legionella* effector SidC defines a unique family of ubiquitin ligases and contains a unique PI(4)P binding domain important for bacterial phagosomal targeting and remodeling.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Yuxin Mao, to provide me the opportunity to work in his lab and for all of his support and advise. Yuxin is a terrific mentor who always teaches me how to think about science independently and to organize my work. He is a hardworking, genius, and talented scientist with easy-going personality. In addition to instructing my scientific development, he also acts like my friend or even my father to take care of all aspects of my career and life. His enthusiasm and motivation for science as well as his intelligence and diligence will always encourage me to be successful in my future research career.

I would like to thank my committee members, Dr. Gerald W Feigenson, Dr. Holger Sondermann and Dr. Brian Crane for their support, suggestions and guidance throughout my thesis project.

I would like to thank my lab mates, Dr. Fenghua Hu, Xiaochun Wu, Dr. Fosheng Hsu, Dr. Leila Toulabi, Dr. Xiaolai zhou, Jon Wasilko, Peter Sullivan, Anil Akturk and all the current and previous lab members in Yuxin and Fenghua's lab.

I would like to thank my wife, Dr. Ning Liu, whose love, smile and support always make me feel warm and be full of energy to overcome all our difficulties and challenges. Whenever I fell depressed or not confident, she always encourages and helps me to regain hope after despair, resume life after obstructions, restart journeys after detours, revive strength after defeat and resurrect dreams after rejection. It's my great fortune to have her to be the second half of life. With her, my moods have never seen the color blue.

I would like to thank all my family members, especially my parents in-law, Shufen Jiang, Chunhe Liu and my parents Lei Han, Qiang Luo. Their constant support and selfless dedication are the solid foundation for us. I am extremely grateful for all of their help and support throughout my whole life. I also want to thank my daughter Ariana Qiaochu and my son Ares Zixiao Luo. They are the angels of us and always the source of happiness. They are also the hope and the future for my life. In addition, I would like to thank all my relatives and friends in China, whom I miss so much and will never forget.

Thanks to all the other people whoever are in the memory of my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLE

CHAPTER I

Introduction

Legionella pneumophila pathogenesis

The gram-negative bacterium *Legionella pneumophila* is ubiquitously found in natural water systems, where it parasitizes free-living amoebae (Fields et al., 2002). It can also invade human macrophages and cause a severe pneumonia, called Legionnaires' disease, which is transmitted only by inhalation of contaminated aerosols (Nash et al., 1984). After internalization through phagocytosis, the bacteria hijack the trafficking of endoplasmic reticulum (ER)-derived vesicle and generate a distinct compartment called the "*Legionella*-containing vacuole" (LCV). This special vacuole avoids the fusion with lysosomes and allows *L. pneumophila* to be alive and replicate within the host cells (Isberg et al., 2009). The formation of the LCVs and the pathogenesis of the bacteria depend on the intracellular multiplication/defective organelle trafficking (Icm/Dot) type IV secretion system (T4SS), which translocates nearly 300 effecter proteins to modulate host signaling and vesicle trafficking pathways (Fig 1.1; Segal et al., 1998). During the process of LCV maturation, a particular phosphoinositide lipid, PI(4)P, is shown to become enriched on the LCV membrane (Weber et al., 2006). In addition, poly-ubiquitin conjugates are also discovered to accumulate on the LCV membrane and may play a key role in the membrane remodeling (Dorer et al., 2006). However, the molecular mechanisms underlying these processes remain elusive.

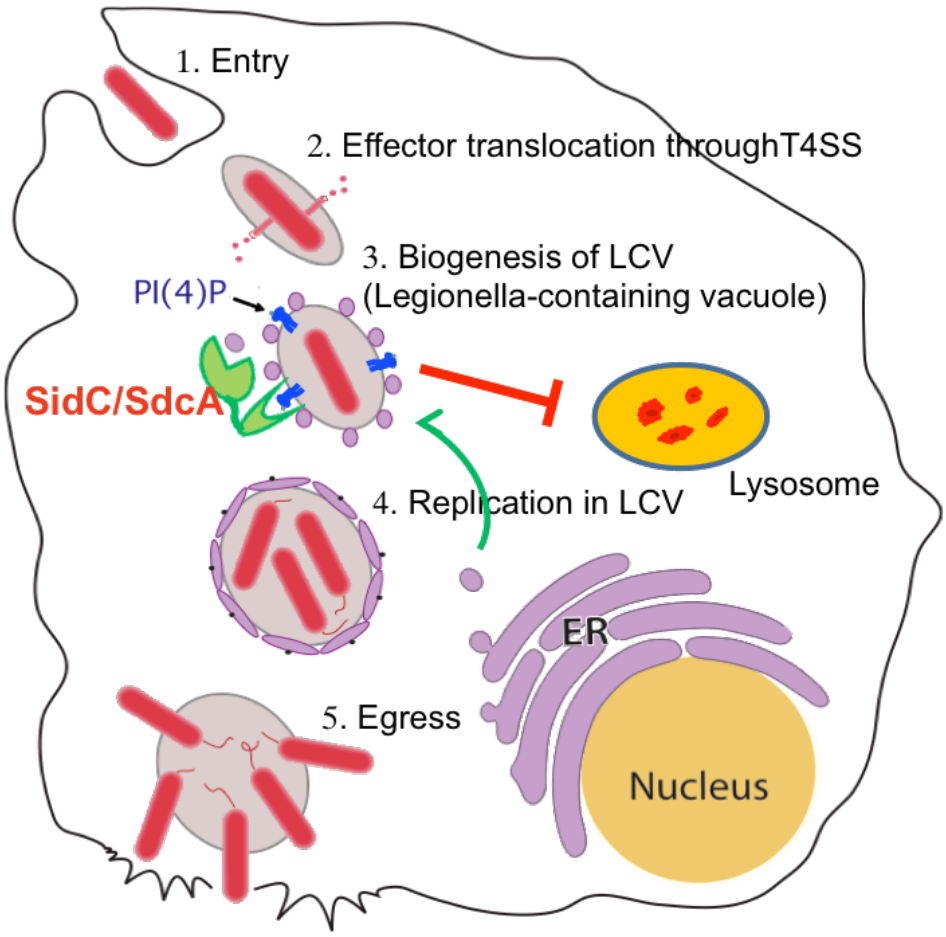


Figure 1.1. Intracellular replication cycle of *L. pneumophila*

After entering the host cells by phagocytosis, the avirulent bacteria injected about 300 bacterium's effector proteins into the host cell through the Dot/Icm type IV secretion system (T4SS). Some of these effector proteins function to recruit host ER derived vesicles and form a unique compartment called *L.pneumophila*–containing vacuoles (LCV), which can escape the default endo-lysosomal pathway and support replication of *L. pneumophila*. Eventually, the bacteria are released from the lyzed host cell and infect neighboring macrophages. During the process of LCV maturation, PI(4)P is found to be enriched on the LCV membrane. SidC and its paralogue SdcA is located on the LCV membrane through the interaction with PI(4)P.

Ubiquintination pathway and E3 Ubiquitin ligases

Ubiquitin is a highly conserved 76-amino-acid protein. The post-translational attachment of Ubiquitin to a lysine residue in its target proteins is a key mechanism in regulating diverse cellular processes, including protein degradation, cell cycle progression, and signal transduction (Laney et al., 1999). Ubiquitylation reactions involve the sequential transfer of activated Ubiquitin between three classes of proteins: ubiquitin-activating enzymes (E1s), ubiquitin-conjugating enzymes (E2s), and ubiquitin-protein ligases (E3s). The E1 uses ATP to activate free Ubiquitin and transfers it to an E2 by forming a thioester bond between the ubiquitin COOH terminus and the active site cysteine of an E1 or E2. When an E3 ubiquitin ligase forms a complex with both substrate and the ubiquitin conjugated E2 (E2~Ub), it directly or indirectly catalyzes the ligation between ubiquitin and the target protein through isopeptide linkage (Fig 1.2; Hershko et al., 1998). There are two major families of E3s classified: the RING domain ligases and the HECT domain ligases. The really interesting new gene (RING) domain functions as a scaffold to mediate the direct transfer of ubiquitin from E2 to substrate. There is no catalytic cysteine in the RING domain containing E3s. The structure contains a conserved motif with cysteine and histidine residues and is maintained through binding two zinc atoms (Deshaies et al., 2009). The E6AP carboxyl terminus (HECT) domain is highly conserved and located at the C-terminus of the protein, whereas their N-terminal domain usually contributes the substrate recognition. The structure of the HECT domain is L-shaped and bi-lobe. It utilizes the catalytic cysteine at the C-terminal C-lobe to form a ubiquitin-E3 thioester intermediate before transferring the ubiquitin from E2 to a lysine on the substrate. The N-terminal N-lobe usually contributes the E2 recognition. Basing on the first crystal structure of a HECT domain protein (E6AP) solved in complex with its E2 (UBCH7), the two lobes are far apart. But during the

ubiquitin transfer reaction, they can come together since they are connected with a flexible hinge (Fig 1.3; Huang et al., 1999).

Bacterial pathogens developed a number of effector proteins to interrupt host ubiquitin pathways. For instance, AvrPtoB from *Pseudomonas syringae* is a RING domain-like E3 ubiquitin ligase, which plays a role in inhibiting host programmed cell death (Janjusevic et al., 2006). The *Salmonella enterica* effector protein SopA is a HECT domain-containing E3 ubiquitin ligase (Diao J et al., 2008) and some bacterial E3s like the *Shigella* effector IpaH even has no structural similarity to any other ubiquitin ligase (Singer AU et al., 2008). Several *L. pneumophila* Icm/Dot effectors are also reported as E3 ubiquitin ligases with a F-box or U-box domain, including LegAU13/AnkB, LegU1 and LicA (Ensminger AW et al., 2010). The identification of a novel bacterial E3 ubiquitin ligase not only benefits the explanation of how pathogens interact with host ubiquitin pathway, but also helps to discover new families of E3 ubiquitin ligase.

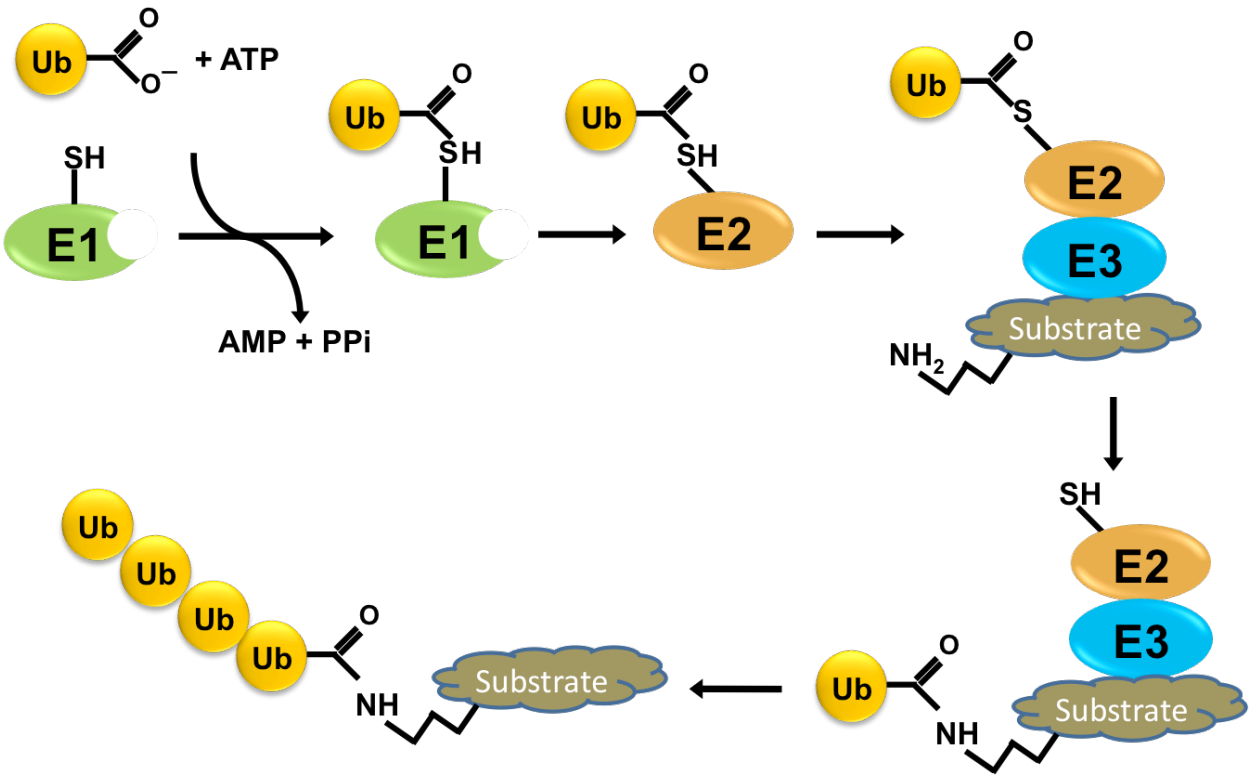


Figure 1.2. The Ubiquitylation system

The protein ubiquitination reaction starts with the ubiquitin activating enzyme E1. E1 utilizes the energy from ATP to activate ubiquitin by linking the carboxyl end of Ub through a thioester bond to its catalytic cysteine. In the presence of the second enzyme, the ubiquitin-conjugating enzyme E2, the activated ubiquitin is transfer to the catalytic cysteine of E2 also through a thioester linkage. Then, the third protein E3 ubiquitin ligase bridges the activated E2 and proper substrate together and catalyses the transfer of ubiquitin to a lysine on the substrate protein forming an iso-peptide bond. This process can be repeated until a multiubiquitin chain is formed.

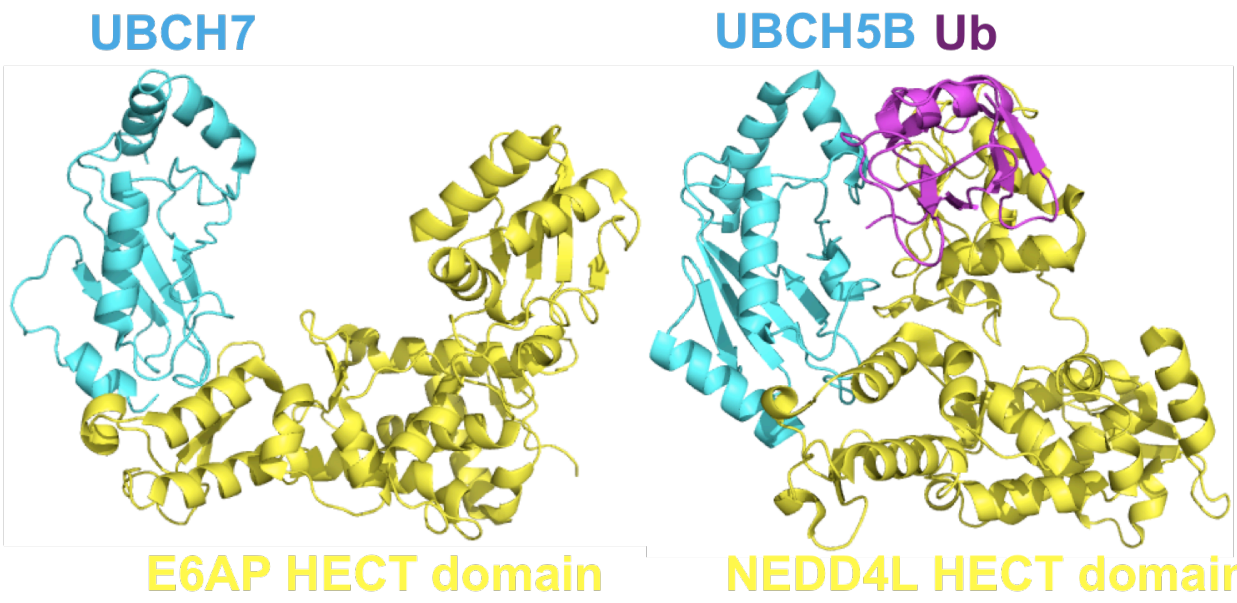


Figure 1.3. HECT E3 structure

The HECT domain consists of two lobes. The N-terminal N-lobe is responsible for the E2 interaction and the C-terminal C-lobe contains the catalytic cysteine that forms the thioester with ubiquitin. The first identified member of the HECT family is E6-associated protein (E6AP). The structure of E6AP in complex with its E2 UBCH7 is shown in the left panel. The C-lobe of E6AP in the E2-E3 complex structure is in a more open conformation where they active site cysteine residues are 41Å apart. The right panel shows the crystal structure of the NEDD4L HECT domain in complex with ubiquitin-conjugated E2 (UBCH5B~Ub). The E2-E3-Ub complex structure displays a more close architecture. The C-lobe of NEDD4L which contacts with the esterified ubiquitin shifts towards the UbcH5B and the distance between the E2 and E3 catalytic cysteines is about ~8Å. All of the above structural analyses indicate that the two lobes of the HECT domain come together during ubiquitin transfer through a flexible hinge.

Phosphoinositides and current PI(4)P binding probes

The phosphoinositides (PIs) are derivatives of phosphatidylinositol (PtdIns) that can be phosphorylated at the 3', 4' and 5' positions of their inositol headgroup in any one of seven combinations. Most phosphoinositides have quite unique intracellular distributions and they function as regulators by targeting effecter proteins to specific compartments (Fig 1.4). Although they are minor component of cellular membranes, PIs play a fundamental role in a broad range of cell signaling and membrane trafficking events (Di Paolo et al., 2006). Phosphatidylinositol-4-phosphate (PI(4)P), the most abundant monophosphorylated phosphoinositide, was originally considered as a simple intermediate for synthesizing PI(4,5)P2 on the plasma membrane(PM) (Doughman et al., 2003). After the identification of multiple widespread PtdIns 4-kinases (PI4K), multiple pools of PI(4)P was discovered, including on the Golgi complex, the endosomal system, and the plasma membrane (D'Angelo et al., 2008). The most critical role of PI(4)P is in Golgi complex involved in the recruitment of effecter proteins by the direct interaction with their PI(4)P binding domain (Graham et al., 2011). Among these PI(4)P interacting proteins, the pleckstrin homology (PH) domain of Four-phosphate-adaptor protein 1 (FAPP1), and the PH domain of Oxysterol-binding protein (OSBP) were previously used as PI(4)P-binding biosensors for the study of PI(4)P intracellular distribution (Balla et al., 2005). However, the FAPP1-PH also strongly interact with another Golgi localized protein ADP-ribosylation factor 1 (ARF1), which bias its localization towards the Golgi complex (Balla et al., 2000). Also, the PH domain of OSBP prefers to target plasma membranes due to its equal affinity for binding PI(4)P and PI(4,5)P2 (Levine et al., 2002). These drawbacks of the two domains limited them from being ideal PI(4)P binding probes. Two Icm/Dot substrates SidM (Brombacher et al., 2009) and SidC (Luo et al., 2004) coordinate and function on the

cytoplasmic surface of the LCV membrane through the interaction with PI(4)P. The structural basis for the Rab1 guanine nucleotide exchange factor SidM to recognize PI(4)P is previously well described based on its crystal structure with (Claudia et al., 2014) or without PI(4)P (Zhu et al., 2010). The PI(4)P binding domain of SidM (P4M) was also developed as a PI(4)P binding probe for the study of PI(4)P intracellular distribution (Gerald et al., 2014). However, the structural mechanism for the PI(4)P dependant membrane targeting of the other *L. pneumophila* PI(4)P interactor SidC and the potentiality of it as a novel PI(4)P specific binding probe has not been established yet.

Legionella Effector SidC

The Icm/Dot substrate SidC is a 917 amino acid protein. The study of the SidC and its paralogue SdcA depletion stain indicated their importance during *Legionella pneumophila* pathogenesis. ΔSidC-SdcA mutant does not impair the process of phagocytosis, intracellular replication of the bacteria, or the ability of it to avoid fusion with lysosomes. But two features of SidC are revealed. The N-terminal domain of SidC is shown to be required for the recruitment of ER derived vesicles during the formation of the LCV and SidC localized on the cytoplasmic side of LCV membrane by binding with PI(4)P through its C-terminal PI(4)P binding domain, which was mapped within the residues 609-776 (Fig 1.5; Ragaz et al., 2008). My thesis project aims to further investigate the detailed mechanisms of these two features of SidC.

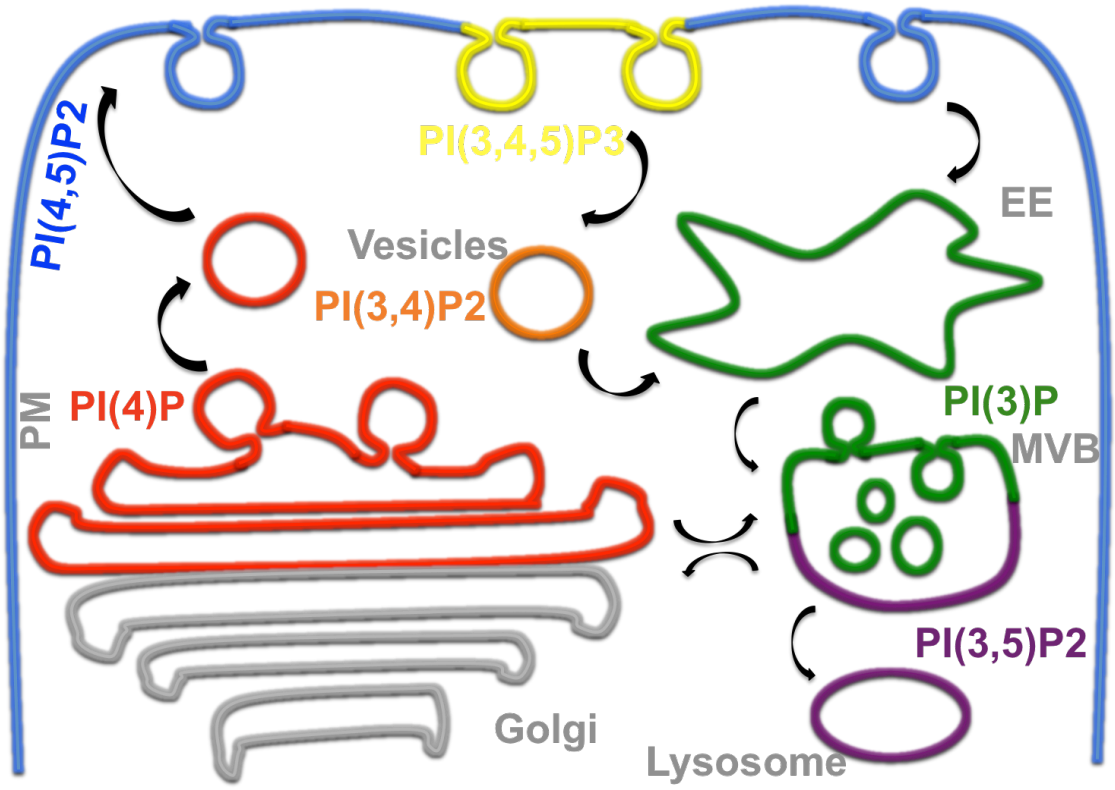
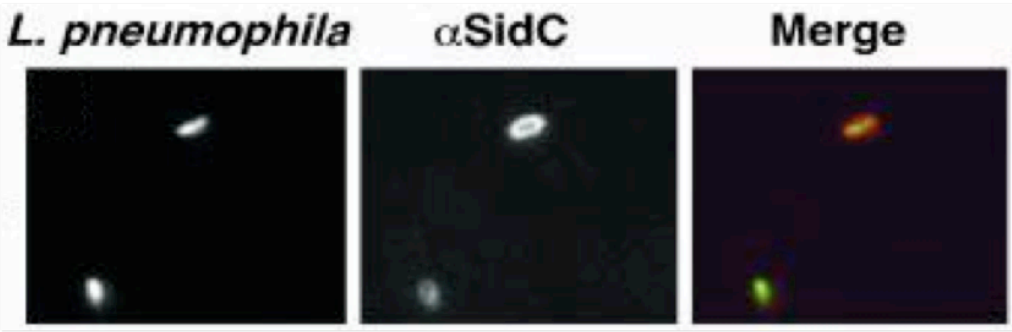


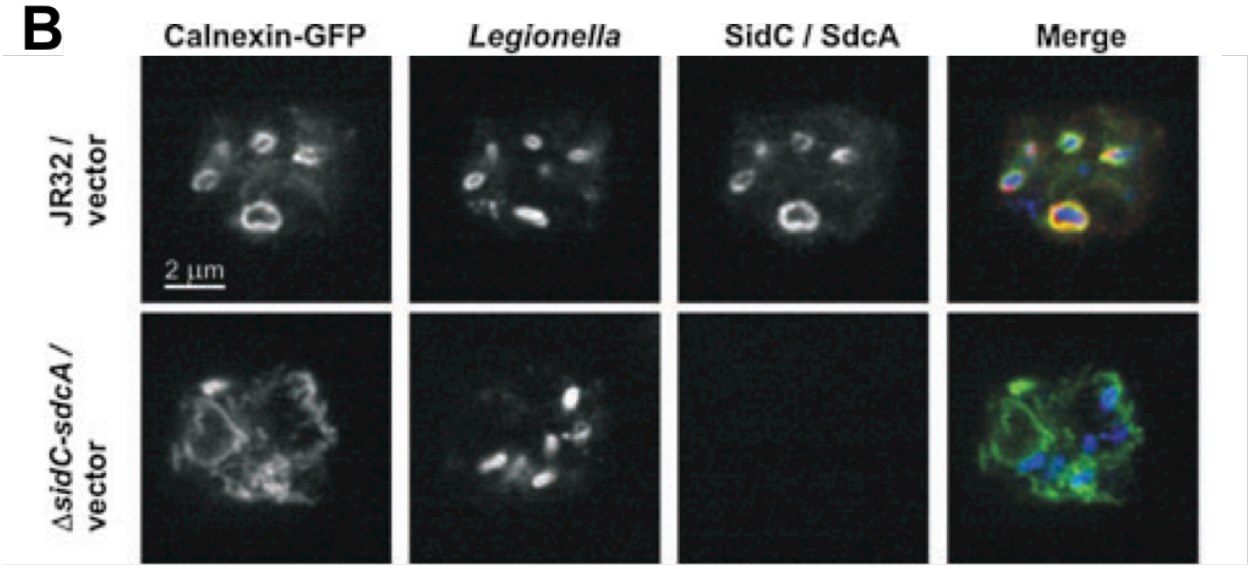
Figure 1.4. Phosphoinositides play a critical role in defining and maintaining organelle identity. PIs play essential roles in a broad spectrum of cellular processes. Each of the PI has its own distinctive subcellular localization. PI(4,5)P2 and PI(3,4,5)P3 are mainly distributed on plasma membrane. PI(4)P is enriched in Golgi complex, plasma membrane and secreted vesicles. PI(3)P is mainly discovered in early or late endosomes and PI(3,5)P2 is majority localized in late endosomes, MVB and lysosomes. PI(3,4)P2 is largely found in trafficking vesicles.

A



Luo and Isberg, PNAS (2004)

B



Curdin Ragaz et al. Cellular Microbiology (2008)

Figure 1.5. SidC localized on the cytoplasmic side of LCV membrane and is required for the recruitment of ER derived vesicles.

(A) The translocation of SidC on the cytoplasmic side of LCV membrane. Bacteria strains Lp02 expressing GFP associated with bone marrow-derived macrophage. Immunoprobing of infected cells with anti-$(His)_6$-SidC. (B) *L. pneumophila* Δ*sidC-sdcA* is defective for the recruitment of calnexin-GFP to LCVs. Confocal laser scanning micrographs of calnexin-GFP-labelled *D. discoideum* AX3 (green), infected at an MOI of 50 for 1 h with *L. pneumophila* wild-type strain JR32 labelled with a serogroup-specific antibody (blue) and immuno-stained for SidC/SdcA with an anti-M45 antibody (red). Bar, 2 μm.

REFERENCES

Balla, A., G. Tuymetova, A. Tsiomenko, P. Várnai, T. Balla. (2005). A plasma membrane pool of phosphatidylinositol 4-phosphate is generated by phosphatidylinositol 4-kinase type-III alpha: studies with the PH domains of the oxysterol binding protein and FAPP1. *Mol. Biol. Cell.* 16:1282–1295. doi:10.1091/mbc.E04-07-0578

Balla, T., T. Bondeva, P. Várnai. (2000). How accurately can we image inositol lipids in living cells? *Trends Pharmacol. Sci*. 21:238–241. doi:10.1016/S0165-6147(00)01500-5

Brombacher, E., S. Urwyler, C. Ragaz, S.S. Weber, K. Kami, M. Overduin, H. Hilbi. (2009). Rab1 guanine nucleotide exchange factor SidM is a major phosphatidylinositol 4-phosphate-binding effector protein of Legionella pneumophila. *J. Biol. Chem*. 284:4846–4856. doi:10.1074/jbc.M807505200

Claudia M. Del Campo, Ashwini K. Mishra, Yu-Hsiu Wang, Craig R. Roy, Paul A. Janmey, David G. Lambright. (2014) Structural Basis for PI(4)P-Specific Membrane Recruitment of the Legionella pneumophila Effector DrrA/SidM. *Structure*. Volume 22, Issue 3, Pages 397–408

D'Angelo, G., Vicinanza, M., Di Campli, A., De Matteis M.A. (2008) The multiple roles of PtdIns(4)P — not just the precursor of PtdIns(4,5)P2. *J. Cell Sci*., 121, pp. 1955–1963

Deshaies, R. J. and Joazeiro, C. A. (2009). RING domain E3 ubiquitin ligases. *Annu. Rev. Biochem*. 78, 399-434.

Di Paolo, G., De Camilli, P. (2006) Phosphoinositides in cell regulation and membrane dynamics. *Nature*, 443, pp. 651–657

Diao J, Zhang Y, Huibregtse JM, Zhou D, Chen J (2008) Crystal structure of SopA, a Salmonella effector protein mimicking a eukaryotic ubiquitin ligase. *Nat Struct Mol Biol* 15(1):65–70.

Dorer MS, Kirton D, Bader JS, & Isberg RR (2006) RNA interference analysis of Legionella in Drosophila cells: exploitation of early secretory apparatus dynamics. *PLoS Pathog* 2(4):e34.

Doughman RL, Firestone AJ, Anderson RA (2003) Phosphatidylinositol phosphate kinases put PI4,5P(2) in its place. *J Membr Biol Jul* 15;194(2):77-89.

Ensminger AW, Isberg RR (2010) E3 ubiquitin ligase activity and targeting of BAT3 by multiple Legionella pneumophila translocated substrates. *Infect Immun* 78(9): 3905–3919.

Fields BS, Benson RF, & Besser RE (2002) Legionella and Legionnaires' disease: 25 years of investigation. *Clin Microbiol Rev* 15(3):506-526.

Graham T.R., Burd C.G. (2011) Coordination of Golgi functions by phosphatidylinositol 4-kinases. *Trends Cell Biol.,* 21, pp. 113–121

Hammond GR, Machner MP, Balla T (2014) A novel probe for phosphatidylinositol 4-phosphate reveals multiple pools beyond the Golgi. *J Cell Biol* 205: 113-126.

Hershko, A., Ciechanover, A. (1998) The ubiquitin system. *Annu. Rev. Biochem.,* 67, pp. 425-479

Huang, L., Kinnucan, E., Wang, G., Beaudenon, S., Howley, P. M., Huibregtse, J. M. and Pavletich, N. P. (1999). Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science* 286, 1321-1326.

Isberg RR, O'Connor TJ, & Heidtman M (2009) The Legionella pneumophila replication vacuole: making a cosy niche inside host cells. *Nat Rev Microbiol* 7(1):13-24.

Janjusevic R, Abramovitch RB, Martin GB, Stebbins CE (2006) A bacterial inhibitor of host programmed cell death defenses is an E3 ubiquitin ligase. *Science* 311(5758): 222–226.

Laney J.D., M. Hochstrasser. (1999) Substrate targeting in the ubiquitin system. *Cell*, 97, pp. 427–430

Levine, T.P., S. Munro. (2002). Targeting of Golgi-specific pleckstrin homology domains involves both PtdIns 4-kinase-dependent and -independent components. *Curr. Biol.* 12:695–704. doi:10.1016/S0960-9822(02)00779-0

Luo ZQ, Isberg RR (2004) Multiple substrates of the Legionella pneumophila Dot/Icm system identified by interbacterial protein transfer. *Proc Natl Acad Sci USA* 101(3): 841–846.

Nash TW, Libby DM, Horwitz MA (1984) Interaction between the legionnaires' disease bacterium (Legionella pneumophila) and human alveolar macrophages. Influence of antibody, lymphokines, and hydrocortisone. *J Clin Invest* 74: 771-782.

Ragaz C, et al. (2008) The Legionella pneumophila phosphatidylinositol-4 phosphate binding type IV substrate SidC recruits endoplasmic reticulum vesicles to a replication permissive vacuole. *Cell Microbiol* 10(12):2416–2433.

Segal G, Purcell M, Shuman HA (1998) Host cell killing and bacterial conjugation require overlapping sets of genes within a 22-kb region of the Legionella pneumophila genome. *Proc Natl Acad Sci U S A* 95: 1669-1674.

Singer AU, et al. (2008) Structure of the Shigella T3SS effector IpaH defines a new class of E3 ubiquitin ligases. *Nat Struct Mol Biol* 15(12):1293–1301.

Weber SS, Ragaz C, Reus K, Nyfeler Y, & Hilbi H (2006) Legionella pneumophila exploits PI(4)P to anchor secreted effector proteins to the replicative vacuole. *PLoS Pathog* 2(5):e46.

Y. Zhu, L. Hu, Y. Zhou, Q. Yao, L. Liu, F. Shao. (2010) Structural mechanism of host Rab1 activation by the bifunctional Legionella type IV effector SidM/DrrA. *Proc. Natl. Acad. Sci. USA*, 107, pp. 4699–4704

CHAPTER II

The *Legionella* effector SidC defines a unique family of ubiquitin ligases important for bacterial phagosomal remodeling

FoSheng Hsu[a,*], Xi Luo[a,*], Jiazhang Qiu[b,*], Yanbin Teng[a], Jianping Jin[c], Marcus Smolka[a], Zhao-Qing Luo[b], and Yuxin Mao[a]

[a]Weill Institute for Cell and Molecular Biology and Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA.

[b]Department of Biological Sciences, Purdue University, 915 West State Street, West Lafayette, Indiana 47907, USA.

[c]Department of Biochemistry & Molecular Biology, Medical School, The University of Texas Health Science Center at Houston, 6431 Fannin Street, Houston, Texas, 77030, USA

[*]These authors contribute equally

Xi Luo contributed significantly to Figure 2.1-2.12 and 2.14-2.16 and assisted in designing experiments and preparation of the manuscript.

Atomic coordinates and structure factors for the reported structures have been deposited into the Protein Data Bank under the accession codes 4TRH (native SidC), 4TRG (Hg-bound form)

ChapterII was originally published in Proc Natl Acad Sci U S A. 111(29):10538-43 (2014)

Summary

The activity of proteins delivered into host cells by the Dot/Icm injection apparatus allows *Legionella pneumophila* to establish a niche called the *Legionella*-containing vacuole (LCV), which is permissive for intracellular bacterial propagation. Among these proteins, SidC anchors to the cytoplasmic surface of the LCV and is important for the recruitment of host endoplasmic reticulum (ER) proteins to this organelle. However, the biochemical function underlying this activity is unknown. Here we determined the structure of the N-terminal domain of SidC, which has no structural homology to any protein. Sequence homology analysis revealed a potential canonical catalytic triad formed by Cys46, His444, and Asp446 on the surface of SidC. Unexpectedly, we found that SidC is an E3 ubiquitin ligase which utilizes the C-H-D triad to catalyze the formation of high molecular weight poly-ubiquitin chains through multiple ubiquitin lysine residues. A C46A mutation completely abolished the E3 ligase activity and the ability of the protein to recruit host ER proteins as well as poly-ubiquitin conjugates to the LCV. Thus, SidC represents a novel E3 ubiquitin ligase family important for phagosomal membrane remodeling by *L. pneumophila*.

Introduction

Legionella pneumophila is a ubiquitous bacterium found in aquatic environments where it infects freshwater protozoa. Development of the potentially fatal Legionnaires' disease occurs when susceptible individuals inhale contaminated aerosols (Fields et al., 2002). After engulfment by phagocytes, *L. pneumophila* uses its Dot/Icm type IV secretion system to deliver a large number of effector proteins that modulate host cellular processes, leading to creation of a specialized *Legionella* containing vacuole (LCV) that provides the environment for robust intracellular bacterial growth (Isberg et al., 2009). The mature LCV is characterized by an enrichment of a particular phosphoinositide lipid PI(4)P (Weber et al., 2006; Hsu et al., 2012; Toulabi et al., 2013) and by the accumulation of endoplasmic reticulum (ER) proteins, presumably captured by intercepting vesicles derived from the ER (kagan & Roy et al., 2002; Tilney et al., 2001). Another feature of the LCV is enrichment of poly-ubiquitin conjugates around the vacuolar membrane (Dorer et al., 2006). This sophisticated membrane remodeling process is achieved by the coordinated activity of effector proteins delivered into the host through the Dot/Icm apparatus (Zhu et al., 2011; Lifshitz et al., 2013). However, the mechanism underlying these processes is still not well established. Recently, the *Legionella* effector proteins SidC (***S***ubstrate of ***I***cm/***D***ot transporter) (Luo & Isberg et al., 2004) and its paralog SdcA were proposed to function as vesicle fusion tethering factors. Both proteins were shown to recruit ER vesicles to the LCV while anchored on the LCV via a specific C-terminal phosphatidylinositol-4-phosphate [PI(4)P]-binding domain (Weber et al., 2006; Luo & Isberg et al., 2004; Ragaz et al., 2008). Bacterial vacuoles harboring the Δ*sidC-sdcA* mutant bacteria recruit ER-derived vesicles less efficiently, and in vitro experiments further showed that the N-terminal 70 kDa fragment of

SidC binds to ER vesicles in *Dictyostelium* and macrophage lysates (Ragaz et al., 2008). However, the biochemical mechanism for SidC-mediated ER recruitment remains unclear.

Post-translational modification by ubiquitin (Ub) regulates a myriad of cellular pathways, including protein homeostasis (Hershko & Ciechanovor et al., 1998), cell signaling (Chen & Sun et al., 2009), and membrane trafficking (Hurley & Stenmark et al., 2011; Haglund & Dikic et al., 2012). Protein ubiquitination requires the sequential activities of a cascade of enzymes known as ubiquitin activating enzyme E1, ubiquitin conjugating enzyme E2, and ubiquitin ligase E3 (Scheffiner & Huibregtse et al., 1995). Ubiquitin E3s belong to two major groups: the RING-domain (Really Interesting New Gene) and the HECT-domain (Homologous to the E6AP Carboxyl Terminus) families (Deshaies & Joazeiro et al., 2009; Rotin & Kumar et al., 2009). E3s in the RING family facilitate the transfer of Ub from the E2 catalytic cysteine directly to a substrate (Deshaies & Joazeiro et al., 2009). However, HECT-domain E3s utilize their own active cysteine to form a thioester bond-linked E3~Ub intermediate before transfering the Ub to specific substrates.

Due to the important role of protein ubiquitination in eukaryotes, it is not surprising that a number of bacterial pathogens and symbionts exploit the host ubiquitin pathway (Hubber & Nagai et al., 2013; Jiang & Chen et al., 2012). For example, the *Pseudomonas syringae* protein AvrPtoB contains a RING-like domain and functions as a ubiquitin ligase to inhibit host programmed cell death (Janjusevic et al., 2006). The effector SopA from *Salmonella enterica* is a HECT-like ubiquitin ligase with a catalytic domain organized in a bilobed architecture similar to the conventional HECT E3s (Diao et al., 2008). The *Shigella* effector IpaH has a C-terminal domain that carries ubiquitin ligase activity but bears no sequence and structural resemblance to other ubiquitin ligases (Singer et al., 2008).

In *L. pneumophila*, a number of Dot/Icm effectors are predicted to contain regions with sequence similarity to the F-box or U-box domain (De Felipe et al., 2005; Cazalet et al., 2004). Several of such predicted F-box containing effectors, including LegAU13/AnkB (Ensminger et al., 2010; Price et al., 2009; Lomma et al., 2010), LegU1, and LicA (Ensminger et al., 2010) interact with components of the Skp-Cullin-F-box (SCF) ubiquitin ligase complex. In vitro assays have further validated the ubiquitin E3 ligase activity for LegU1, LegAU13/AnkB (Ensminger et al., 2010), and LubX (Kubori et al., 2008). Using yeast two-hybrid or other biochemical methods, several specific host and bacterial proteins that are targeted for ubiquitination by *Legionella* E3 ligases have been identified. The U-box type E3 ligase LubX polyubiquitinates the host kinase Clk1 (Kubori et al., 2008) and the *Legionella* effector SidH (Kubori et al., 2010). The F-box protein LegU1 specifically directs the ubiquitination of the host chaperone protein BAT3 (Ensminger et al., 2010). The co-option of the host ubiquitin pathway by *L. pneumophila* is further supported by a seminal discovery that the bacterium recruits polyubiquitinated species around the bacterial phagosome shortly after bacterial uptake (Dorer et al., 2006). Despite these findings, much remains to be discovered regarding effectors and their targets in the exploitation of the host ubiquitin pathway by *L. pneumophila*.

Here we report the crystal structure of the N-terminal portion of SidC, which revealed a canonical catalytic triad containing a cysteine, a histidine, and an aspartate residue. We found that ectopic expression of this domain altered the intracellular ubiquitination pattern. In vitro experiments demonstrated the formation of high molecular weight ubiquitinated conjugates in a manner that is dependent on the catalytic residue C46. We further showed that the SidC paralog SdcA has ubiquitin ligase activity but with a preference for a different E2. Finally, wild type SidC, but not the C46A mutant, can fully complement the defect of *ΔsidC-sdcA* mutant in the

recruitment of ER markers during infection. Our results demonstrate that the *Legionella* effector SidC defines a new family of ubiquitin ligases, the activity of which facilitates the maturation of LCV by remodeling its protein composition.

Results

Crystal structure of the N-terminal domain of SidC revealed a canonical catalytic triad

SidC is a large protein containing 917 amino acid residues with a conserved N-terminal domain and a C-terminal PI(4)P binding domain (Fig 2.1A). The N-terminal portion of SidC has sequence homology to other *Legionella* proteins lacking this PI(4)P binding domain, indicating the presence of a functional independent module (aa. 1-542) within the N-terminus (Figs. 2.2-2.3). We named this conserved domain as SNL domain (**S**idC **N**-terminal ubiquitin **L**igase domain, see below).

The structure of the SNL domain was determined by single isomorphous replacement with anomalous signal (SIRAS) using crystals soaked with a mercury containing compound (Fig. 2.4, Table 2.1). The structure reveals that the SNL domain has a crescent-like overall shape consisting of two sub-domains (Fig. 2.1A and B). A small all-alpha helical domain (in green) is inserted into the rest of the protein between residues 224-327 (Figs. 2.3 and 2.5). The main domain contains a core of two layered β sheets, which is sandwiched between two clusters of α helices (Fig. 2.1B and Fig. 2.5). Structural homology search with the Dali server (Holm & Rosenstrom et al., 2010) did not yield any significant hits, which restricts the functional assignment based on structural homology. Since many *Legionella* effector proteins are enzymes, we hypothesized that SidC may be an enzyme. In this case, the catalytic residues responsible for

the presumed catalytic function should be conserved across all SidC homologs. To test this hypothesis, we performed multiple sequence alignment analyses. The alignment revealed two clusters of conserved residues within the SNL domain (Fig. 2.1C and Fig. 2.3). Strikingly, when all the identical residues across the SNL domain family members were mapped to the crystal structure, some of these residues emerging on the surface of the structure formed a continuous patch (Fig. 2.1C). Within this patch, three completely identical residues, C46, H444, and D446, are arranged in a way that is reminiscent of the classical catalytic triad found in cysteine-based proteases (Shao et al., 2002) and deubiquitinases (DUB) (Reyes-Turcu et al., 2009) and other cysteine-based enzymes. Interestingly, the area encompassing this conserved residue patch shows a concentrated negative electrostatic potential (Fig. 2.6), suggesting a potential site for protein-protein or protein-substrate interactions. Taken together, the sequence and structure analyses of the SNL domain raised the possibility that this domain is an enzyme containing a conserved Cys-His-Asp catalytic triad.

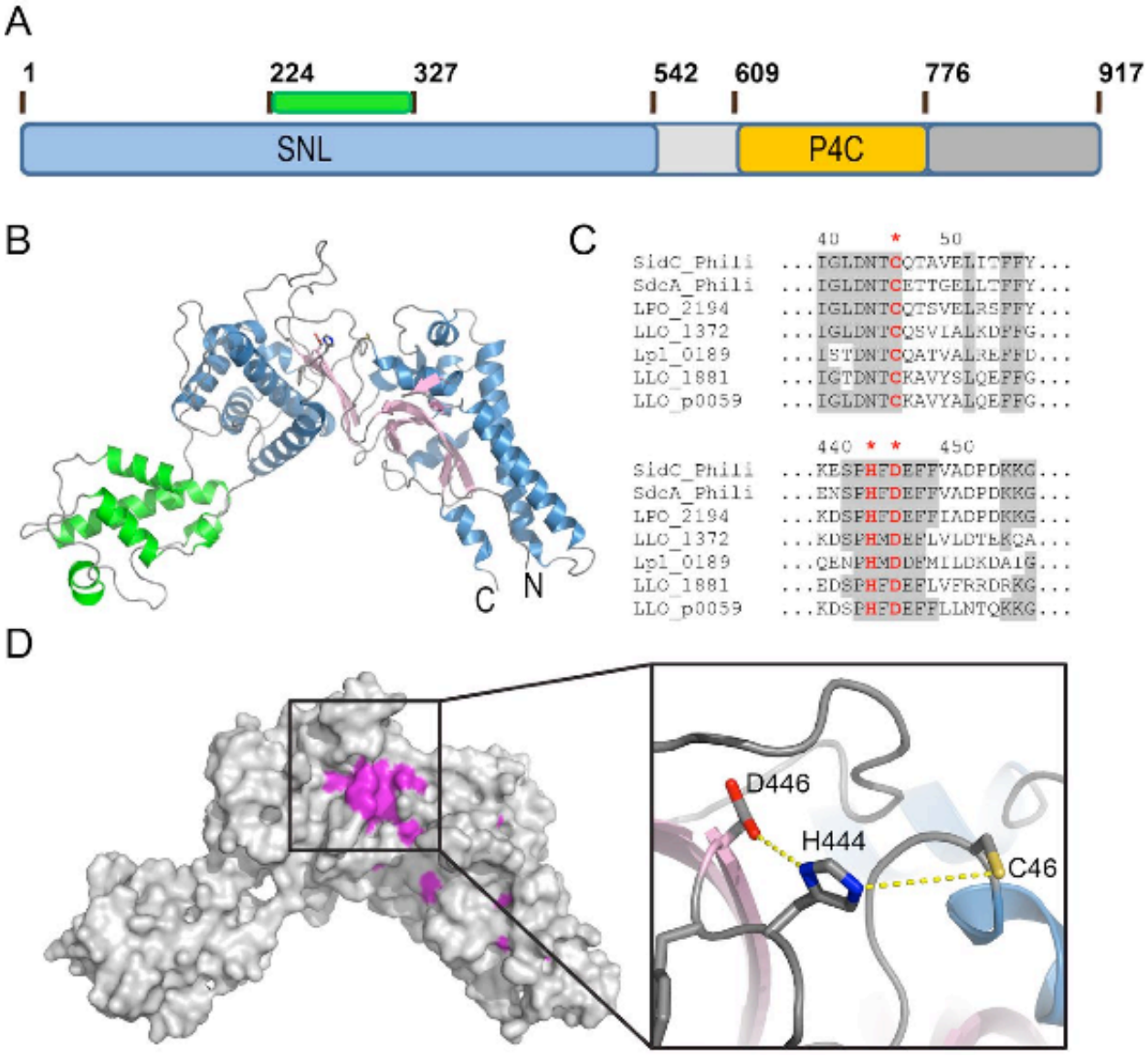

A
1
224
327
542
609
776
917
SNL
P4C
B
C
N
C
40
*
50
SidC_Phili ...IGLDNTCQTAVELITFFY...
SdcA_Phili ...IGLDNTCETTGELLTFFY...
LPO_2194 ...IGLDNTCQTSVELRSFFY...
LLO_1372 ...IGLDNTCQSVIALKDFFG...
Lpl_0189 ...ISTDNTCQATVALREFFD...
LLO_1881 ...IGTDNTCKAVYSLQEFFG...
LLO_p0059 ...IGLDNTCKAVYALQEFFG...
440
*
*
450
SidC_Phili ...KESFHFDEFFVADPDKKG...
SdcA_Phili ...ENSFHFDEFFVADPDKKG...
LPO_2194 ...KDSFHFDEFFIADPDKKG...
LLO_1372 ...KDSFHMDEFLVLDTEKQA...
Lpl_0189 ...QENFHMDDFMLLDKDAIG...
LLO_1881 ...EDSFHFDEFLVFRRDRKG...
LLO_p0059 ...KDSFHFDEFFLLNTQKKG...
D
D446
H444
C46

Figure 2.1. Crystal structure of the N-terminal SNL domain of SidC (aa. 1-542).

(*A*) Schematic diagram of the domain structure of SidC. SidC contains an N-terminal SNL domain (blue) and a C-terminal PI(4)P binding domain, P4C (yellow). The green bar indicates the position of an inserted sub-domain within the SNL domain. (*B*) Ribbon diagram of the overall structure of the SNL domain. The SNL domain has two sub-domains. The main sub-domain contains a two layered β-sheet (pink) flanked by two clusters of α-helices (blue). The all α-helical insertion sub-domain is shown in green. (*C*) Multiple sequence alignment reveals two clusters of conserved residues. The first cluster contains an invariable cysteine C46, and the second cluster contains invariable H444 and D446. (*D*) Mapping of the invariable residues across all SNL domains to the structure. Identical residues in the two conserved residues clusters shown in (*C*) form a patch on the surface. **Inset**: A zoom-in view of this conserved residue patch unveils the presence and the structural arrangement of a canonical catalytic triad formed by C46, H444, and D446.

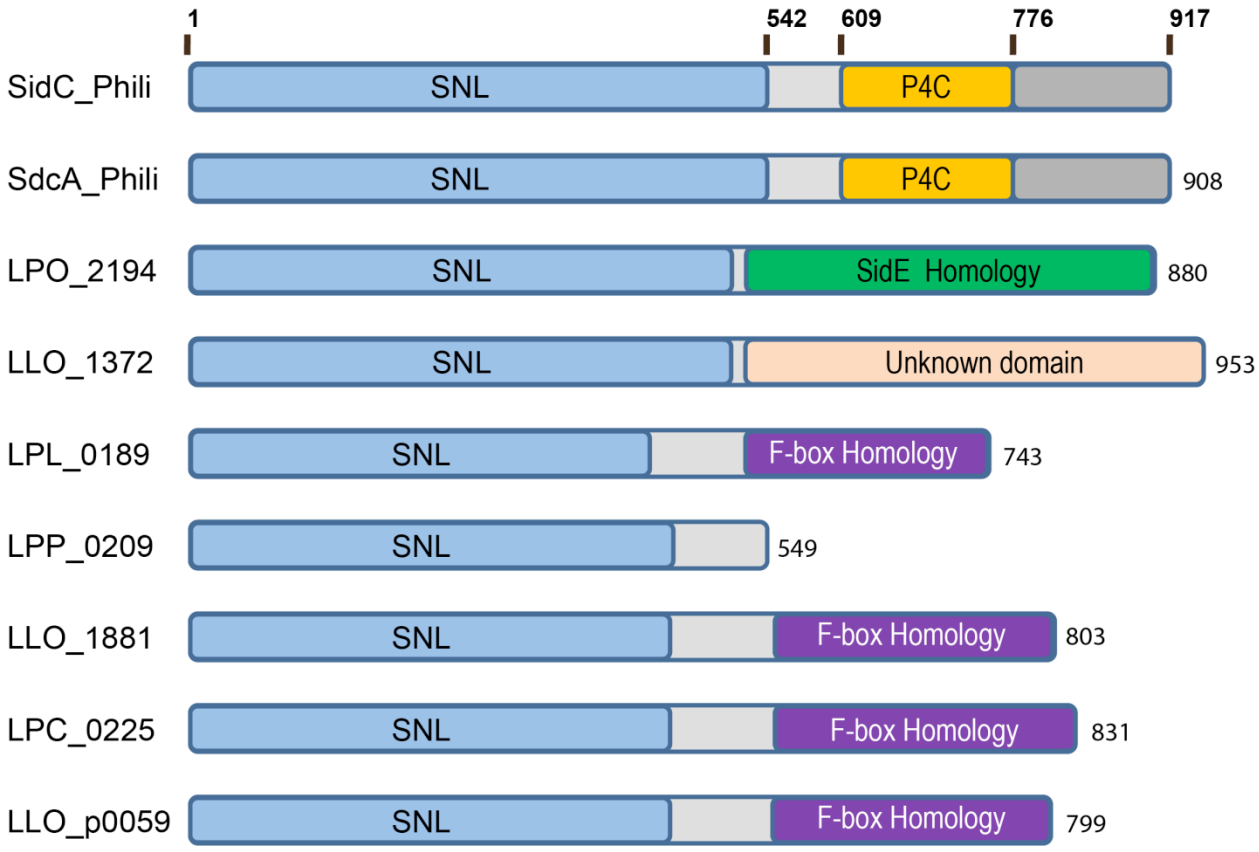


Figure 2.2. Schematic domain structures of selective members of the SidC family.

SidC is a protein comprised of 917 residues. It contains an N-terminal SNL domain (**S**idC **N**-terminal **Li**gase domain) and a C-terminal P4C domain (**p**hosphatidylinositol-**4**-phosphate binding of Sid**C**). The SNL domain is located at the N-terminus of all members of the SidC family. Entrez database accession numbers are as follow: SidC_Phili, gi: 52842719; SdcA__Phili, gi: 52842718; LPO_2194, gi: 397664568; LLO_1372, gi: 289164709; LPL_0189, gi: 54293148; LPP_0209, gi: 54296184; LLO_1881, gi: 289165217; LPC_0225, gi: 148358361; and LLO_p0059, gi: 308051561. The C-terminal portion of LPO_2194 contains a domain, which is conserved with the C-terminal part of SidE family proteins. The C-terminal part of LPL_0189, LLO_1881, LPC_0225, and LLO_p0059 is homologous to F-box domain-containing proteins.

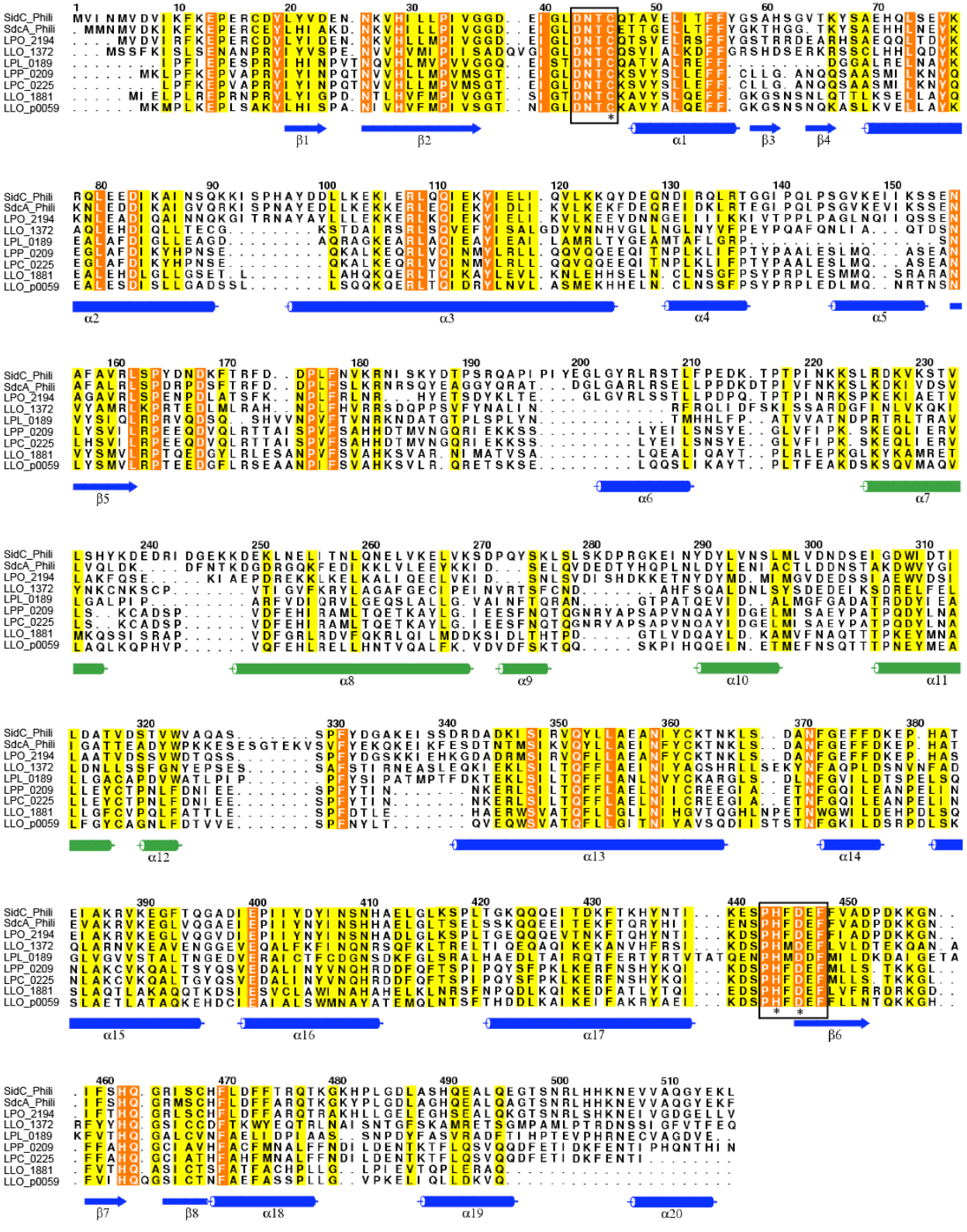

SidC_Phili
SdcA_Phili
LPO_2194
LLO_1372
LPL_0189
LPP_0209
LPC_0225
LLO_1881
LLO_p0059
β1
β2
α1
β3
β4
α2
α3
α4
α5
β5
α6
α7
α8
α9
α10
α11
α12
α13
α14
α15
α16
α17
β6
β7
β8
α18
α19
α20

Figure 2.3. Multiple sequence alignment of the SNL domains of SidC.

The sequences corresponding to the SNL domain of SidC (aa. 1-542) from different *Legionella* species were aligned by Clustal Omega (Sievers et al., 2011) and colored by ALSCRIPT (Barton et al., 1993). Secondary elements are drawn below the alignment. Two conserved sequence clusters are marked with a square. The predicted catalytic triad consisting of C46, H444, and D446 are highlighted with stars. Entrez database accession numbers are listed in the legend of Fig. 2.2.

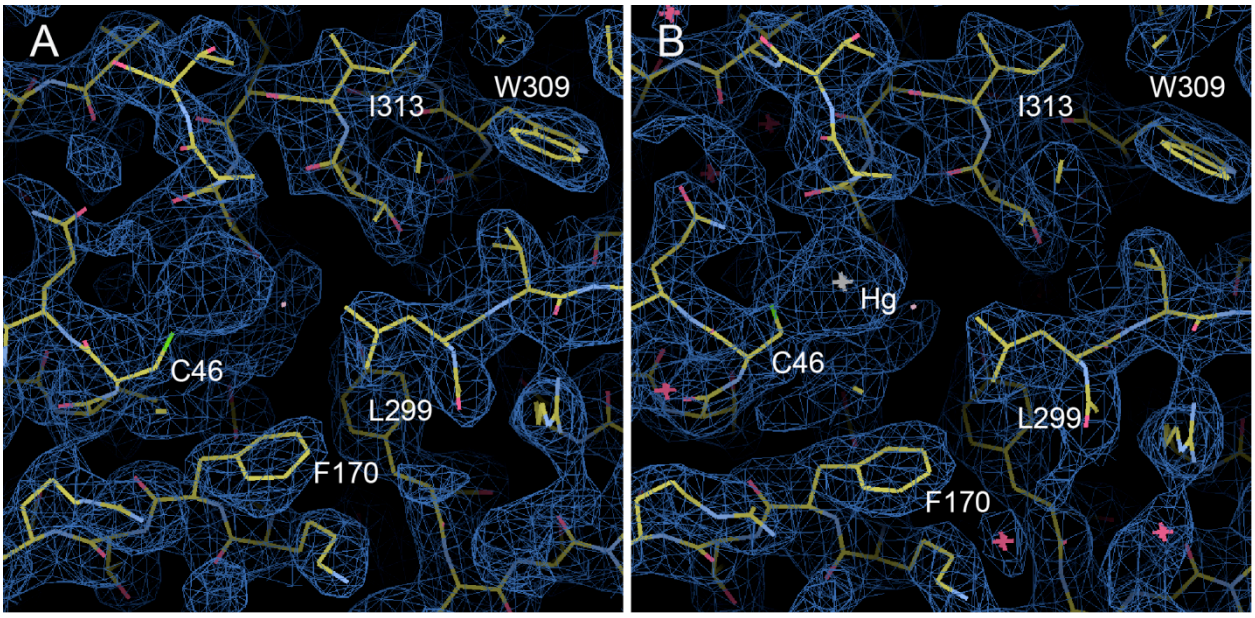


Figure 2.4. Representative experimental electron density maps.

(*A*) Electron density map contoured at 1σ level after SIRAS phasing with the HKL2MAP program at 2.6 Å resolution. This map covers the area around one of the Hg atoms near C46. A partial polypeptide model automatically built with ARP/wARP is shown. (*B*) Electron density of the same area after the final cycle of refinement against the same 2.8 Å resolution data.

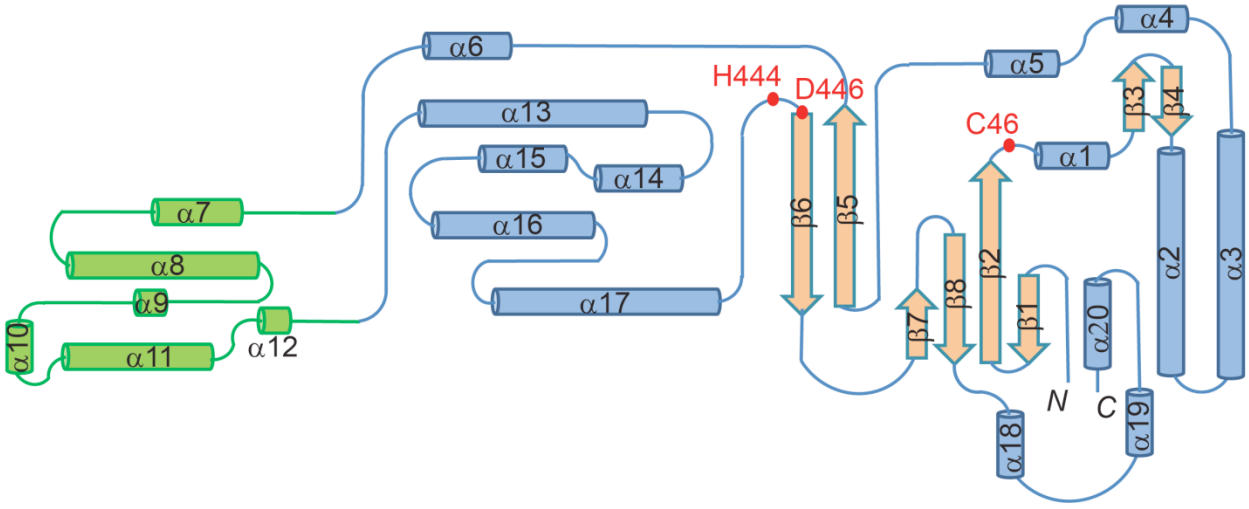


Figure 2.5. Schematic diagram of the secondary structure topologies of the SNL domain of SidC. Alpha helixes are represented by cylinders and beta strands are represented by arrows. The predicted catalytic triad are labeled and marked by red circles.

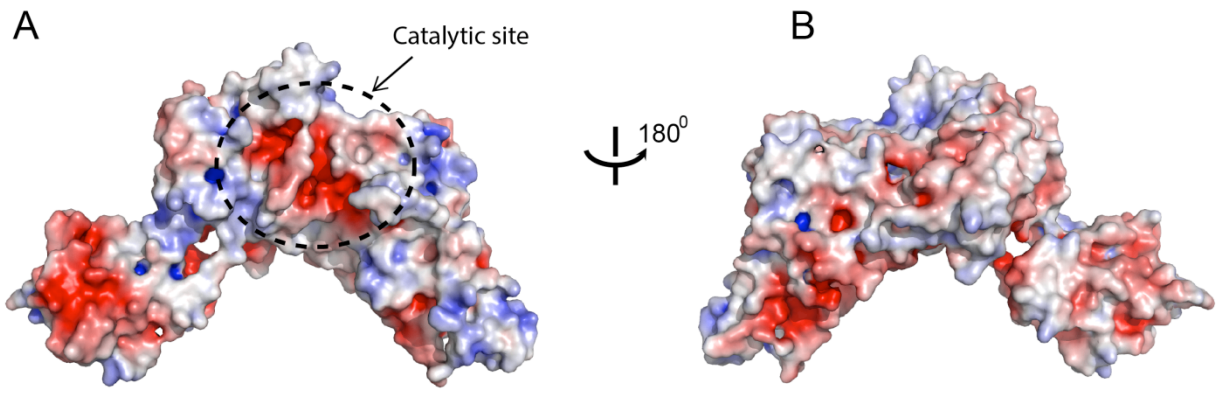


Figure 2.6. Molecular surface of the SNL domain of SidC.

(*A*) Front view of the SNL domain. The orientation of this view is the same as that in Fig. 1B and D. The dash-lined circle indicates the area containing the catalytic C46, H444, and D446 catalytic triad. (*B*) Back view of the domain. The surface is colored based on electrostatic potential with positively charged regions in blue (+5 kcal/electron) and negatively charged surface in red (-5 kcal/electron).

Table 2.1. Data collection and structural refinement statistics

A. Data collection statistics		
Space group	$P2_12_12_1$	
Cell dimensions	a = 61.53 Å, b = 133.44 Å, c = 170.25 Å, $\alpha = 90^0$, $\beta = 90^0$, $\gamma = 90^0$	
	Native	Hg
Synchrotron beam lines	MCCHESS A1	NSLS X4C
Wavelength (Å)	0.9789	0.9789
Maximum resolution (Å)	2.03	2.6
Observed reflections	400,367	651,503
Unique reflections	90,100	83,467
Completeness (%)[a]	95.2(93.0)	99.9(99.8)
$\langle I \rangle / \langle \sigma \rangle$[a]	25.0(1.74)	31.1(10.9)
R_{sym}[a,b] (%)	10.7(72.4)	5.9(17.9)
Number of Hg sites		4
B. Refinement statistics	Native	Hg
Resolution (Å)[a]	35.88-2.03(2.055-2.03)	36-2.59(2.66-2.59)
R_{crys} / R_{free} (%)[a,c]	21.2/24.8(28.8/31.0)	18.9/23.1(25.4/30.3)
Rms bond length (Å)	0.018	0.016
Rms bond angles (°)	1.767	1.637
Ramachandran plot		
Most favored/Additional (%)	97.4/2.6	97.1/2.9
Generous/Disallowed (%)	0/0	0/0

[a]Values in parenthesis are for the highest resolution shell.

[b]$R_{sym} = \Sigma_h \Sigma_i |I_I(h) - \langle I(h)| / \Sigma_h \Sigma_i I_I(h)$.

[c]$R_{crys} = \Sigma(|F_{obs}| - k|F_{cal}|) / \Sigma|F_{obs}|$. R_{free} was calculated for 5% of reflections randomly excluded from the refinement.

Ectopic expression of SidC alters the pattern of intracellular ubiquitinated species

The enzymatic activity of SidC was tested using commercially available cysteine protease kits or poly-ubiquitin chains as potential DUB substrates. However, we did not detect either kind of activity (Fig. 2.7). To interrogate the potential cysteine-related enzymatic activities of SidC, we co-expressed the GFP-tagged wild type SNL domain of SidC or its C46A mutant with human HA-tagged ubiquitin (HA-Ub) in 293T cells. Whole cell lysates were separated in SDS-PAGE gels and probed for HA. The pattern of ubiquitinated species was significantly different in cells expressing the wild type SNL domain from that in cells expressing the SNL C46A mutant or GFP control (Fig. 2.8A). Some ubiquitinated species were more prominent (indicated by *) while others were absent (indicated by arrow heads) in transfected cells (Fig. 2.8A). This result suggests that SidC affects the host ubiquitin pathway. To identify potential proteins with an altered ubiquitination state, we used the Stable Isotope Labeling by Amino acids in Cell culture (SILAC) method (Oda et al., 1999). HEK 293T cells were first co-transfected with HA-Ub and GFP-tagged SNL domain (either the wild type or its C46A mutant). After labeling the cells with heavy or light medium, cell lysates were prepared and ubiquitinated species were immuno-precipitated with anti-HA antibody coated beads. The samples were further analyzed by mass spectrometry (Fig. 2.9). Strikingly, SILAC analysis showed that 5 out of the 9 top hits with reduced ubiquitination in cells expressing the wild type SNL domain belong to the ubiquitin-conjugating enzyme E2 family (Table 2.2).

To further validate the mass spectrometry results, a similar experiment was performed and the immunoprecipitated proteins were solubilized with SDS with or without DTT and analyzed by western blot using specific antibody against the E2 enzyme UBE2L3/UbcH7. Indeed, the amount of ubiquitinated UbcH7 was significantly reduced in the presence of the wild

type SNL domain compared with the GFP control or the SNL C46A mutant (Fig. 2.8B). In the presence of DTT, the E2~Ub bands were shifted down relative to their corresponding E2 band (Fig. 2.8B, last three lanes). This observation implies that the linkage between the C-terminal carboxyl group of the Ub molecule and E2 is not an isopeptide bond, which would be resistant to DTT treatment. Instead, the Ub of the immunoprecipitated E2~Ub appears to be linked via a thioester bond to the catalytic cysteine of the E2, which is readily reduced by DTT (Fig. 2.8B, last three lanes). These data further suggest that SidC is not a DUB, which cleaves the isopeptide linkage between Ub and target proteins or within Ub chains. Given the fact that the ectopically expressed SNL domain can discharge the activated Ub (linked to the E2 through a thioester bond) from the E2, we hypothesize that SidC functions as a ubiquitin ligase that efficiently hijacks the charged Ub from E2s and then attaches the Ub moiety to other SidC-specific targets. In agreement with this hypothesis, ectopic expression of the SNL domain did not change the total levels of UbcH7; instead, it only exhausted the pool of E2 charged with ubiquitin (Fig. 2.10). This hypothesis also explains the observation shown in Fig. 2.8A that some ubiquitinated species were diminished (presumably due to competition with endogenous E3s for charged E2s), while others appeared more prominent in the presence of the SNL domain (presumably due to the ubiquitin ligase activity of SidC).

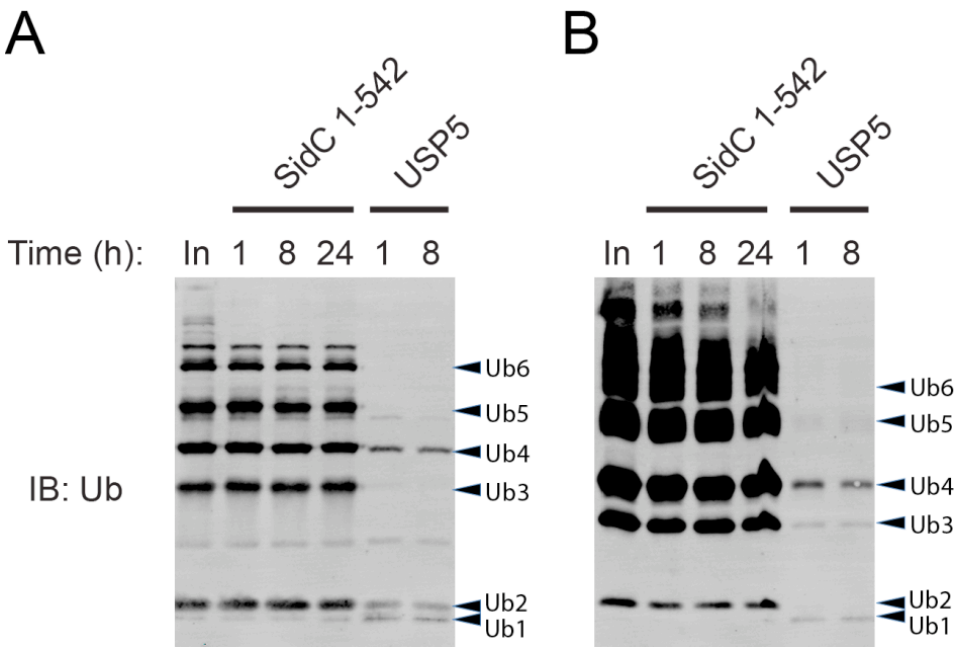


Figure 2.7. In vitro deubiquitination assay of SidC.

(*A*) Western blot of reaction products of SidC with mixture of K48 Ub(1-6) poly-ubiquitin chains. (*B*) Similar reactions with K63 poly-ubiquitin chains. Incubation with wild type SidC from 1 to 24 hours. USP5 (isopeptidase T, from Boston Biochem) was used as a positive control. In: Input.

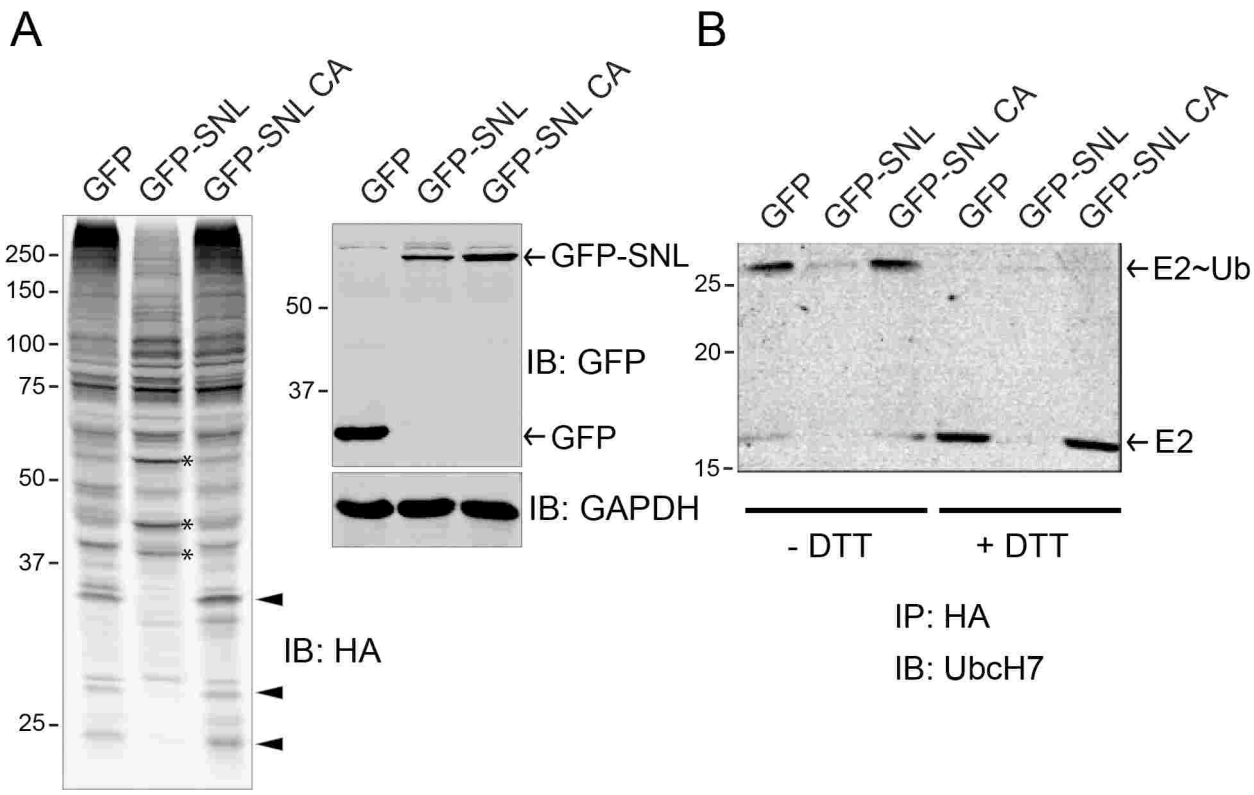

A
GFP
GFP-SNL
GFP-SNL CA
250
150
100
75
50
37
25
*
*
*
IB: HA
GFP
GFP-SNL
GFP-SNL CA
50
37
←GFP-SNL
IB: GFP
←GFP
IB: GAPDH
B
GFP
GFP-SNL
GFP-SNL CA
GFP
GFP-SNL
GFP-SNL CA
25
20
15
←E2~Ub
←E2
- DTT
+ DTT
IP: HA
IB: UbcH7

Figure 2.8. Ectopic expression of SidC altered the intracellular ubiquitination pattern.

(*A*) HEK293T cells were co-transfected with HA-ubiquitin and GFP control, GFP tagged wild type SidC SNL domain, or its C46A mutant. Whole cell lysates were prepared and analyzed by western blot with anti-HA (left panel), anti-GFP, and anti-GAPDH as a loading control (bottom right). Arrow heads highlight several bands that are positive in GFP and GFP-SNL C46A controls but are diminished in the presence of wild type SNL domain. * denotes bands that are more prominent in the sample expressing wild type SNL domain. (*B*) HEK 293T cells were co-transfected with HA-ubiquitin and other indicated plasmids. Cells lysates were prepared and incubated with anti-HA beads to immunoprecipitate HA-Ub tagged species. The precipitated samples were prepared in SDS loading buffer without (three lanes on the left) or with (three lanes on the right) DTT and subjected to western blot analysis. In the presence of DTT, the E2~Ub complex was fully reduced to E2, indicating a thioester linkage between E2 and ubiquitin.

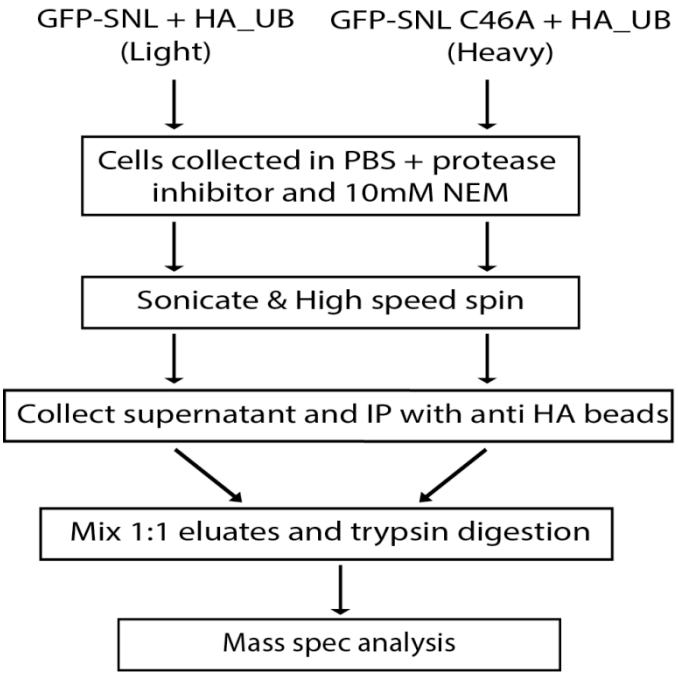


Figure 2.9. Experimental flow chart of SILAC sample preparation.

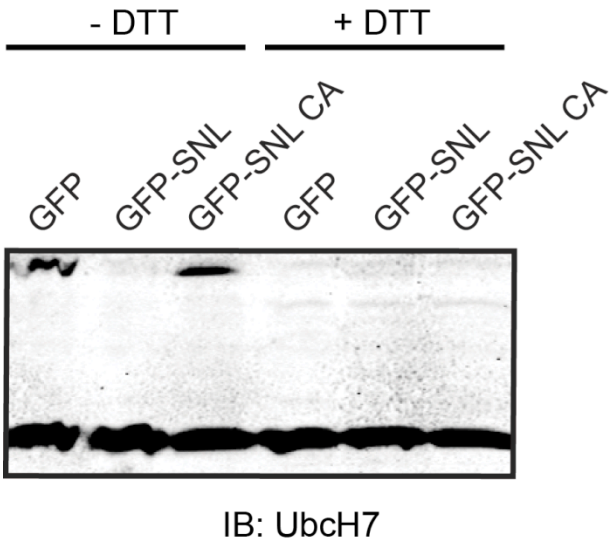


Figure 2.10. The protein levels of the ubiquitin conjugating E2 enzyme, UbcH7, are not changed in the presence of the SNL domain of SidC. 293T cells were co-transfected with HA-ubiquitin and other indicated plasmids. Cells lysates were prepared in the presence (three left lanes) or absence of DTT (three lanes on the right) and analyzed by specific UbcH7 antibody.

Table 2.2. A list of SILAC hits of the proteins with altered ubiquitination.

	experiment 1 Wt = light; mut = heavy		experiment 2 Wt = heavy; mut = light	
protein	peptide count	gmean of log2 ratio (Wt/mut)	peptide count	gmean of log 2 ratio (mut/Wt)
UBE2NL	9	-3.60566	4	3.08539
UBE2L3/UbcH7	11	-3.46593	12	6.35531
UBE2K/Ubc1	10	-2.64236	7	5.3087
UBE2S	9	-2.56502	7	3.80895
UBA1/Ube1	97	-2.53918	97	2.70132
BRAP	4	-2.37871	2	4.29822
UBE2T/HSPC150	9	-2.02673	3	2.64006
USP5/ISOT	35	-2.02175	34	3.8251
AUP1	8	-1.34807	2	2.34722

The N-terminal domain of SidC/SdcA is a ubiquitin E3 ligase

To test whether SidC is a ubiquitin E3 ligase, we performed in vitro ubiquitination assays using purified recombinant proteins (see Materials and Methods). Assays were performed with the SNL domain of SidC (aa. 1-542), E1, and a set of 4 representative E2s (bovine E2-25K, human Cdc34/UBE2R1, UbcH7/UBE2L3, and UbcH5c/UBE2D3). As expected, in the absence of SidC, di-ubiquitin (di-ub) chains were formed in the presence of E2-25K or Cdc34, demonstrating that the purified ubiquitin E1 and E2 enzymes are functional (Fig. 2.11). By contrast, characteristic ubiquitin ladders were observed when the wild type SNL domain was incubated with E2-25K, UbcH7, and UbcH5c but were barely visible with Cdc34 (Fig. 2.12A). Since SidC appeared most active in the presence of UbcH7 (Fig. 2.12A), which was also among the top SILAC hits (Table 2.2), we used UbcH7 in all our following ubiquitin E3 ligase assays. Next, we examined the ubiquitin ligase activity of SidC mutants, in which the proposed catalytic triad was disrupted. Compared to the wild type SNL domain protein, both the H444A and D446A showed reduced activity, whereas the activity was completely abolished in the C46A mutant (Fig. 2.12B).

To analyze the nature of the poly-ubiquitinated species, time-dependent reactions with physiologically relevant amounts of enzymes were performed. The time-course experiments showed the gradual appearance of high molecular weight poly-ubiquitinated species and a small amount of di-Ub chains. The majority of these high molecular weight ubiquitin species was retained in the stacking gel (Fig. 2.12C). Probing the blot with a SidC-specific antibody revealed that the ubiquitin species retained in the stacking gel and at the top part of the separation gel were products of SidC autoubiquitination (Fig. 2.12D). Western blot with antiserum against UbcH7 also revealed poly- or multiple mono- ubiquitination of this E2 (Fig. 2.12E). These

ubiquitinated E2s contributed to the ubiquitin signals derived from ubiquitin species in the middle range of molecular weights observed in Fig. 2.12C. E3 ligases often display autoubiquitination in in vitro assays, and in most instances this activity is considered as a mechanistically relevant activity readout (Pickart et al., 2001). In the case of SidC, autoubiquitination was not detected under infection conditions (Fig. 2.13).

We further compared the ubiquitin ligase activity of the SNL domain (1-542) with the full-length SidC and the full-length SidC paralog SdcA. Both the SNL domain and the full-length SidC protein showed comparable activity (Fig. 2.14). Intriguingly, SdcA exhibited a ubiquitin ligase activity, but with a different E2 preference. Unlike SidC, SdcA efficiently catalyzed ubiquitin polymerization in the presence of UbcH5 (Fig. 2.15). This observation suggests that *L. pneumophila* encodes two seemingly redundant genes in order to maximize its ability to hijack the host ubiquitin system. Together, these data demonstrate that SidC and its paralog SdcA are *bona fide* ubiquitin ligases that have a broad and non-overlapping specificity for ubiquitin-conjugating E2 enzymes.

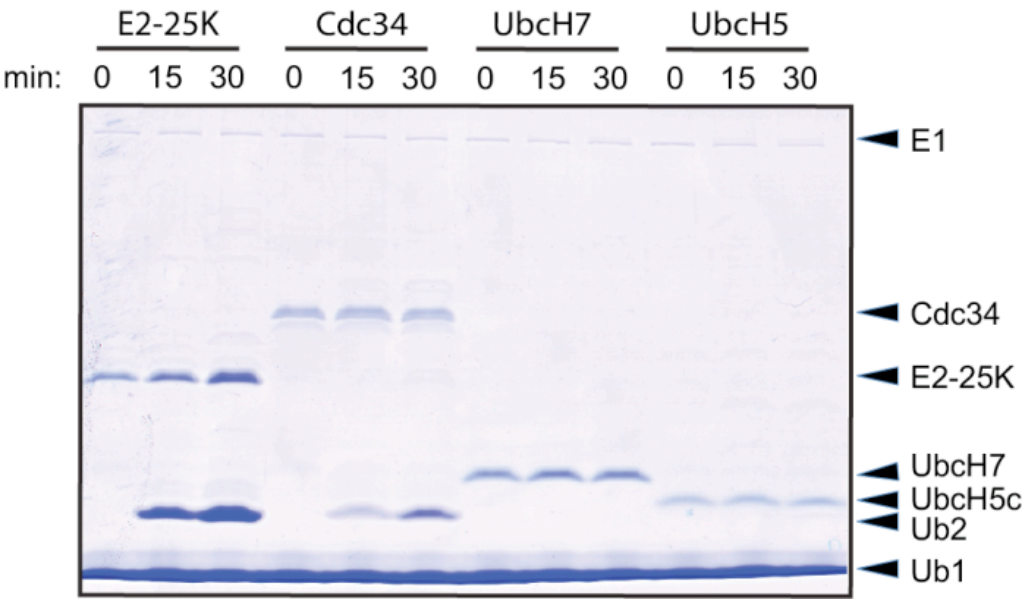


Figure 2.11. SDS-gel analysis of ubiquitin activation enzyme E1, ubiquitin conjugation E2 enzymes, including 4 representative E2s: E2-25K, Cdc34, UbcH7, and UbcH5.

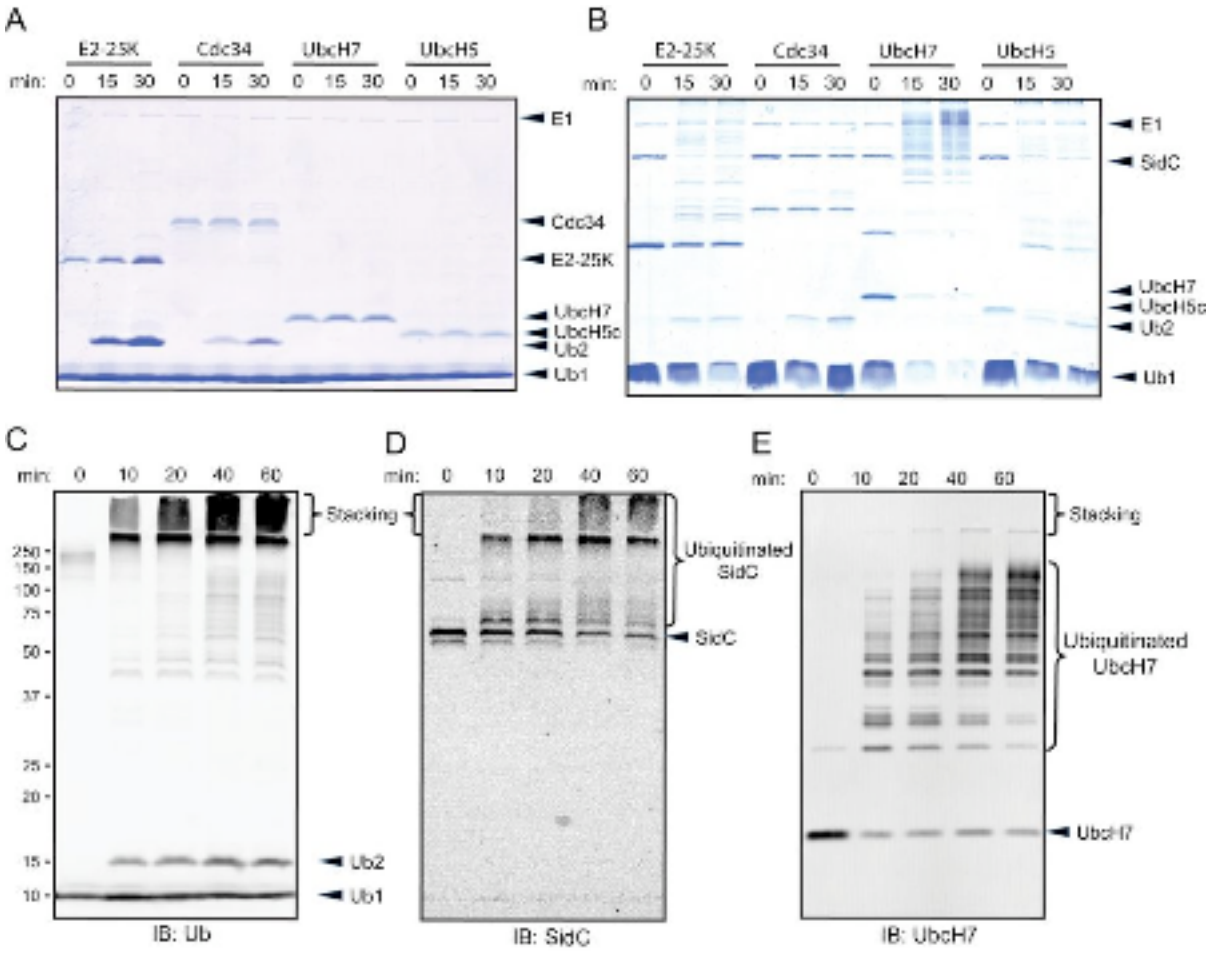

A
E2-25K
Cdc34
UbcH7
UbcH5
min: 0 15 30 0 15 30 0 15 30 0 15 30
E1
Cdc34
E2-25K
UbcH7
UbcH5c
Ub2
Ub1
B
E2-25K
Cdc34
UbcH7
UbcH5
min: 0 15 30 0 15 30 0 15 30 0 15 30
E1
SidC
UbcH7
UbcH5c
Ub2
Ub1
C
min: 0 10 20 40 60
Stacking
250
150
100
75
50
37
25
20
15
10
Ub2
Ub1
IB: Ub
D
min: 0 10 20 40 60
Stacking
Ubiquitinated SidC
SidC
IB: SidC
E
min: 0 10 20 40 60
Stacking
Ubiquitinated UbcH7
UbcH7
IB: UbcH7

Figure 2.12. The SNL domain of SidC has E3 ubiquitin ligase activity.

(*A*) In vitro ubiquitination assays with the SNL domain of SidC (aa. 1-542) and 4 representative E2s: E2-25K, Cdc34, UbcH7 and UbcH5. (*B*) E3 activity assay of SidC H444A, D446A, and C46A mutants. The activity of the C46A mutant was completely abolished while the activities of the H444A and D446A mutants were significantly reduced. (*C*) Time-dependent ubiquitination assay. The reactions were performed with E1, UbcH7, SNL domain, and wild type ubiquitin. The reaction mixtures were analyzed by western blot with anti-ubiquitin antibody. Within the stacking gel, an increased amount of heavily poly-ubiquitinated species was observed during the time course. (*D*) Western blot of the same materials as in (*C*) with anti-SidC antibody. The high molecular weight ubiquitin species were positive for SidC. (*E*) Western blot of the same samples as in (*C*) with anti-UbcH7. UbcH7 was also ubiquitinated during the in vitro reaction.

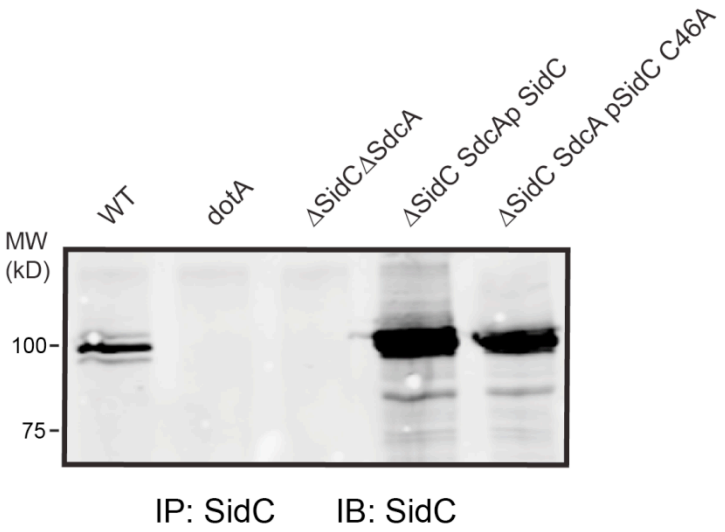


Figure 2.13. SidC was not autoubiquitinated under infection conditions.

$2X10^7$ U937 cells were infected with designated *Legionella* strains at moi = 5. After 2 hr infection, cells were collected and lysed by 50 μl 0.2% saponin for 1h at on ice. The supernatants were collected and immunoprecipitated with anit-SidC antibody. Precipitated materials were analyzed by western blot using anti-SidC antibody.

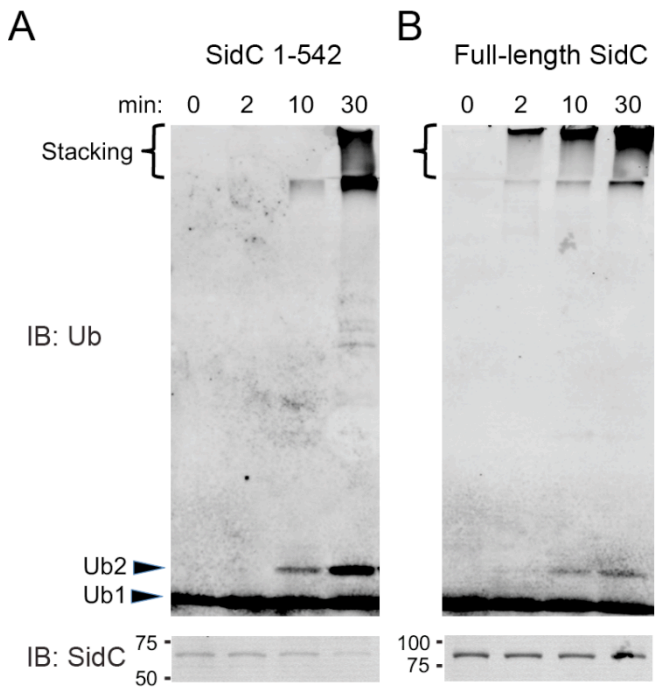


Figure 2.14. Ubiquitin ligase activity assay of the SNL domain and the full length SidC.

(*A*) Time-dependent in vitro ubiquitination of the SNL domain of SidC (1-542). (*B*) Time-dependent in vitro ubiquitination of the SidC full length (1-917). (Note that the full length SidC has a comparable activity as that of its N-terminal SNL domain).

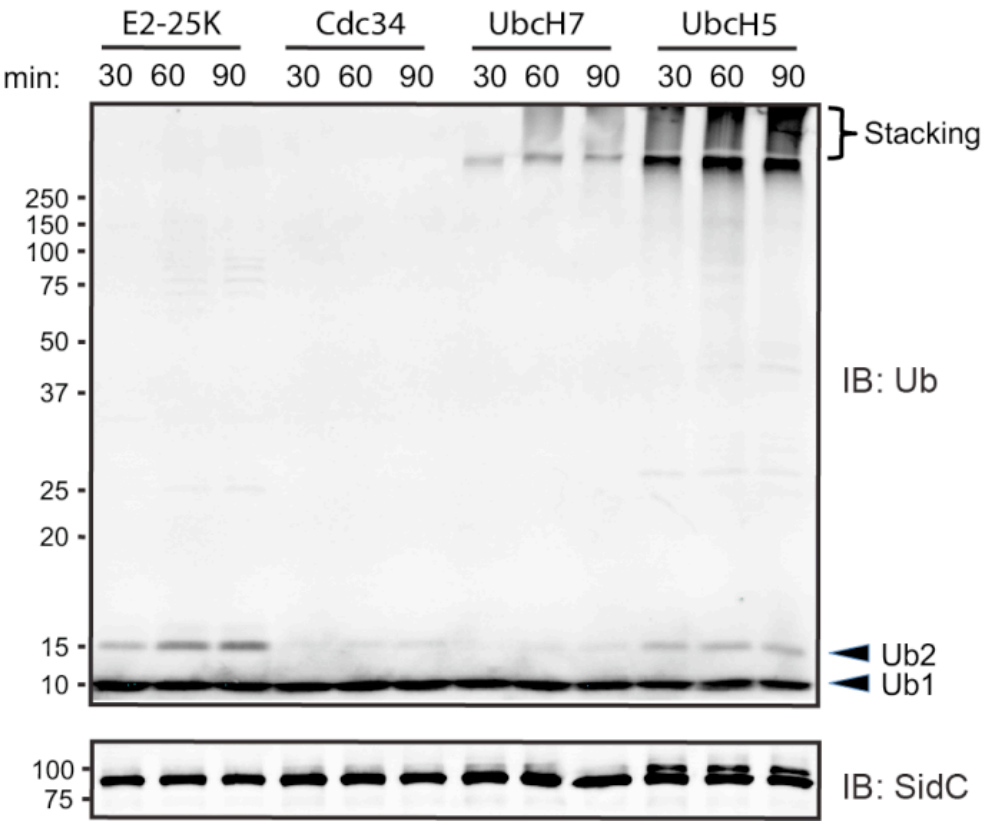


Figure 2.15. In vitro ubiquitination assays with full length SdcA (1-908) and four representative E2s: E2-25K, Cdc34, UbcH7 and UbcH5. SdcA was detected using the poly-clonal antibody against full length SidC, which cross-reacts with SdcA. (Note that unlike SidC, SdcA prefers UbcH5 for efficient poly-ubiquitin chain assembly).

The N-terminal domain of SidC catalyzes multiple types of ubiquitin chain linkage

We next investigated the linkages of the heavily poly-ubiquitinated species made by SidC. The linkage within Ub changes is formed between the C-terminal carboxyl group of a distal Ub and the ε-amino group on a lysine or the N-terminal amino group of the proximal Ub. Ub has seven lysine residues in addition to the N-terminal methionine, implying eight possible linkages in poly-Ub chains. We first tested the activity of SidC with seven ubiquitins, each carrying a single lysine to arginine mutation. Similar to wild type ubiquitin, all single K to R ubiquitin mutants produced high molecular weight poly-ubiquitin species that were retained within the stacking gel. By contrast, a Ub mutant carrying all seven K to R mutation (K0) did not show evidence of poly-Ub formation (Fig. 2.16A). These data suggest that SidC is capable of catalyzing the formation of poly-Ub chain through multiple lysine residues, but not via the N-terminal methionine. To further dissect the linkage preference of SidC, we assayed the activity of SidC with ubiquitin mutants containing six K-to-R mutations with only one native lysine remaining. Although less efficient compared to wild type ubiquitin, high molecular weight poly-ubiquitinated species did form in the presence of Ub-11K and Ub-33K(Ub mutant with only one native lysine at positions 11 and 33, respectively). Much weaker poly-Ub signals were observed with mutant Ub-63K and Ub-48K (Fig. 2.16B). These results indicate that SidC is able to use multiple lysine residues for ubiquitin polymerization with a preference for K11 and K33-linked chains.

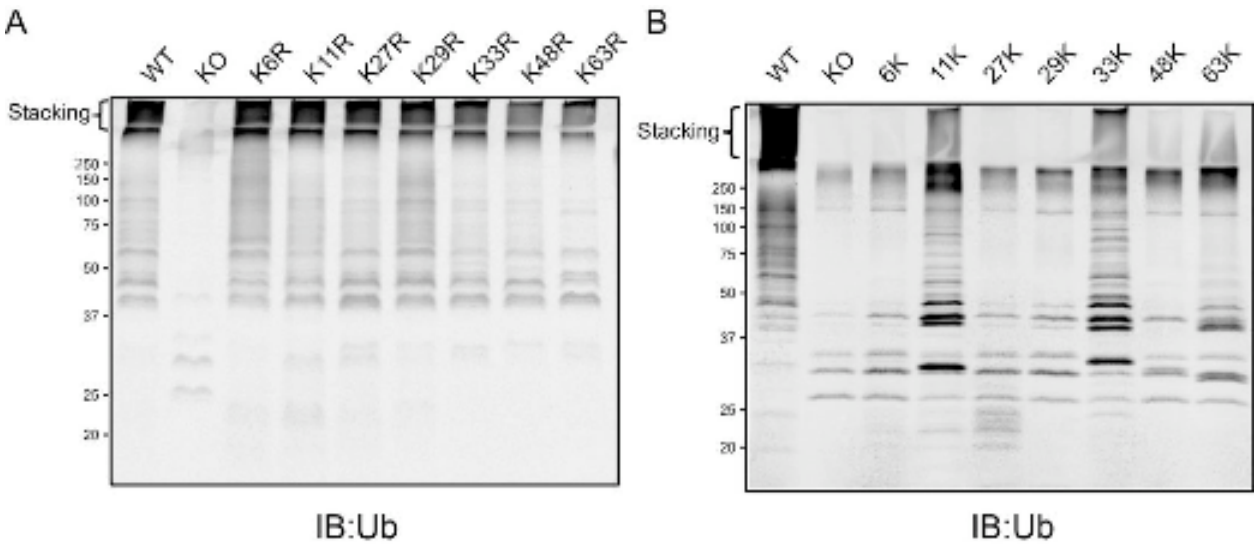


Figure 2.16. The ubiquitin linkage preference by the SNL domain of SidC.

(*A*) In vitro ubiquitination assays using E1, UbcH7, SNL domain, and a set of ubiquitin mutants with a single K-to-R mutation. K0 represents the ubiquitin mutant carrying all seven K-to-R to mutations. (*B*) In vitro ubiquitination assays using E1, UbcH7, SNL domain, and a set of ubiquitin containing a single native lysine with six K-to-R mutations.

The ubiquitin ligase activity of SidC is required for the recruitment of ER proteins and ubiquitin to the LCV

SidC and its paralog SdcA have been shown to recruit ER vesicles to the LCV (Ragaz et al., 2008). We thus investigated the relevance of the E3 ligase activity of SidC in *Legionella* infection. A *Dictyostelium discoideum* strain stably expressing GFP-HDEL (Liu & Luo et al., 2007) (a GFP fusion with the ER retention marker HDEL) was infected with relevant *Legionella* strains and the bacteria were immunostained with an anti-*Legionella* antibody. In *D. discoideum*-infected with wild type *L. pneumophila*, more than 50% of the LCVs were positive for GFP-HDEL 2 hrs after uptake (Fig. 2.17A and B). In contrast, such recruitment did not occur in infections with a Dot/Icm deficient strain. The association of GFP-HDEL with the LCVs created by the *ΔsidC-sdcA* mutant was significantly reduced. Importantly, this defect was almost fully restored by a plasmid expressing wild type SidC but not by its catalytically inactive C46A mutant (Fig. 2.17A and B). The C46A mutation did not have detectable effects on the expression nor the translocation of the protein by the bacteria (Fig. 2.18). These results demonstrate that the ubiquitin ligase activity of SidC is critical for its role in the recruitment of ER components to the bacterial phagosome.

It has been reported that *L. pneumophila* recruits polyubiquitin conjugates around the bacterial phagosomes shortly after infection (Dorer et al., 2006). This recruitment is nearly abolished in the infection with the Δ*sidC-sdcA* mutant strain (Horenkamp et al., 2014). We tested whether the E3 ligase activity of SidC/SdcA is responsible for the recruitment of ubiquitin conjugates. U937 cells were infected with relevant *Legionella* strains and the bacteria were immunostained with an anti-*Legionella* antibody and the specific poly-ubiquitin antibody FK1. The recruitment of poly-ubiquitin species was observed in cells infected with wild type bacteria

but not with the Δ*dotA* or the Δ*sidC-sdcA* mutants. This defect was reversed by expression of wild type but not of the catalytically inactive C46A mutant in a plasmid. Previous studies showed that in strain AA100/130b, the F-box protein AnkB appears to be essential for the recruitment of ubiquitin materials to LCVs (Price et al., 2009). However, here we showed that in strain Lp02 (Berger & Isberg et al., 1993), the recruitment of poly-ubiquitin species was not impaired. (Fig. 2.19). Thus, the E3 ligase activity of SidC/SdcA, but not AnkB, is required for the recruitment of ubiquitinated species to the LCV at least in strain Lp02.

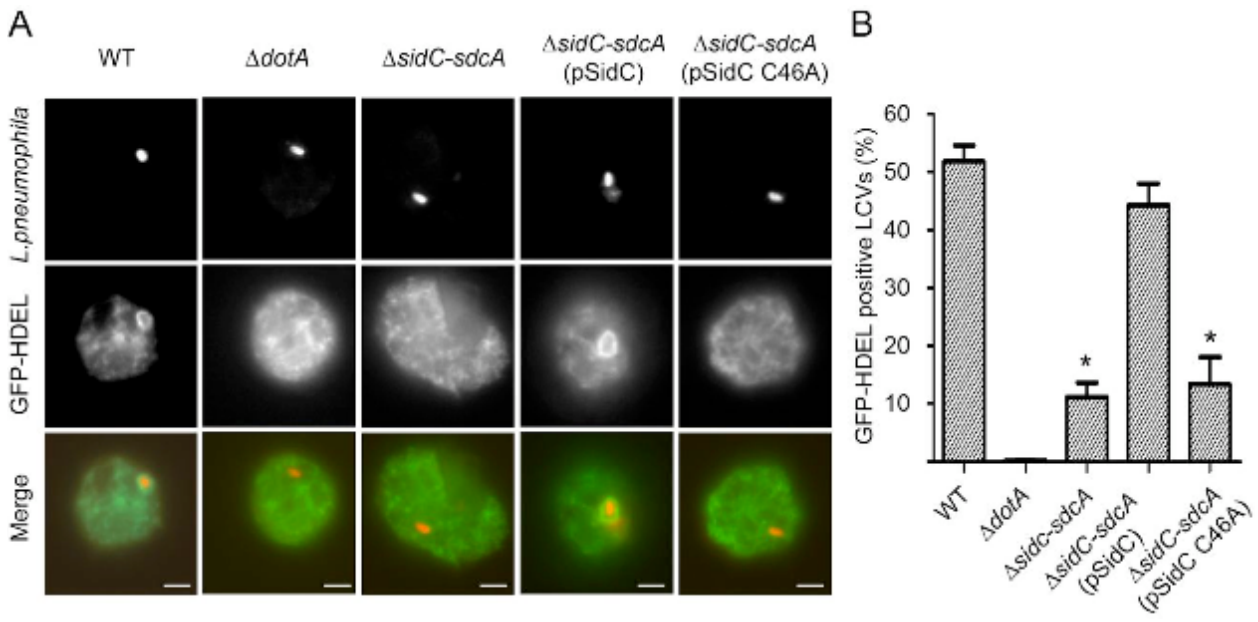


Figure 2.17. The E3 ubiquitin ligase activity is required for the ER marker recruitment to the bacterial phagosome.

(*A*) Images show the recruitment of the ER marker GFP-HDEL (green) to the LCVs in *D. discoideum* cells infected with the indicated *Legionella* strains (red). Scale bars, 2 μm. WT: *L. pneumophila* Philadelphia-1 wild type strain Lp02; *dotA*: the type IV secretion system defective strain Lp03; Δ*sidC-sdcA*: the *SidC* and *SdcA* double deletion mutant of the Lp02 strain, Δ*sidC-sdcA*(pSidC) and Δ*sidC-sdcA*(pSidC C46A): Δ*sidC-sdcA* strain complemented with a plasmid expressing wild type or C46A mutant SidC. (*B*) Percent of cells containing GFP-HDEL positive LCVs counted from three independent assays under the conditions infected with the indicated *Legionella* strains. *P < 0.01 compared to the infection with wild type *Legionella* strain.

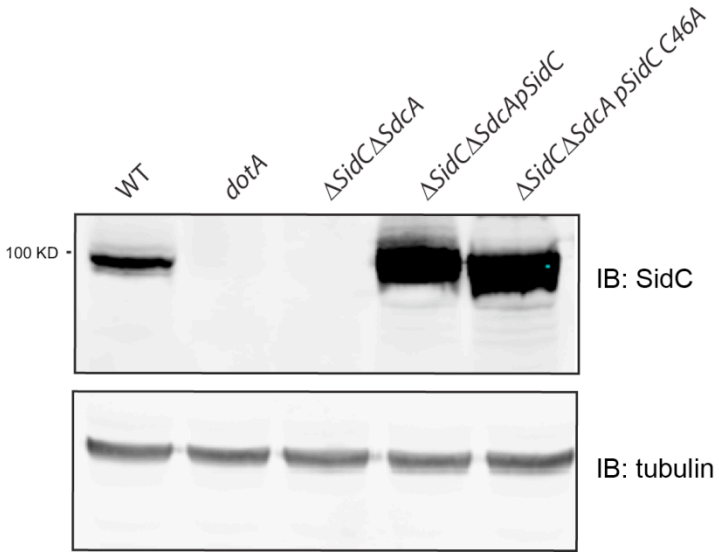


Figure 2.18. SidC translocation after *Legionella* infection.

$2X10^7$ U937 cells were infected with designated *Legionella* strains at moi = 5. After 2 hr infection, cells were collected and lysed by 50 µl 0.2% saponin for 1h at on ice. After 10 min centrifugation at 12,000g, the supernatants were collected and mixed with SDS loading buffer. These samples were blotted by SidC antibody to detect the levels of SidC protein translocated into the host cells (top panel). Tubulin was used as a loading control (bottom panel).

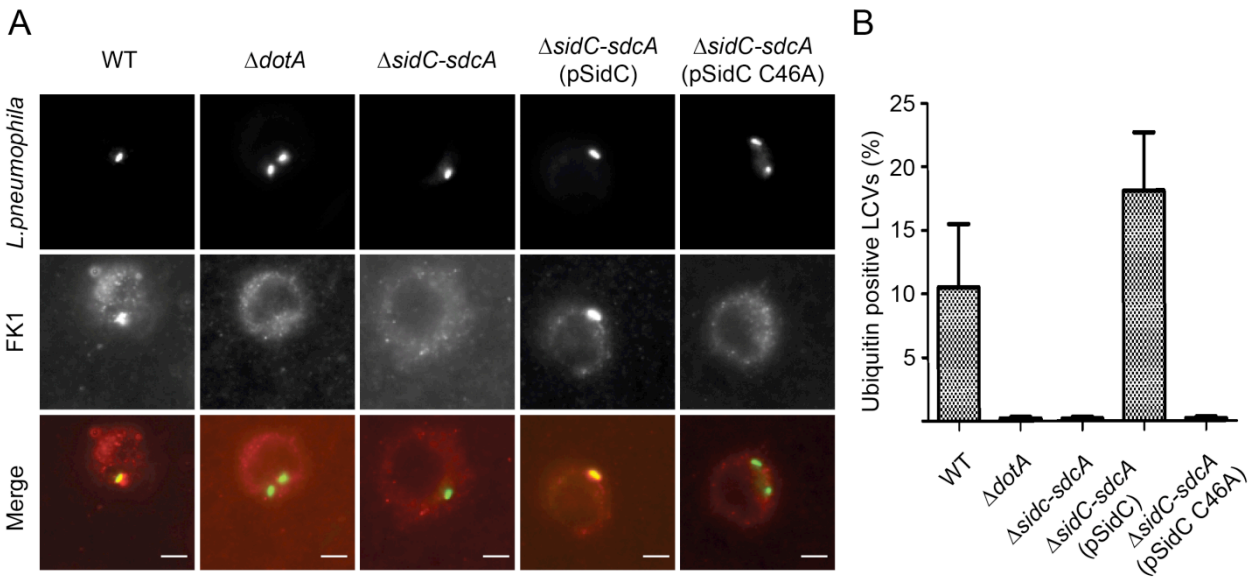


Figure 2.19. The E3 ubiquitin ligase activity is required for the recruitment of ubiquitin conjugates to the bacterial phagosome.

(A) Images show the recruitment of ubiquitinated species (red) to the LCVs in U937 cells infected with the indicated *Legionella* strains (green). Scale bars, 3 μm. WT: *L. pneumophila* Philadelphia-1 wild type strain Lp02; dotA: the type IV secretion system defective strain Lp03; Δ*sidC-sdcA*: the SidC and SdcA double deletion mutant of the Lp02 strain; Δ*sidC-sdcA*(pSidC) and Δ*sidC-sdcA*(pSidC C46A): Δ*sidC-sdcA* strain complemented with a plasmid expressing wild type or C46A mutant SidC. (B) Percent of cells containing ubiquitin positive LCVs counted from three independent assays under the conditions infected with the indicated *Legionella* strains.

Discussion

Through the Dot/Icm transporter, *L. pneumophila* delivers approximately 300 experimentally verified substrates into the host (Zhu et al., 2011; Lifshitz et al., 2013). However, it has been a vast challenge to assign them an exact function in infection due to the scarcity of conserved functional motifs within these effector proteins. The *Legionella* effector SidC was such an example. SidC and its paralog SdcA were first assigned as vesicle tethering factors for promoting the recruitment and fusion of ER-derived vesicles to the bacterial phagosome (Ragaz et al., 2008). Recently, crystal structures of the N-terminal portion of SidC were reported (Horenkamp et al., 2014; Gazdag et al., 2014). In both publications, the N-terminal portion of SidC was concluded to function as an ER vesicle tethering factor. Here, we determined the crystal structure of the N-terminal portion of SidC (1-542), which we named SNL (SidC N-terminal E3 Ligase) domain. The structure of the SNL domain is very similar to the recently reported structures with an rmsd of less than 1.2 Å to both structures. However, by careful structural and biochemical analysis, we discovered that the N-terminal SNL domain of SidC possesses a cysteine based ubiquitin ligase activity.

Compared to other cysteine based ubiquitin ligases, the SNL domain family has some unique features. First, the SNL domain has no detectable primary and tertiary structural similarity to any known protein, suggesting a novel fold of this domain, as also suggested in the two previous publications (Horenkamp et al., 2014; Gazdag et al., 2014). Second, the SNL domain has a conserved catalytic C-H-D triad. The aspartic and histidine residues within the triad likely render the catalytic cysteine a stronger nucleophile than just a single catalytic cysteine found in other ubiquitin ligases. This feature may enhance the kinetics of Ub transfer from E2s to SidC and may also be advantageous during the competition with host Ub ligases for

the pool of activated E2~Ub. Third, in HECT E3s, the catalytic cysteine is located at the C-terminal small lobe, which can move a large distance towards the N-terminal lobe through a hinge-like motion. This conformational flexibility is believed to allow the transfer of Ub from E2 to the E3 catalytic cysteine (Huang et al., 1999; Kamadurai et al., 2009; Verdecia et al., 2003). However, the catalytic cysteine of the SNL domain is localized at the center of a surface of its main sub-domain (Fig. 2.1B and D). This structural organization suggests that the Ub transfer by the SNL domain may not involve large movements of the catalytic cysteine, which further indicates that the binding site for E2~Ub is in close proximity to the catalytic cysteine. To address this hypothesis, future structural and biochemical experiments are warranted to identify the E2 binding site on the SNL domain. Nevertheless, given these unique features, we conclude that the SNL domain of SidC defines a novel family of ubiquitin E3 ligases.

Although SidC and SdcA share 72% sequence identity, they exhibit differential preference for ubiquitin E2 conjugating enzymes. This discrimination of E2s suggests that these two proteins have evolved to maximally exploit components in the host ubiquitin pathway. This idea could explain why *L. pneumophila* has maintained two such highly similar proteins in evolution.

Modulation of multiple host cellular processes is essential for the intracellular life cycle of *L. pneumophila* (Hubber & Roy et al., 2010). Given the fact that ubiquitination regulates a myriad of cellular functions, including cell cycle, cell death, trafficking, and immune responses (Hershko & Ciechanover et al., 1998), it is not surprising that *L. pneumophila* codes for at least six proteins containing F-box domains or U-box domains, which are hallmarks of multi-component E3 ligase complexes (Hubber & Nagai et al., 2013). Among these, LubX and LegU1 are *bona fide* E3 ligases that target SidH, a Dot/Icm substrate involved in cell death, and BAT3, a

protein involved in the cell cycle, respectively (Ensminger et al., 2010; Kubori & Nagai et al., 2008). Our observation that the E3 ligase activity of SidC is essential for its role in the recruitment of ER proteins to the bacterial phagosome indicates its role in modulating vesicle trafficking, which is distinct from that of LubX and LegU1. However, the intriguing question of how the E3 ubiquitin activity of SidC connects with ER vesicle trafficking remains to be explored. It is likely that SidC/SdcA rewires the functions of specific host proteins by ubiquitination. Recent data have shown that the small GTPase Rab1, which regulates the vesicular trafficking step from the ER to the *cis*-Golgi, is mono-ubiquitinated when macrophage cells are infected with wild type *L. pneumophila*, but not with a Δ*sidC-sdcA* mutant strain (Horenkamp et al., 2014). However, SidC apparently does not ubiquitinate Rab1 directly, since co-transfection of SidC and Rab1 fails to cause mono-ubiquitination of Rab1 (Horenkamp et al., 2014). In agreement with this observation, we also did not detect direct ubiquitination of Rab1 by SidC. When wild type SidC was co-transfected with either GFP-tagged wild type Rab1, the dominant negative S22N-Rab1, or the constitutive active Q70L-Rab1, no ubiquitinated Rab1 was detected (Fig. 2.20A). Furthermore, in vitro ubiquitination experiments using recombinant GST-tagged Rab1 failed to detect direct ubiquitination of Rab1 (Fig. 2.20B). These data suggest that the mono-ubiquitination of Rab1 is mediated through an indirect unknown mechanism as proposed by the previous report (Horenkamp et al., 2014).

An exciting and challenging future direction is to identify specific host factors that are ubiquitinated by SidC/SdcA. We expect that proteins involved in the host ER-related membrane trafficking events are the potential targets of SidC/SdcA. The identification of these unknown factors would not only help explain the molecular mechanisms of pathogen-host interactions, but

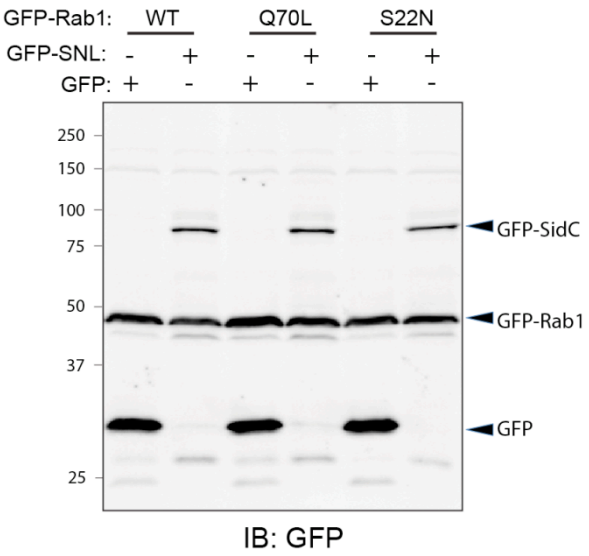


Figure 2.20. Rab1 is not ubiquitinated by SidC.

GFP-tagged wild type Rab1, its constitutive active form (Q70L), or dominant negative form (S22N) were co-transfected with the SNL domain of SidC GFP control in 293T cells. Cell lysate were prepared and analyzed by western blot using anti-GFP antibody.

also would enhance our knowledge of basic cellular processes, in particular, membrane trafficking.

Materials and Methods

Cloning and mutagenesis. PCR products for SidC (aa. 1-542), full length SidC (aa. 1-917), and full-length SdcA (aa. 1-908) amplified from *L. pneumophila* genomic DNA were digested with BamHI and XhoI restriction enzymes and inserted into a pET28a-based vector in frame with an N-terminal His-SUMO tag (Zuo et al., 2005). Single amino acid substitution of SidC was introduced by in vitro site-directed mutagenesis using oligonucleotide primer pairs containing the appropriate base changes. For mammalian expression, corresponding fragments of SidC were PCR subcloned into pEGFP-C1 vector. The plasmid pZL199 (VanRheenen et al., 2006) inserted with wild type or C46A mutant SidC was used to complement the *ΔsidC-sdcA* mutant (Luo & Isberg et al., 2004). PCR products for all single lysine-containing human ubiquitin mutants were amplified from pET30a-hUb containing the corresponding mutations (from Dr. Shu-bing Qian, Cornell University). PCR products for all single lysine to arginine human ubiquitin mutants were amplified from pcDNA-His-hUb containing the corresponding mutation. All ubiquitin DNA products were digested with NdeI and XhoI restriction enzymes and ligated into pET21a plasmid digested with the same enzymes. All constructs were confirmed by DNA sequencing.

Protein Expression and Purification. Tni insect cell line was used for expression of E1 proteins. The E1 protein was first affinity purified by cobalt resins (Clonetech) followed by gel filtration chromatography. For protein expression, *E. coli* Rosetta strains harboring the

expression plasmids were grown in Luria-Bertani medium supplemented with 50 μg/ml kanamycin to mid-log phase. Protein expression was induced for overnight at 18°C with 0.1 mM isopropyl-B-D-thiogalactopyranoside (IPTG). Harvested cells were resuspended in a buffer containing 20 mM Tris-HCl (pH 8.0), 200 mM NaCl, and protease inhibitor cocktail (Roche) and were lysed by sonication. Soluble fractions were collected by centrifugation at 18,000 rpm for 20 min at 4°C and incubated with cobalt resins (Clonetech) for 1 h at 4°C. Protein bound resins were extensively washed with lysis buffer. The SUMO-specific protease Ulp1 was then added to the resin slurry to release SidC from the His-SUMO tag. Eluted protein samples were further purified by FPLC size exclusion chromatography. The peak corresponding to SidC was pooled and concentrated to 10 mg/ml in a buffer containing 20 mM Tris, pH 7.7, and 200 mM NaCl, 14 mM β-mercaptoethanol. For ubiquitin preparation, bacterial cells expressing ubiquitin or its derivatives were harvested in a buffer containing 20 mM ammonium acetate (pH 5.1) and were lysed by sonication. Soluble fractions were collected by centrifugation at 18,000 rpm for 20 min at 4°C. The resulting lysates were treated with acetic acid to adjust the pH to 4.7. The solutions were centrifuged at 18,000 rpm for 10 min to remove the white flocculent precipitant. The pH of the final clear lysates was re-adjusted to 5.1 with NaOH. The supernatant was collected and purified by cation exchange HiTrap SP column (GE healthcare) with the buffer gradient from 20 mM ammonium acetate (pH 5.1) to 0.5 M ammonium acetate (pH 5.1). Ubiquitin peaks were pooled and further purified by size-exclusion chromatography in 20 mM Tris pH 8.0 and 50 mM NaCl. The expression plasmids encoding E2-25K (Raasi & Pickart et al., 2005) , hUbcH7, hUbcH5 and hCdc34 were first purified as His-sumo tagged fusion by cobalt resins and the tagged was removed by Ulp1 protease. A size exclusion column was used for the final stage purification of these E2 enzymes.

Crystallization. Crystals were grown at room temperature by the hanging-drop vapor diffusion method by mixing 1 μl of protein (10 mg/ml) with an equal volume of reservoir solution containing 0.1 M cacodylate pH 5.6, 7.5% PEG6000, and 10 mM DTT. Rod-shaped crystals were formed within 2-3 days. For phase determination, protein crystals were soaked in cryoprotectant (0.1 M cacodylate pH 5.6, 7.5% PEG6000, and 25% (v/v) glycerol) with the addition of 10 mM ethylmercury chloride (kind gift from Dr. Steve Ealick at Cornell University) for 10 min at room temperature.

Data collection and processing. Diffraction data sets for native protein crystals were collected at the Cornell synchrotron light source, MacCHESS beam line A1. Data set for mercury derivative SidC crystals was collected at Brookhaven National Laboratory, X4C beamline. All data sets were indexed, integrated and scaled with HKL-2000 (Otwinowski & Minor et al., 1997). The crystals belong to space group $P2_12_12_1$ with a = 68.64 Å; b = 134.45 Å; c = 172.68 Å; $\alpha = \beta = \gamma = 90^0$ (Table 2.1). The calculated Matthews coefficient V_m= 3.32 and with 62.9% of solvent in the crystal and two protein molecules in an asymmetric unit (Matthews et al., 1968).

Structure determination and refinement. Four mercury sites corresponding to residues C17 and C46 in both molecules were identified in the crystal using the program HKL2MAP (Pape & Schneider et al., 2004). The initial phase was calculated by single isomorphous replacement with anomalous scattering (SIRAS) method and was improved by solvent flattening in HKL2MAP. The *ab initio* protein model was then built with ARP/wARP program (Langer et al., 2008) Iterative cycles of model building and refinement were carried out with the program COOT

(Emsley & Cowtan et al., 2004). and the refmac5 program (Murshudov et al., 1997) in the CCP4 suite (Collaborative Computational Project N, 1994) to complete the final model.

In vitro deubiquitination assay. Lys-48- or Lys-63-linked polyubiquitin chains (Ub_{1-7}) were purchased from Boston Biochem. The in vitro de-ubiquitination reactions were carried out in a buffer containing 50mM Tris (pH 8.0), 50mM NaCl, 1mM EDTA, and 5 mM DTT. The reaction mixture has a final volume of 20 µl with a final concentration of 40 ng/µl ubiquitin chains and with 1 µg of the SNL domain of SidC and 1 µg of USP5 (positive control; from Boston Biochem). The reactions were stopped at the indicated time points with SDS loading dye and the samples were separated in a 12% acrylamide gel and western blotted with anti-ubiquitin (Covance).

In vitro E3 ubiquitin ligase assay. Ubiquitination assays were performed at 37°C in the presence of 50 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 0.5 mM DTT, 50 mM creatine phosphate (Sigma P7396), 3 U/ml of pyrophosphatase (Sigma I1643), 3 U/ml of creatine phosphokinase, 150 nM (or 100 nM) E1, 200 nM (or 100 nM) E2, 0.5 µM (or 200 nM) SidC and 100 µM (or 1 µM) ubiquitin. All reactions were stopped by the addition of 5X SDS-PAGE loading buffer containing 250 mM BME and analyzed by either Coomassie Brilliant blue stain or Western blot with mouse anti-ubiquitin, rabbit anti-UbcH7 (BostonBiochem) and anti-SidC.

Cell culture and transfection. HEK293T cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS (Cellgro), and 0.1% Pen/Strep (Cellgro) at 37°C in 5% CO_2 atmosphere. Cells were co-transfected with pCDNA3-HA-Ubiquitin (Addgene) and

pEGFP-C1, pEGFP-C1-SidC-542 or pEGFP-C1-SidC-542 CA plasmids for 24 hours. Transfection was performed using polyethyleneimine (PEI) reagent. Cells were harvested with 1X SDS-PAGE loading buffer containing 100 mM BME. The samples were subsequently probed with mouse anti-HA (Sigma), rabbit anti-GFP or mouse anti-GAPDH.

SILAC labeling and immunoprecipitation. HEK293T cells were grown for 2 weeks in complete media containing normal lysine and arginine ("light") or [$^{13}C_6$,$^{15}N_2$] lysine and [$^{13}C_6$,$^{15}N_4$] arginine ("heavy", Sigma) before proceeding to DNA transfection. For HA-immunoprecipitation, cells were collected in lysis buffer containing 20 mM Tris pH 7.5, 150 mM NaCl, 1% Triton and EDTA-free protease inhibitor cocktail (Roche). Soluble fractions were collected by centrifugation at 18,000 rpm for 15 min at 4°C and incubated with anti-HA affinity gel (Sigma) for 2 h at 4°C. Immunoprecipitates were eluted in 1% SDS and 100 mM Tris pH 8.0, boiled for 5 min and then precipitated with 49.9% acetone, 50% ethanol, and 0.1% acetic acid for removal of SDS. Proteins were digested with trypsin (Promega) and desalted in a C18 column. The peptides were dried in a speed-vac. The final sample was dissolved in 0.1% trifluoroacetic acid, and analyzed by LC-MS/MS using a Q-Exactive mass spectrometer (Thermo). Data analysis was carried out using a Sorcerer system (Sage-N) running Sequest for protein identification and Xpress for peptide quantitation. The protocol was adapted from Ohouo et al. (Ohouo et al., 2010).

***Legionella* strains and infection.** Strains of *L. pneumophila* used include the wild type Lp02 (Berger & Isberg et al., 1993), the Dot/Icm deficient Lp03 (Berger & Isberg et al., 1993) and the *sidc-sdcA* mutant (Luo & Isberg et al., 2004). *Dictyostelium discoideum* strain AX4 stably

expresseing HDEL-GFP was cultured at 21.5 ºC in HL-5 medium supplemented with penicillin and streptomycin (100 U/ml), and 10 μg/ml of G418. *D. discoideum* cells were seeded onto poly-lysine coated coverslips at $2x10^5$ cells per well. After incubation for 2 hrs at 25 ºC, cells were infected with *L. pneumophila* grown to post-exponential phase for 2 hrs at an MOI of 2. To detect translocated SidC, 2×10^7 U937 cells were plated onto 10-cm petri dish 12 hours before infection. Cells were infected with post-exponential *L. pneumophila* strains for 2 hours at an MOI of 5. Cells collected by centrifugation were resuspended and lysed in 50 μL of PBS containing 0.2% saponin. After 30 min incubation on ice, lysates were cleared by centrifugation at 10,000 *g* for 10 min at 4 ºC. The supernatants were collected analyzed by SDS-PAGE followed by Western blot with appropriate antibodies. U937 cells were cultured and prepared for infection as described in the Materials and Methods section of our manuscript. For ubiquitin recruitment assay, $2X10^5$ of U937 cells were seeded into 24-well plates with coverslips 12 hours before infection. Cells were infected with the indicated *L. pneumophila* strains for 2 hours at an MOI of 2.

Antibodies, immunostaining and Western blot. Anti-*L. pneumophila* (Xu et al., 2010) and anti-SidC (Luo & Isberg et al., 2004) were described previously. Anti-tubulin antibody was purchased from (DSHB, University of Iowa). Infected *D. discoideum* samples were fixed and stained as described earlier (Liu & Luo et al., 2007). For ubiquitin immunofluorescent analyses, U937 cells were fixed and stained by standard procedures (McLean & Nakane et al., 1974). Anti-*L. pneumophila* antibodies were used at a dilution of 1:20,000. Poly-ubiquitinated proteins were stained by FK1 antibody at a dilution of 1:50 followed by 555-conjugated goat anti-mouse IgM (Invitrogen, Carlsbad, CA). Recruitment of GFP-HDEL and poly-ubiquitinated proteins was

examined by Olympus IX-81 fluorescence microscope. Images obtained from an Orca camera were processed with the IPlab software package (Scanalytic, Inc. Fairfax, VA). Western blots were performed following standard protocols (Xu et al., 2010); anti-SidC and anti-tubulin antibodies were used at 1:10,000; 1: 20,000 and 1:5000, respectively.

ACKNOWLEDGEMENTS

We thank Dr. S.E. Ealick (Cornell Univ.) for heavy atom containing reagents. Dr. S. Qian (Cornell Univ.) for ubiquitin constructs. We also thank Drs. S. D. Emr, V. Vogt (Cornell Univ.) for critical reading of the manuscript. This work was supported by the Harry Samuel Mann Award (F.H.) and National Institutes of Health (NIH) Grants R01-GM094347 (Y.M.), K02AI085403 (Z.-Q.L) and R56AI103168 (Z.-Q.L). R01-GM102529 (J.J.), the Welch Foundation grant AU-1711 (J.J.), R01-GM097272 (M.S.), and the American Cancer Society #RSG-11-146-01-DMC (M.S.). The X-ray data were collected at MacChess beamline A1 and NSLS beamline X4C. 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.

AUTHOR CONTRIBUTIONS

Y.M. and Z.Q.L. conceived the project. F.H., X.L., J.Q., Y.T., J.J., M.S., and Y.M. performed the experiments. F.H., X.L., J.Q., Y.T., J.J., M.S., and Y.M. analyzed the data. F.H., Z.Q.L., and Y.M. wrote the paper.

REFERENCES

Barton GJ (1993) ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng* 6(1):37-40.

Berger KH & Isberg RR (1993) Two distinct defects in intracellular growth complemented by a single genetic locus in Legionella pneumophila. *Mol Microbiol* 7(1):7-19.

Cazalet C*, et al.* (2004) Evidence in the Legionella pneumophila genome for exploitation of host cell functions and high genome plasticity. *Nat Genet* 36(11):1165-1173.

Chen ZJ & Sun LJ (2009) Nonproteolytic functions of ubiquitin in cell signaling. *Mol Cell* 33(3):275-286.

Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. *Acta Cryst.* D(50):760-763.

de Felipe KS*, et al.* (2005) Evidence for acquisition of Legionella type IV secretion substrates via interdomain horizontal gene transfer. *J Bacteriol* 187(22):7716-7726.

Deshaies RJ & Joazeiro CA (2009) RING domain E3 ubiquitin ligases. *Annu Rev Biochem* 78:399-434.

Diao J, Zhang Y, Huibregtse JM, Zhou D, & Chen J (2008) Crystal structure of SopA, a Salmonella effector protein mimicking a eukaryotic ubiquitin ligase. *Nat Struct Mol Biol* 15(1):65-70.

Dorer MS, Kirton D, Bader JS, & Isberg RR (2006) RNA interference analysis of Legionella in Drosophila cells: exploitation of early secretory apparatus dynamics. *PLoS Pathog* 2(4):e34.

Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.

Ensminger AW & Isberg RR (2010) E3 ubiquitin ligase activity and targeting of BAT3 by multiple Legionella pneumophila translocated substrates. *Infect Immun* 78(9):3905-3919.

Fields BS, Benson RF, & Besser RE (2002) Legionella and Legionnaires' disease: 25 years of investigation. *Clin Microbiol Rev* 15(3):506-526.

Gazdag EM, Schobel S, Shkumatov AV, Goody RS, & Itzen A (2014) The structure of the N-terminal domain of the Legionella protein SidC. *J Struct Biol*.

Haglund K & Dikic I (2012) The role of ubiquitylation in receptor endocytosis and endosomal sorting. *J Cell Sci* 125(Pt 2):265-275.

Hershko A & Ciechanover A (1998) The ubiquitin system. *Annu Rev Biochem* 67:425-479.

Holm L & Rosenstrom P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38(Web Server issue):W545-549.

Horenkamp FA*, et al.* (2014) Legionella pneumophila Subversion of Host Vesicular Transport by SidC Effector Proteins. *Traffic*.

Hsu F*, et al.* (2012) Structural basis for substrate recognition by a unique Legionella phosphoinositide phosphatase. *Proc Natl Acad Sci U S A* 109(34):13567-13572.

Huang L*, et al.* (1999) Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science* 286(5443):1321-1326.

Hubber A & Roy CR (2010) Modulation of host cell function by Legionella pneumophila type IV effectors. *Annu Rev Cell Dev Biol* 26:261-283.

Hubber A, Kubori T, & Nagai H (2013) Modulation of the Ubiquitination Machinery by Legionella. *Curr Top Microbiol Immunol*.

Hurley JH & Stenmark H (2011) Molecular mechanisms of ubiquitin-dependent membrane traffic. *Annu Rev Biophys* 40:119-142.

Isberg RR, O'Connor TJ, & Heidtman M (2009) The Legionella pneumophila replication vacuole: making a cosy niche inside host cells. *Nat Rev Microbiol* 7(1):13-24.

Janjusevic R, Abramovitch RB, Martin GB, & Stebbins CE (2006) A bacterial inhibitor of host programmed cell death defenses is an E3 ubiquitin ligase. *Science* 311(5758):222-226.

Jiang X & Chen ZJ (2012) The role of ubiquitylation in immune defence and pathogen evasion. *Nat Rev Immunol* 12(1):35-48.

Kagan JC & Roy CR (2002) Legionella phagosomes intercept vesicular traffic from endoplasmic reticulum exit sites. *Nat Cell Biol* 4(12):945-954.

Kamadurai HB*, et al.* (2009) Insights into ubiquitin transfer cascades from a structure of a UbcH5B approximately ubiquitin-HECT(NEDD4L) complex. *Mol Cell* 36(6):1095-1102.

Kubori T, Hyakutake A, & Nagai H (2008) Legionella translocates an E3 ubiquitin ligase that has multiple U-boxes with distinct functions. *Mol Microbiol* 67(6):1307-1319.

Kubori T, Shinzawa N, Kanuka H, & Nagai H (2010) Legionella metaeffector exploits host proteasome to temporally regulate cognate effector. *PLoS Pathog* 6(12):e1001216.

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

Lifshitz Z*, et al.* (2013) Computational modeling and experimental validation of the Legionella

and Coxiella virulence-related type-IVB secretion signal. *Proc Natl Acad Sci U S A* 110(8):E707-715.

Liu Y & Luo ZQ (2007) The Legionella pneumophila effector SidJ is required for efficient recruitment of endoplasmic reticulum proteins to the bacterial phagosome. *Infect Immun* 75(2):592-603.

Lomma M*, 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(9):1272-1291.

Luo ZQ & Isberg RR (2004) Multiple substrates of the Legionella pneumophila Dot/Icm system identified by interbacterial protein transfer. *Proc Natl Acad Sci U S A* 101(3):841-846.

Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* 33(2):491-497.

McLean IW & Nakane PK (1974) Periodate-lysine-paraformaldehyde fixative. A new fixation for immunoelectron microscopy. *J Histochem Cytochem* 22(12):1077-1083.

Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3):240-255.

Oda Y, Huang K, Cross FR, Cowburn D, & Chait BT (1999) Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A* 96(12):6591-6596.

Ohouo PY, Bastos de Oliveira FM, Almeida BS, & Smolka MB (2010) DNA damage signaling recruits the Rtt107-Slx4 scaffolds via Dpb11 to mediate replication stress response. *Mol Cell* 39(2):300-306.

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

Pape T & Schneider TR (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Cryst.* 37:843-844.

Pickart CM (2001) Mechanisms underlying ubiquitination. *Annu Rev Biochem* 70:503-533.

Price CT*, et al.* (2009) Molecular mimicry by an F-box effector of Legionella pneumophila hijacks a conserved polyubiquitination machinery within macrophages and protozoa. *PLoS Pathog* 5(12):e1000704.

Raasi S & Pickart CM (2005) Ubiquitin chain synthesis. *Methods Mol Biol* 301:47-55.

Ragaz C*, et al.* (2008) The Legionella pneumophila phosphatidylinositol-4 phosphate-binding type IV substrate SidC recruits endoplasmic reticulum vesicles to a replication-permissive vacuole. *Cell Microbiol* 10(12):2416-2433.

Reyes-Turcu FE, Ventii KH, & Wilkinson KD (2009) Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. *Annu Rev Biochem* 78:363-397.

Rotin D & Kumar S (2009) Physiological functions of the HECT family of ubiquitin ligases. *Nat Rev Mol Cell Biol* 10(6):398-409.

Scheffner M, Nuber U, & Huibregtse JM (1995) Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature* 373(6509):81-83.

Shao F, Merritt PM, Bao Z, Innes RW, & Dixon JE (2002) A Yersinia effector and a Pseudomonas avirulence protein define a family of cysteine proteases functioning in bacterial pathogenesis. *Cell* 109(5):575-588.

Sievers F*, et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.

Singer AU*, et al.* (2008) Structure of the Shigella T3SS effector IpaH defines a new class of E3 ubiquitin ligases. *Nat Struct Mol Biol* 15(12):1293-1301.

Tilney LG, Harb OS, Connelly PS, Robinson CG, & Roy CR (2001) How the parasitic bacterium Legionella pneumophila modifies its phagosome and transforms it into rough ER: implications for conversion of plasma membrane to the ER membrane. *J Cell Sci* 114(Pt 24):4637-4650.

Toulabi L, Wu X, Cheng Y, & Mao Y (2013) Identification and structural characterization of a Legionella phosphoinositide phosphatase. *J Biol Chem* 288(34):24518-24527.

VanRheenen SM, Luo ZQ, O'Connor T, & Isberg RR (2006) Members of a Legionella pneumophila family of proteins with ExoU (phospholipase A) active sites are translocated to target cells. *Infect Immun* 74(6):3597-3606.

Verdecia MA*, et al.* (2003) Conformational flexibility underlies ubiquitin ligation mediated by the WWP1 HECT domain E3 ligase. *Mol Cell* 11(1):249-259.

Weber SS, Ragaz C, Reus K, Nyfeler Y, & Hilbi H (2006) Legionella pneumophila exploits PI(4)P to anchor secreted effector proteins to the replicative vacuole. *PLoS Pathog* 2(5):e46.

Xu L*, et al.* (2010) Inhibition of host vacuolar H+-ATPase activity by a Legionella pneumophila effector. *PLoS Pathog* 6(3):e1000822.

Zhu W*, et al.* (2011) Comprehensive identification of protein substrates of the Dot/Icm type IV transporter of Legionella pneumophila. *PLoS One* 6(3):e17638.

Zuo X*, et al.* (2005) Enhanced expression and purification of membrane proteins by SUMO fusion in Escherichia coli. *J Struct Funct Genomics* 6(2-3):103-111.

CHAPTER III

Structure of the *Legionella* virulence factor, SidC reveals a unique PI(4)P-specific binding domain essential for its targeting to the bacterial phagosome

Xi Luo[1], David J. Wasilko[1], Yao Liu[2], Jiayi Sun[2], Xiaochun Wu[1], Zhao-Qing Luo[2], and Yuxin Mao[1]

[1]Weill Institute for Cell and Molecular Biology and Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853

[2]Department of Biological Sciences, Purdue University, 915 West State Street, West Lafayette, Indiana 47907, USA.

Xi Luo contributed significantly to Figure 3.1-3.13 and wrote and prepared the manuscript.

Atomic coordinates and structure factors for the reported structures have been deposited into the Protein Data Bank under the accession codes 4ZUZ.

Chapter III was originally published in PLOS Pathogens 11(6):e1004965. (2015)

Abstract

The opportunistic intracellular pathogen *Legionella pneumophila* is the causative agent of Legionnaires' disease. *L. pneumophila* delivers nearly 300 effector proteins into host cells for the establishment of a replication permissive compartment known as the *Legionella*-containing vacuole (LCV). SidC and its paralog SdcA are two effectors that have been shown to anchor on the LCV via binding to phosphatidylinositol-4-phosphate [PI(4)P] to facilitate the recruitment of ER proteins to the LCV. We recently reported that the N-terminal SNL (SidC N-terminal E3 Ligase) domain of SidC is a ubiquitin E3 ligase and its activity is required for the recruitment of ER proteins to the LCV. Here we report the crystal structure of SidC (1-871). The structure revealed that SidC contains four domains that are packed into an arch-like shape. The P4C domain (PI(4)P binding of SidC) is comprised of a four α-helix bundle and covers the ubiquitin ligase catalytic site of the SNL domain. Strikingly, a pocket with characteristic positive electrostatic potentials is formed at one end of this bundle. Liposome binding assays of the P4C domain further identified the determinants of phosphoinositide recognition and membrane interaction. Interestingly, we further found that binding with PI(4)P stimulates the E3 ligase activity, presumably due to a conformational switch induced by PI(4)P from a closed form to an open active form. Mutations of key residues involved in PI(4)P binding significantly reduced the association of SidC to the LCV and abolished its activity in the recruitment of ER proteins and ubiquitin signals, highlighting that PI(4)P-mediated targeting of SidC is critical for its function in the remodeling of bacterial phagosome membrane. Finally, a GFP-fusion with the P4C domain was demonstrated to be specifically localized to PI(4)P-enriched compartments in mammalian cells. This domain shows potential to be developed as a sensitive and accurate PI(4)P probe in living cells.

Author Summary

The Legionnaires' disease is caused by the intracellular bacterial pathogen *Legionella pneumophila*. Successful infection by this bacterium requires a special secretion system that injects nearly 300 effector proteins into the cytoplasm of host cells. The effector SidC and its paralog SdcA anchor on the *Legionella*-containing vacuole (LCV) and are important for the recruitment of ER proteins to the LCV. Recent data demonstrated that SidC and SdcA are ubiquitin E3 ligases and that their activity is required for the enrichment of ER proteins and ubiquitin conjugates on the LCV. Here we present the crystal structure of SidC revealing the architecture of a novel PI(4)P binding module. Biochemical and cell biological studies highlighted key determinants involved in PI(4)P binding and membrane insertion. Characterization of this novel PI(4)P binding module paved a potential avenue for the development of an accuratein vivo PI(4)P probe. Our data also revealed a distinct regulatory mechanism of the ubiquitin E3 ligase activity of SidC, which is activated by the lipid molecule, PI(4)P. Furthermore, our results suggest that proper spatial localization of SidC to the cytoplasmic surface of the bacterial phagosome through the binding with PI(4)P is crucial for its function.

Introduction

The gram-negative bacterium *Legionella pneumophila* is ubiquitously found in natural water systems, where it is a parasite of free-living amoebae (Fields et al., 2002). Upon inhalation of contaminated aerosols by susceptible individuals, this bacterium replicates within human macrophages and causes a severe pneumonia called Legionnaires' disease (Nash et al., 1984). During its interaction with host cells, *L. pneumophila* delivers approximately 300 effector proteins (Zhu et al., 2011) into the host via its Dot/Icm (defective organelle trafficking/ intracellular multiplication) type IV secretion system (Segal et al., 1998; Vogel et al., 1998). Through the collective action of this large array of virulence proteins, *L. pneumophila* co-opts various host processes to facilitate bacterial survival and replication within the host. Among the many host cellular pathways, host membrane trafficking is one of the most studied processes manipulated by *L. pneumophila* (Asrat et al., 2014).

Upon internalization, *L. pneumophila* forms a specialized membrane-bound compartment named the *Legionella*-containing vacuole (LCV). Previous data have shown that LCVs undergo a delicately programmed maturation process to evade fusion with lysosomes. The LCV is first covered by ER-derived vesicles at the early stage of uptake (Tilney et al., 2001), and later is enriched with resident ER proteins (Swanson et al., 1998; Kagan et al., 1998), and then studded with ribosome's (Horwitz et al., 1983). This biphasic maturation process converts the plasma membrane-derived LCV into an ER-like compartment and thus allows for the establishment of a replication-permissive niche within host cells (Isberg et al., 2009). The recruitment of ER and ER-derived vesicles requires sophisticated interactions between *Legionella* effectors and host factors. For example, Rab1, a Rab family small GTPase involved in vesicle trafficking between the ER and Golgi (Stenmark et al., 2009) has been shown to be

recruited to the LCV by SidM (DrrA) (Machner et al., 2006; Murata et al., 2006). Following Rab1 recruitment, a cascade of *Legionella* effectors, including SidM (Muller et al., 2010), AnkX (Mukherjee et al., 2011), SidD (Neunuebel et al., 2011; Tan et al., 2011), Lem3 (Tan et al., 2011), and LepB (Ingmundson et al., 2007), regulate the spatial and temporal dynamics of Rab1 through canonical GEF, GAP activities, as well as protein posttranslational modifications (Sherwood et al., 2013).

Lipid molecules, particularly phosphoinositides (PIs) play crucial roles during the transition of LCVs into ER-like compartments (Weber et al., 2014). PIs are a collection of lipids that have their inositol headgroup reversibly phosphorylated at the 3', 4' and 5' positions. Although PIs are minor components of cellular membranes, they play fundamental roles in a broad range of cell signaling and membrane trafficking events (Di Paolo et al., 2006). One of the PI species, phosphotidylinositiol 4-phosphate (or PI(4)P) has been shown to accumulate on the LCV (Weber et al., 2006). The enrichment of PI(4)P on the LCV is facilitated by both *Legionella* effectors and host PI metabolizing enzymes. SidF, the first *Legionella* protein shown to directly modify host PIs (Hsu et al., 2012), anchors on the LCV and specifically hydrolyzes PI(3,4)P_2 and PI(3,4,5)P_3 to PI(4)P and PI(4,5)P_2, respectively. Another *Legionella* encoded PI phosphatase SidP likely prevents the accumulation of PI(3)P on the LCV by hydrolyzing this lipid into phosphoinositol (Toulabi et al., 2013). The *Legionella* effector protein LpnE appears to recruit the host PI-5-phosphatase OCRL to the LCV, which converts PI(4,5)P_2 to PI(4)P (Weber et al., 2009). Meanwhile, host phosphatidylinositol 4-kinases (PI4Ks) also play a role in the establishment of PI(4)P-enriched vacuoles (Brombacher et al., 2009; Hubber et al., 2014). Lipid remodeling on the LCV is critical for the selective anchoring of effectors to its surface. The specific recruitment of the Rab1 modulator SidM to the LCV is mediated by a C-terminal unique

PI(4)P-binding P4M domain (Brombacher et al., 2009; Schoebel et al., 2010; Del Campo et al., 2014). Other LCV-localizing effectors, such as LidA (Machner et al., 2006) and LpnE (Weber et al., 2009) are likely also mediated by the binding to PI(4)P.

Among the currently identified PI(4)P-binding *Legionella* effectors, the protein SidC and its paralog SdcA were shown to anchor on the LCV through a 20 kDa PI(4)P-binding domain named P4C (PI(4)P binding of SidC) to facilitate the recruitment of ER proteins to the LCV (Ragaz et al., 2008). Deletion of *sidC* and *sdcA* resulted in a delay in the establishment of the replicative vacuole and a delayed appearance of ubiquitin signals on the LCV (Horenkamp et al., 2014). Recent structural studies of SidC revealed a novel N-terminal SNL domain that represents a unique family of ubiquitin E3 ligases (Hsu & Luo et al., 2014). Interestingly, the ubiquitin ligase activity is required for the early recruitment of ER vesicles and the ubiquitinated protein species to the LCV (Hsu & Luo et al., 2014). With the deciphering of the biochemical mechanisms of functional domains, an intriguing question emerged as to how the P4C domain of SidC targets and likely regulates the E3 ligase activity of the SNL domain.

Here we report the crystal structure of SidC (aa 1-871) containing the SNL, the P4C and a C-terminal undefined domain. Although the structure of the SNL domain is similar to the structure of the isolated SNL domain reported previously (Horenkamp et al., 2014; Hsu & Luo et al., 2014; Gazdag et al., 2014), significant conformational changes were observed, including a hinge motion between the SNL and the insertion domains, as well as a rearrangement of residues near the catalytic triad. We also revealed the structure of the P4C domain. Unlike any other known PI(4)P binding domain, the P4C domain of SidC is comprised of a four α-helix bundle and the PI(4)P binding site resides in a highly positively charged pocket formed at one end of the bundle. Further analyses of the P4C domain both in vitro and in vivo shed light on the molecular

mechanism of PI(4)P binding and LCV anchoring of SidC. Our data further provide a mechanistic background for the potential usage of the P4C domain as an accurate PI(4)P probe in general cell biology studies.

Results

Overall crystal structure of SidC

Full length SidC from the Philadelphia 1 strain of *L. pneumophila* is a 106 kDa protein comprised of 917 residues (Luo et al., 2004). For insights into the molecular mechanisms of the biological functions of SidC, we carried out structural studies of SidC. The N-terminal conserved SNL domain was crystallized and our structure-driven approach successfully revealed that the SNL domain is a novel E3 ligase (Hsu & Luo et al., 2014). We extended our structure studies to full length SidC. Although thin plate-shaped crystals were obtained, these crystals gave poor diffractions against X-ray beams. Crystals were also obtained with a truncated form of SidC (aa 1-871, from now on referred as SidC871) of which 46 residues were deleted from the C-terminus. These crystals diffracted up to 2.9 Å and the structure was solved by the molecular replacement method using our previously determined SidC (aa 1-542, SidC542) crystal structure as the search model. Residues of the C-terminal portion of SidC871 were built de novo to the model interspersed with iterative refinement. The final refined model contains two molecules of SidC in the asymmetric unit and each molecule contains all residues from 8 to 864. The structure was refined to 2.9 Å resolution with R_{work}/R_{free} = 22.7/28.1 and no significant stereochemistry violations (Figure 3.1 and Table 3.1).

The crystal structure of SidC871 reveals that SidC is comprised of four distinct domains including the N-terminal SNL domain, the INS domain (which is inserted within the SNL domain), the P4C domain, and a C-terminal unknown domain named CTD (Figure 3.2A-E). These four domains are packed in a single arch-like shape with the P4C and the CTD domains in close contact with the SNL domain (Figure 3.2B-E). Strikingly, the ubiquitin E3 ligase catalytic site (colored in red in Figure 3.2C and 3.2E) on the SNL domain is covered by the P4C domain. This observation suggests that the P4C domain is required to move away from the SNL domain in order to access the active site of the ubiquitin ligase. A plausible hypothesis is that the binding of PI(4)P by the P4C may cause the P4C domain to shift away from the catalytic site allowing the exposure of the ubiquitin ligase catalytic site. This hypothesis will be tested below.

Conformational dynamics of the SNL and INS domains

Structural comparison of the SNL and INS domains in SidC871 with the previously determined SidC542 structure, which contains only the SNL and INS domains (Hsu & Luo et al., 2014) revealed three regions with major conformational changes (Figure 3.3A). The first area is the loop (residues 42-49) that contains the catalytic cysteine C46 (Figure 3.3A and B). In the SidC542 structure, C46 is close to the D446 and H444 to form a catalytic triad. However, C46 in SidC871 is shifted away from H444 by about 6 Å (Figure 3.3B, Figure 3.1). This conformational switch involves large Psi-Phi angle changes of L42, T45, and C46 by the flipping of main chain peptide bonds (Figure 3.4). The second region is a non-conserved loop (residue 59-66) on the interface between the SNL domain and the P4C domain (Figure 3.3C). The conformational change of this loop makes room for the P4C domain to pack against the SNL domain. However, the biological role of the structural flexibility of this loop is not clear.

The third major conformational change occurs at the two flexible linkers between the SNL and INS domains. The SNL domain is rotated along a hinge (residue S224 and S328) by about 30^0 in SidC871 compared to the orientation in SidC542 (Figure 3.3A). This hinge bending motion is reminiscent of a similar displacement between the N- and the C-lobes of the HECT family ubiquitin ligases (Lin et al., 2012; Verdecia et al., 2003). These observations imply that the INS domain may interact with E2s in the ubiquitin conjugating reaction. To test this hypothesis, we used size exclusion chromatography to analyze the interaction between E2 and SidC. The catalytically inactive C85S mutant UbcH7 was stably charged with ubiquitin (UbcH7~Ub) and incubated with SidC542. The protein mixture was then separated by size exclusion chromatography and elution fractions were analyzed by SDS-PAGE. UbcH7~Ub, but not UbcH7 alone was able to form a stable complex with SidC542 as indicated by the co-fractionation of SidC542 with UbcH7~Ub, but not with UbcH7. In contrast, although the SidC542 ΔINS mutant well behaved as a mono-dispersed protein (Horenkamp et al., 2014), it failed to interact with Ub-UbcH7 (Figure 3.3E). These data suggest that the INS domain mediates the binding with E2s. Furthermore, like the E2~Ub-HECT E3 ternary complex (Kamadurai et al., 2009), the covalently linked ubiquitin likely interacts with an area near the E3 catalytic site, thus stabilizing the E2~Ub-SidC ternary complex. Consistent with these results, the SidC542 ΔINS mutant, which failed to bind E2, is completely inactive in ubiquitination assays (Figure 3.5).

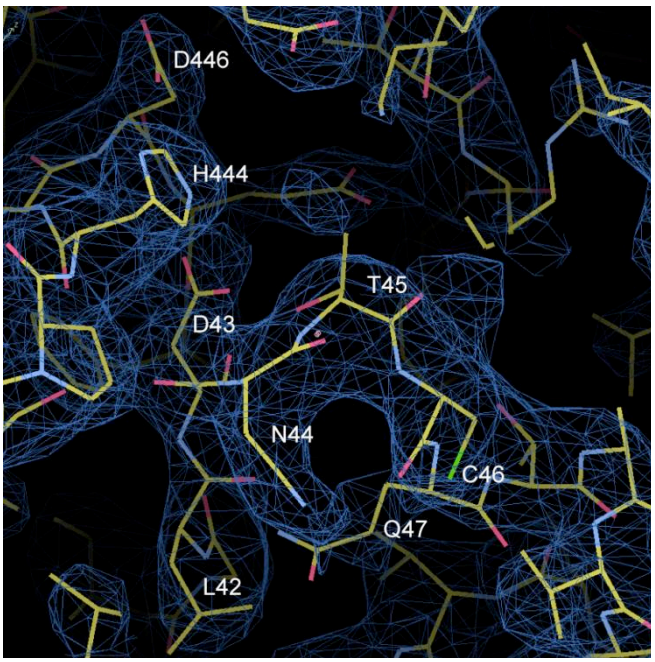


Figure 3.1. Representative electron density map.

Electron density of the area of the ubiquitin ligase catalytic site contoured at 1σ after the final cycle of refinement.

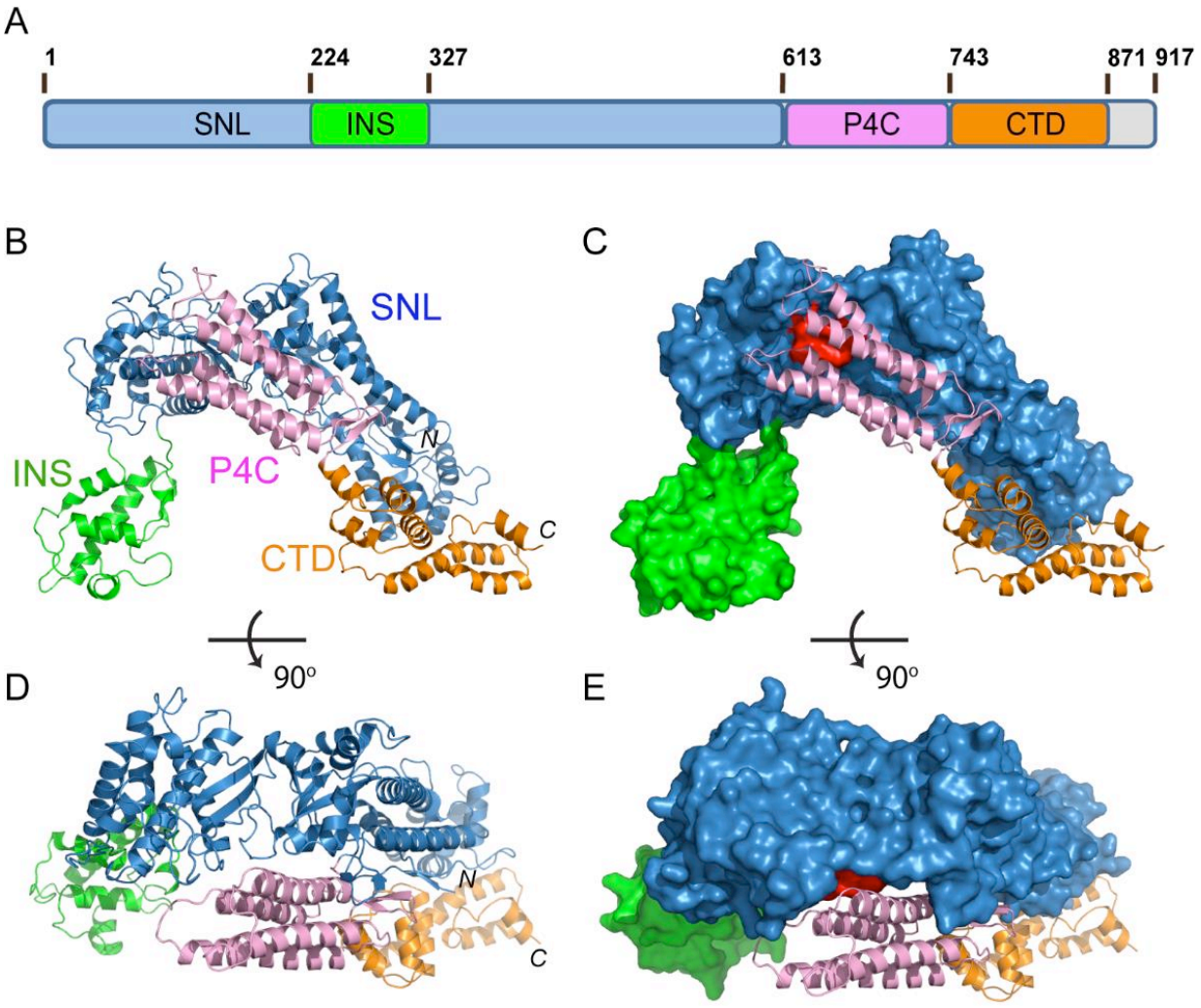


Figure 3.2. Crystal structure of SidC871 (aa. 1-871).

(A) Schematic diagram of the domain structure of SidC. SidC contains an N-terminal SNL domain, an insertion domain (INS), a C-terminal PI(4)P binding domain (P4C), and a C-terminal domain with unknown function (named CTD, colored in brown). (B) Ribbon diagram of the overall structure of SidC. (C) Overall structure of SidC with the SNL and INS domains shown in surface and the P4C and CTD in ribbon. The ubiquitin ligase active site in the SNL domain is colored in red. (C) and (B) have the same orientation. (D) A 90^0 rotated view of (B). (E) A 90^0 rotated view of (C). Color scheme from (A) to (E): the SNL domain in blue; the INS domain in green; the P4C domain in pink; the CTD domain in brown.

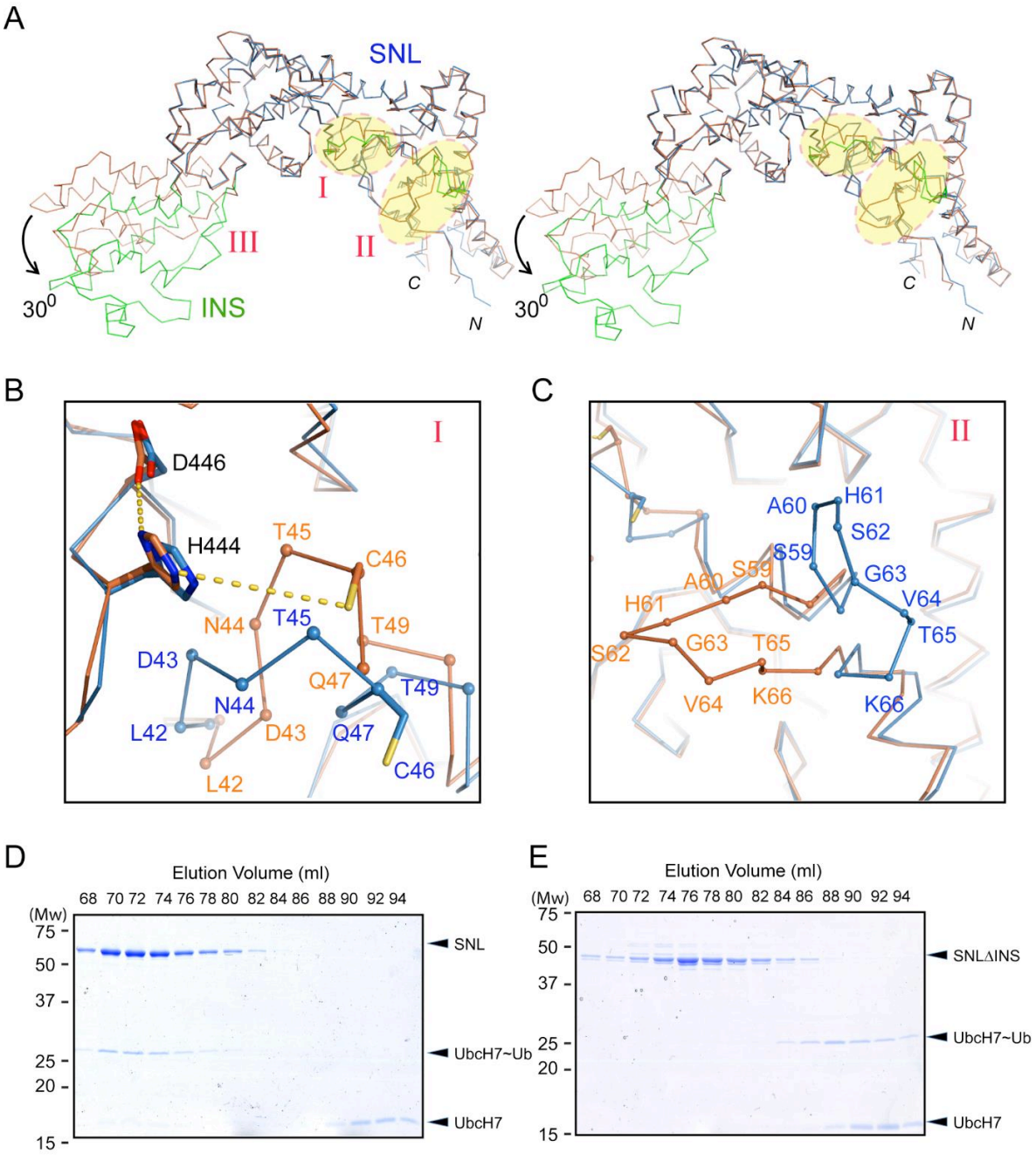

A
SNL
I
III
II
INS
30^0
C
N
B
I
D446
H444
T45
C46
N44
T45
T49
D43
Q47
T49
N44
L42
D43
Q47
L42
C46
C
II
A60
H61
S62
S59
G63
V64
T65
H61
A60
S59
S62
G63
T65
V64
K66
K66
D
Elution Volume (ml)
68 70 72 74 76 78 80 82 84 86 88 90 92 94
(Mw)
75
50
37
25
20
15
SNL
UbcH7~Ub
UbcH7
E
Elution Volume (ml)
(Mw)
68 70 72 74 76 78 80 82 84 86 88 90 92 94
75
50
37
25
20
15
SNLΔINS
UbcH7~Ub
UbcH7

Figure 3.3. Conformational dynamics of the SNL and INS domains.

(A) Stereo view of the Cα trace of the SNL (blue) and INS (green) domains of SidC in overlay with our previously reported SNL-INS domain structure (light brown; PDB ID: 4TRH). The three areas that have major conformational changes are labeled with I, II, and III, respectively. The INS domain in SidC871 is bent by about 30^0 relative to the SNL domain. (B) Zoom in view of the conformational changes at the catalytic site. C46 is shifted away from H444 and D446 in SidC871 (blue) compared with the SidC542 structure (brown). (C) Zoom in view of the conformational change at the non-conserved loop (residue 59-66). (D) The SNL domain (SidC542) forms stable complex with UbcH7~Ub. SDS-gel of the size exclusion chromatography fractions from the sample containing SNL domain with ubiquitin-charged UbcH7. (E) The INS domain is involved in the binding of the SNL domain with UbcH7~Ub. SDS-gel of the size exclusion chromatography fractions from the sample containing the SNLΔINS domain with ubiquitin-charged UbcH7. UbcH7~Ub did not co-migrate with the SNLΔINS domain.

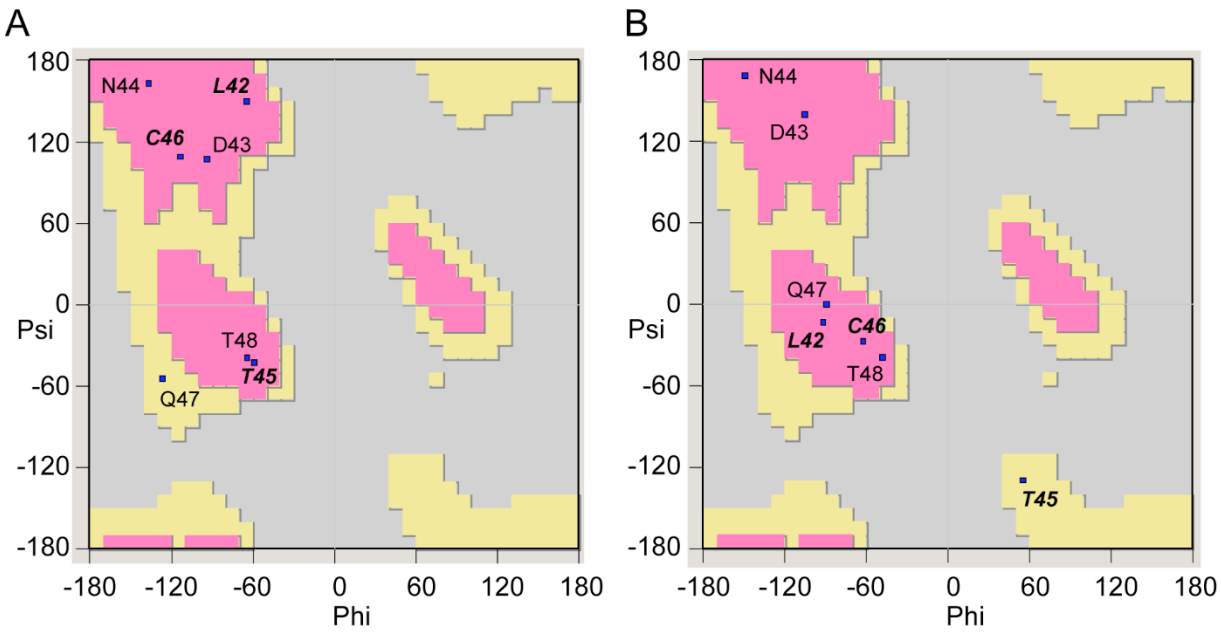


Figure 3.4. Ramachandran plot of residues at the ubiquitin ligase catalytic site.

(A) Ramachandran angles of the catalytic site residues (42-47) from SidC871 structure. (B) Ramachandran angles of the catalytic site residues (42-47) from SidC542 structure. Note the peptide flipping in residues L42, T45, and C46.

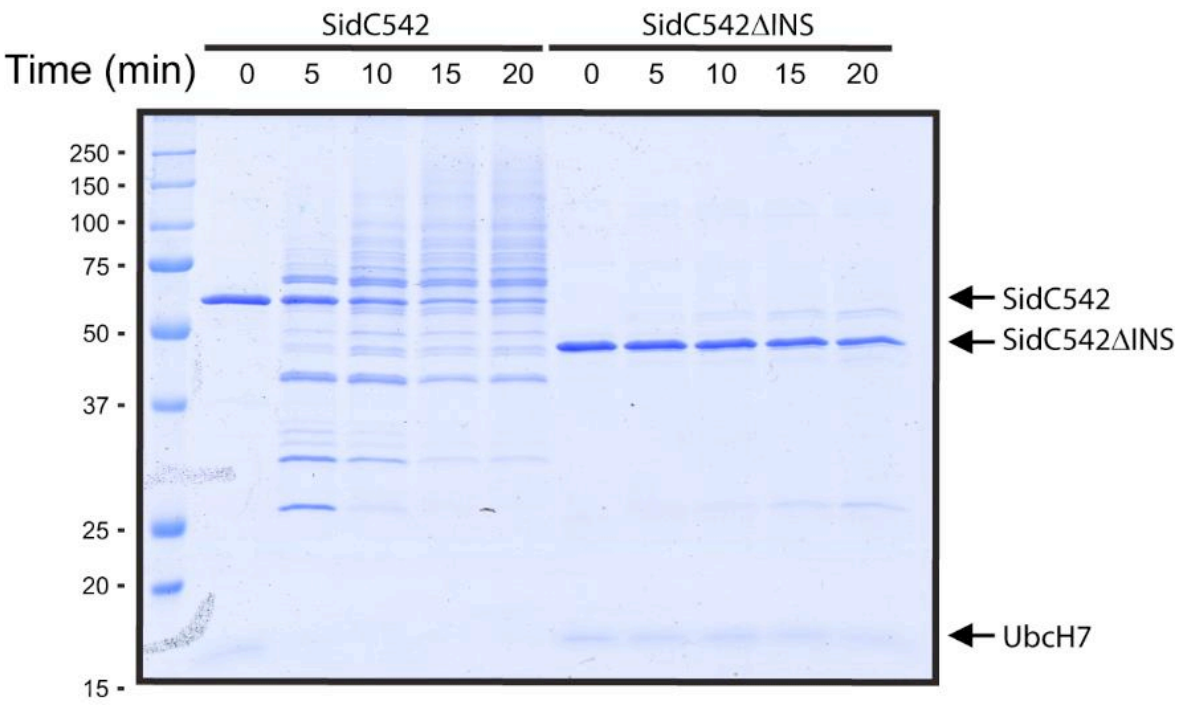


Figure 3.5. In vitro ubiquitination assays with SidC542 and SidC542ΔINS.

Polyubiquitinated species accumulated in reactions with SidC542, however, polyubiquitinated protein bands are absent in reactions with Sidc542ΔINS.

Table 3.1. Data collection, phasing and structural refinement statistics

A. Data collection statistics	
Space group	C2
Cell dimensions	a = 228.16 Å, b = 83.934 Å, c = 129.4 Å, $\alpha = 90^0$, $\beta = 108.82^0$, $\gamma = 90^0$
	Native
Synchrotron beam lines	MCCHESS A1
Wavelength (Å)	0.9759
Maximum resolution (Å)	2.86
Observed reflections	266,725
Unique reflections	53,345
Completeness (%)[a]	100.0(99.9)
$\langle I \rangle / \langle \sigma \rangle$[a]	46.5(2.7)
R_{sym}[a,b] (%)	9.0(79.0)
B. Refinement statistics	Native
Resolution (Å)[a]	50-2.86(3.01-2.86)
R_{crys} / R_{free} (%)[a,c]	22.0/28.3(38.5/44.5)
Rms bond length (Å)	0.01
Rms bond angles (°)	1.24
Ramachandran plot	
Most favored/Additional (%)	95.0/5.0
Generous/Disallowed (%)	0/0

[a]Values in parenthesis are for the highest resolution shell.

[b]$R_{sym} = \Sigma_h \Sigma_i |I_I(h) - \langle I(h) | / \Sigma_h \Sigma_i I_I(h)$.

[c]$R_{crys} = \Sigma(|F_{obs}| - k|F_{cal}|)/\Sigma|F_{obs}|$. R_{free} was calculated for 5% of reflections randomly excluded from the refinement.

A novel highly specific PI(4)P-binding domain

It has been reported that SidC contains a specific PI(4)P binding domain that lies between residues 609-776 and this 20 kDa region was named P4C (PI(4)P binding of SidC) (Ragaz et al., 2008). Our crystal structure of SidC871 revealed the architecture of this unique domain. The P4C domain (aa. 614-743) is mainly comprised of four α-helices running in an antiparallel way. These four α-helices form a bundle with one end sealed by a C-terminal short β hairpin. A pocket is created at the other end by two loops (L1 and L2) connecting α1, α2 and α3, α4, respectively (Figure 3.6A). Strikingly, the calculated electrostatic surface potentials revealed that this pocket is highly positively charged (Figure 3.6B) and it is the only area clustered with positive charges in the domain (Figure 3.6C). Our structural observations suggest that the positively charged pocket is likely the binding site for phosphoinositides (PI).

To test the specificity of PI binding, we performed liposome floatation assays (Matsuoka et al., 2000). Liposomes were prepared with 80% phosphatidylcholine (PC), 10% Phosphatidylserine (PS) and 10% either PI(4)P or other phospholipids. After incubation with purified GFP-tagged P4C proteins for 30 min, a sucrose gradient centrifugation was performed to separate liposomes from unbound proteins. The floated liposomes were pooled and normalized, and bound proteins were analyzed by Western blot followed by densitometry quantification. A large amount of GFP-P4C proteins were recovered in the liposome samples containing PI(4)P. However, almost no GFP-P4C co-floated with liposomes in the absence of PI(4)P or with liposomes containing either PI(3)P, PI(4,5)P_2, or phosphatidic acid (PA) (Figure 3.6D and 3.6E). The specific interaction with PI(4)P by the P4C domain can also be directly visualized by fluorescence microscopy. Large unilamellar vesicles (LUV) were prepared with lipid compositions similar to those used in the liposome floatation assays. In addition, a Dil dye

(Invitrogen) was incorporated into the LUV for fluorescent detection. After a 30 min incubation of GFP-P4C with the liposomes, the samples were imaged under a fluorescence microscope. In agreement with the results from the liposome floatation assay, strong GFP signals were observed on the surface of the liposomes containing PI(4)P, whereas there was almost no detectable GFP signal for the liposomes without PI(4)P (Figure 3.6F and 3.6G). Together, these results demonstrate that the P4C domain of SidC binds to PI(4)P with a high selectivity.

Determinants of PI(4)P recognition and membrane targeting by the P4C domain

Membrane association by PI binding proteins is generally governed by the stereospecific recognition of a distinct PI headgroup and the nonspecific insertion of hydrophobic residues into the membrane bilayer (Cho et al., 2005; Moravcevic et al., 2012). The structure of the P4C domain highlights these key features for its association with PI(4)P-containing membranes. First, there are two conserved arginine residues (R638 and R652) located in the positively charged pocket of the P4C domain (Figure 3.7A and Figure 3.8). These two arginine residues contribute to the overall positive charge of the pocket and may be directly involved in PI(4)P binding by making hydrogen bonds and salt bridges with the headgroup of PI(4)P. To test this hypothesis, we constructed mutants in which each of the two arginine residues was replaced with a glutamine. The R638Q mutation significantly reduced the binding to PI(4)P-containing liposomes to ~30% of that of the wild type P4C, while the R652Q mutation nearly completely abolished the PI(4)P binding by the P4C domain (Figure 3.7B and 3.7C). The second feature is the hydrophobic nature of the two loops L1 and L2 that surround the PI(4)P binding pocket (Figure 3.7A). L1 contains three consecutive hydrophobic residues, W642, W643, and F644 while L2 contains two, W704 and F705. Although these residues are variable, their hydrophobic

nature is conserved (Figure 3.8). These hydrophobic residues likely function as a “membrane insertion motif” (MIM). Mutation of these hydrophobic residues to the polar residue serine either in L1 (W642S/W643S/F644S) or in L2 (W704S/F705S) significantly reduced the binding to PI(4)P-containing liposomes, whereas the binding was completely abolished when all five hydrophobic residues were mutated (Figure 3.7B and 3.7C). In agreement with the floatation assays, the effects of P4C mutations on liposome binding were further confirmed by direct fluorescence imaging (Figure 3.7D and 3.7E). These data demonstrate that the high affinity PI(4)P binding by the P4C domain of SidC requires both cationic residues, which mediate the recognition of the headgroup of PI(4)P and the MIM, which enhances membrane association by direct insertion of hydrophobic side chains into the membrane bilayer.

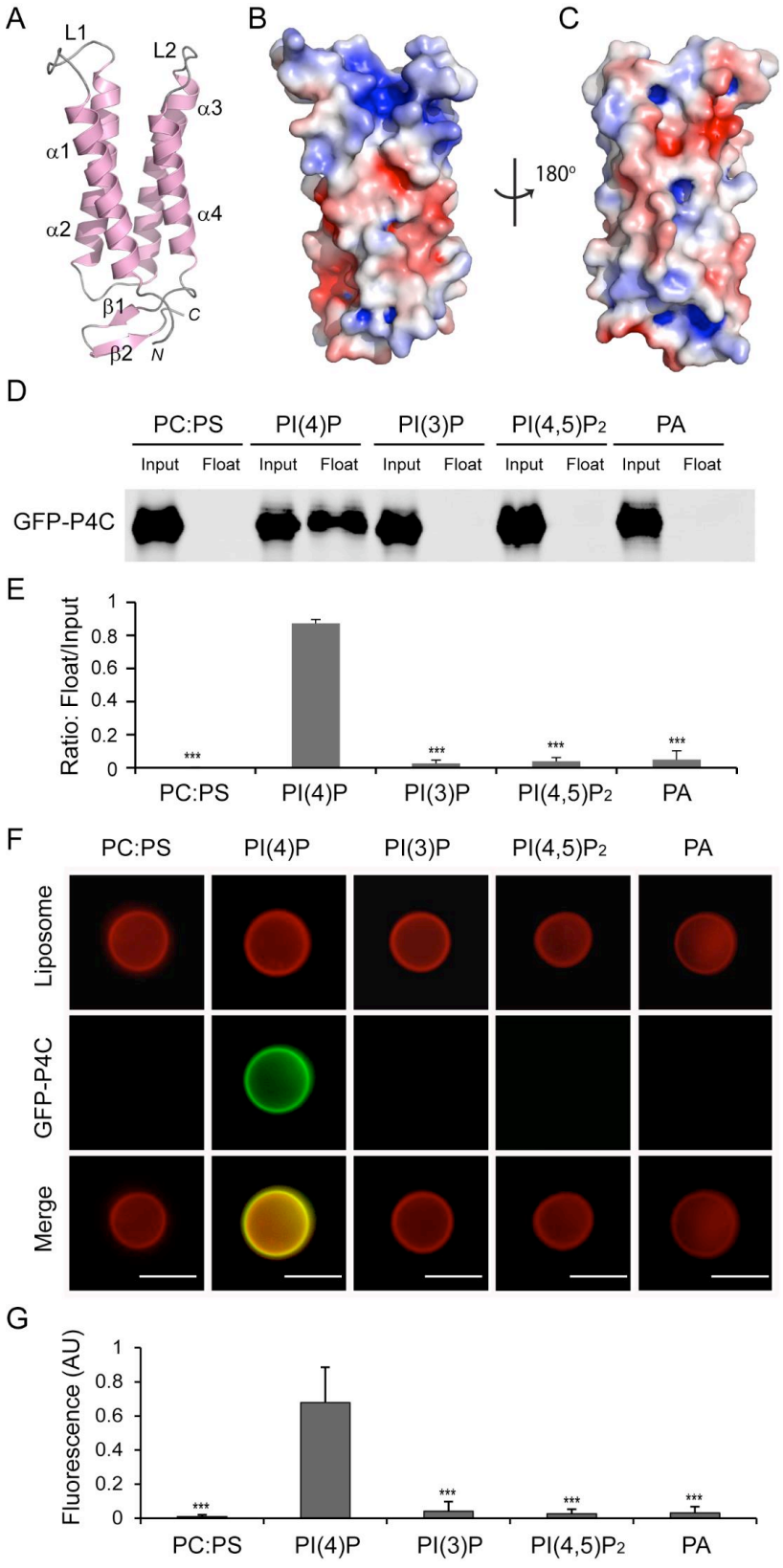

A
L1
L2
α3
α1
α4
α2
β1
C
β2
N
B
C
180°
D
PC:PS
PI(4)P
PI(3)P
PI(4,5)P2
PA
Input
Float
GFP-P4C
E
Ratio: Float/Input
1
0.8
0.6
0.4
0.2
0
***
PC:PS
PI(4)P
PI(3)P
PI(4,5)P2
PA
F
PC:PS
PI(4)P
PI(3)P
PI(4,5)P2
PA
Liposome
GFP-P4C
Merge
G
Fluorescence (AU)
1
0.8
0.6
0.4
0.2
0
***
PC:PS
PI(4)P
PI(3)P
PI(4,5)P2
PA

Figure 3.6. The P4C domain of SidC specifically binds with PI(4)P.

(A) Ribbon diagram of the P4C domain. (B) Molecular surface of the P4C domain. The surface is colored based on electrostatic potential with positively charged region in blue (+5 kcal/electron) and negatively charged surface in red (-5 kcal/electron). (C) Back view of the surface model of the P4C domain. (D) Liposome floatation assays. Input and float samples were analyzed by SDS-PAGE and immunoblotted with anti-GFP antibodies. Recombinant GFP-P4C showed selective binding to PI(4)P-positive liposomes but not the liposomes with other components. (E) Quantification of liposome floatation assays from three independent experiments. Error bars represent standard deviation. (F) Fluorescent images of liposome binding by GFP-P4C. Only the liposomes containing PI(4)P showed strong binding of GFP-P4C. Scale bar = 10 μm (G) Quantification of liposome binding of GFP-P4C. GFP fluorescent signals were normalized to red Dil dye signals on the same liposome and averaged on three randomly picked liposomes. Error bars represent standard deviation. ** $P < 0.01$; *** $P < 0.001$.

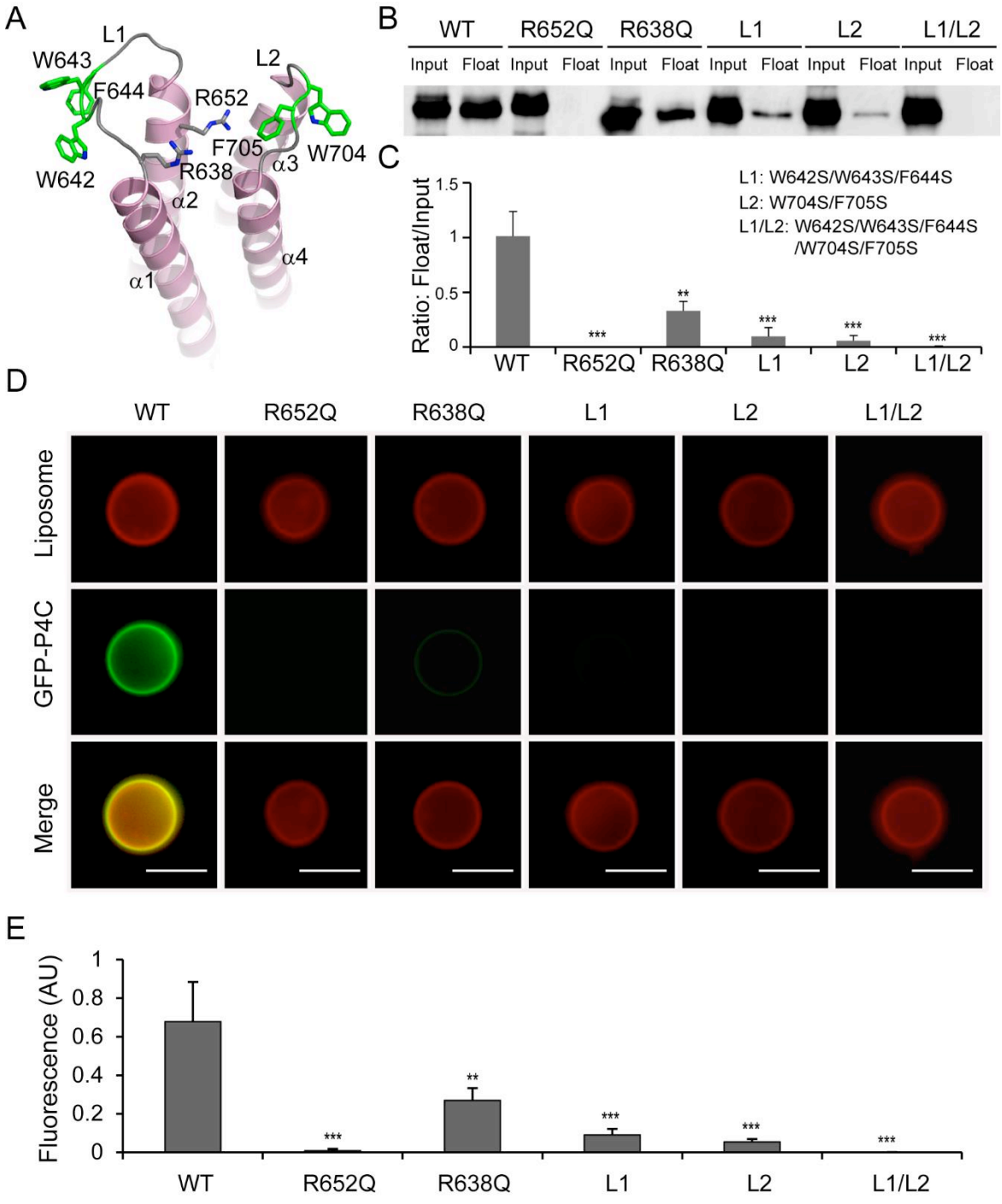

A
L1
W643
F644
R652
L2
F705
W704
α3
R638
W642
α2
α4
α1
B
WT
R652Q
R638Q
L1
L2
L1/L2
Input
Float
C
L1: W642S/W643S/F644S
L2: W704S/F705S
L1/L2: W642S/W643S/F644S
/W704S/F705S
Ratio: Float/Input
1.5
1
0.5
0
**
***
WT
R652Q
R638Q
L1
L2
L1/L2
D
WT
R652Q
R638Q
L1
L2
L1/L2
Liposome
GFP-P4C
Merge
E
Fluorescence (AU)
1
0.8
0.6
0.4
0.2
0
**
***
WT
R652Q
R638Q
L1
L2
L1/L2

Figure 3.7. Determinants of PI(4)P recognition and membrane targeting by the P4C domain

(A) Ribbon diagram of the P4C domain. Residues that play a role in membrane interaction are highlighted in sticks. R652 and R638 form a pocket for the binding of the PI(4)P headgroup. The L1 (W642, W643, and F644) and L2 (W704 and F705) loops are colored in green, and form the membrane interacting motif (MIM). (B) Liposome floatation assay for P4C mutants. The R652Q and the L1/L2 (W642S/W643S/F644S/W704S/F705S) mutants completely abolished PI(4)P-liposome binding. (C) Quantification of liposome floatation assays of P4C mutants averaged from three independent assays. (D) Fluorescent images of liposome binding by GFP-P4C mutants. Mutations of cationic residues in the PI(4)P binding pocket and hydrophobic residues at the two membrane insertion loops significantly reduce the binding to PI(4)P-containing liposome. Scale bar = 10 μm. (E) Quantification of liposome binding of GFP-P4C mutants. GFP fluorescent signals were normalized to red Dil dye signals on the same liposome and averaged on three randomly picked liposomes. Error bars represent standard deviation. ** $P < 0.01$; *** $P < 0.001$.

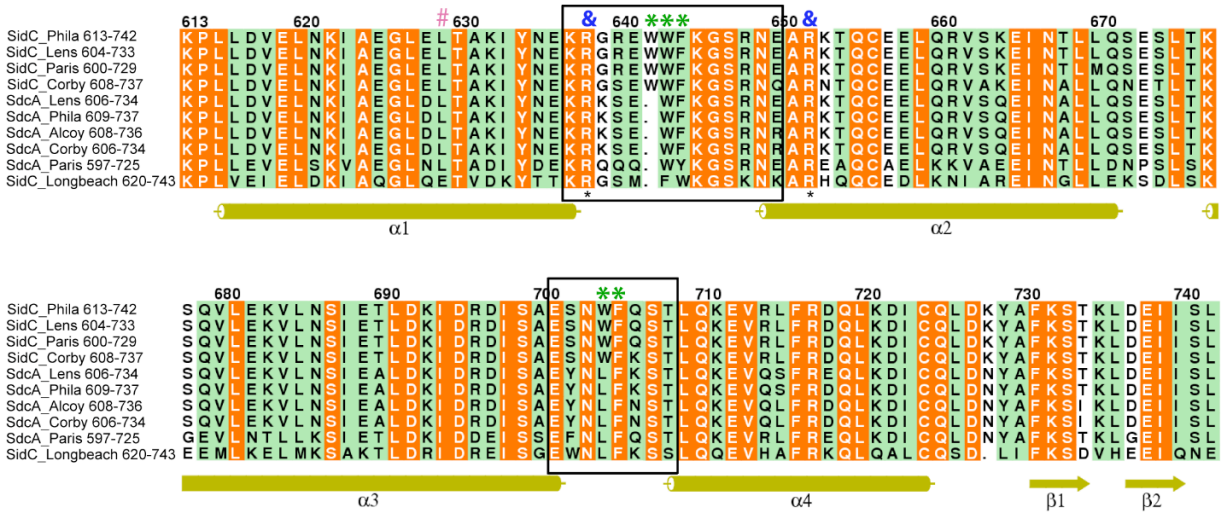


Figure 3.8. Multiple sequence alignment of the P4C domains of SidC.

The sequences corresponding to the P4C domain of SidC (aa. 613-742) from different Legionella species were aligned by Clustal Omega [1] and colored by ALSCRIPT [2]. Secondary elements are drawn below the alignment. The L1 and L2 loop are marked with squares. Two arginine residues forming the cationic binding pocket are marked by "&". Hydrophobic residues (W641, W642, F643, W704, and F705) that form the MIM motif are highlighted with "*". The interface residue L629 is indicated by "#". Entrez database accession numbers are as follows: SidC_Phili, gi: 52842719; SidC_Lens, gi: 54295348; SidC_Paris, gi: 54298515; SidC_Corby, gi: 148360028; SdcA__Lens, gi: 54295347; SdcA__Phili, gi: 52842718; SdcA__Alcoy, gi: 296108150; SdcA__Corby, gi: 148360029; SdcA__Paris, gi: 54298514; SidC__Longbeach, gi: 289166408.

Intracellular localization of the P4C domain in living cells

We further examined the specific PI(4)P binding by the P4C domain in living cells. Wild type and mutant P4C domains tagged with EGFP were expressed in N2A cells by transient transfection (Figure 3.9A). Wild type GFP-P4C displayed enrichment at both the plasma membrane (PM) and the perinuclear region, presumably the Golgi complex, which are the two major PI(4)P-rich compartments in the cell (Balla et al., 2005). Consistent with the in vitro liposomal binding results, the R652Q mutant showed a cytosolic and nuclear localization, suggesting PI(4)P binding is severely hampered by this mutation. The R638Q, which showed a reduced affinity for PI(4)P in vitro, displayed a similar localization compared to wild type but with a higher cytosolic distribution. Individual MIM mutations, L1 and L2 also displayed high level of cytosolic localization, indicating a reduced affinity for membrane. Strikingly, the L1 and L2 double mutation completely abolished the membrane localization of P4C, rendering the mutant to a complete cytosolic protein (Figure 3.9A and 3.9B). Similar intracellular localization of wild type and mutant GFP-P4C domains was also observed in other cell lines such as Cos7 (Figure 3.10). These in vivo data further validate the key determinants for the binding of the P4C domain to PI(4)P-enriched membranes suggesting a potential application of this domain as a general PI(4)P probe in living cells.

To explore the possibility of developing a new PI(4)P probe, we compared the PI(4)P binding and intracellular localization of the P4C domain with other reported PI(4)P probes. GFP-tagged PH domains of FAPP1 and OSBP (Balla et al., 2005; Dowler et al., 2000) and the GFP-tagged P4M domain from the *Legionella* effector protein SidM/DrrA (Schoebel et al., 2010; Del Campo et al., 2014; Hammond et al., 2014) were expressed and purified from *E.coli*. The binding of PI(4)P was analyzed with liposome floatation assays (Figure 3.11). In the floatation

experiments, a higher percentage of GFP-P4C co-floated with PI(4)P-containing liposomes than any of the other PI(4)P probes tested (Figure 3.11A and 3.11B). This observation was further supported by direct visualization of the binding with fluorescence microscopy (Figure 3.11C and 3.11D).

We next compared the intracellular localization of these PI(4)P probes in living cells. Constructs expressing GFP-tagged probes were co-transfected with either the PM marker, mCherry-PLCδ-PH (Yagisawa et al., 1994) or the Golgi marker, DsRed-GalT (Watzele et al., 1991) in N2A cells. The GFP-P4C probe displayed a significant co-localization with both the PM marker and the Golgi marker (Figure 3.9C-E), indicating that the P4C domain is able to faithfully report the two major pools of PI(4)P in living cells. Although all the other three probes also co-localized with GalT at the Golgi apparatus, their PM signals were weaker compared to those of P4C. In summary, our data suggest that the P4C probe from SidC has the potential to be developed as a more sensitive and unbiased biosensor for detecting PI(4)P in living cells.

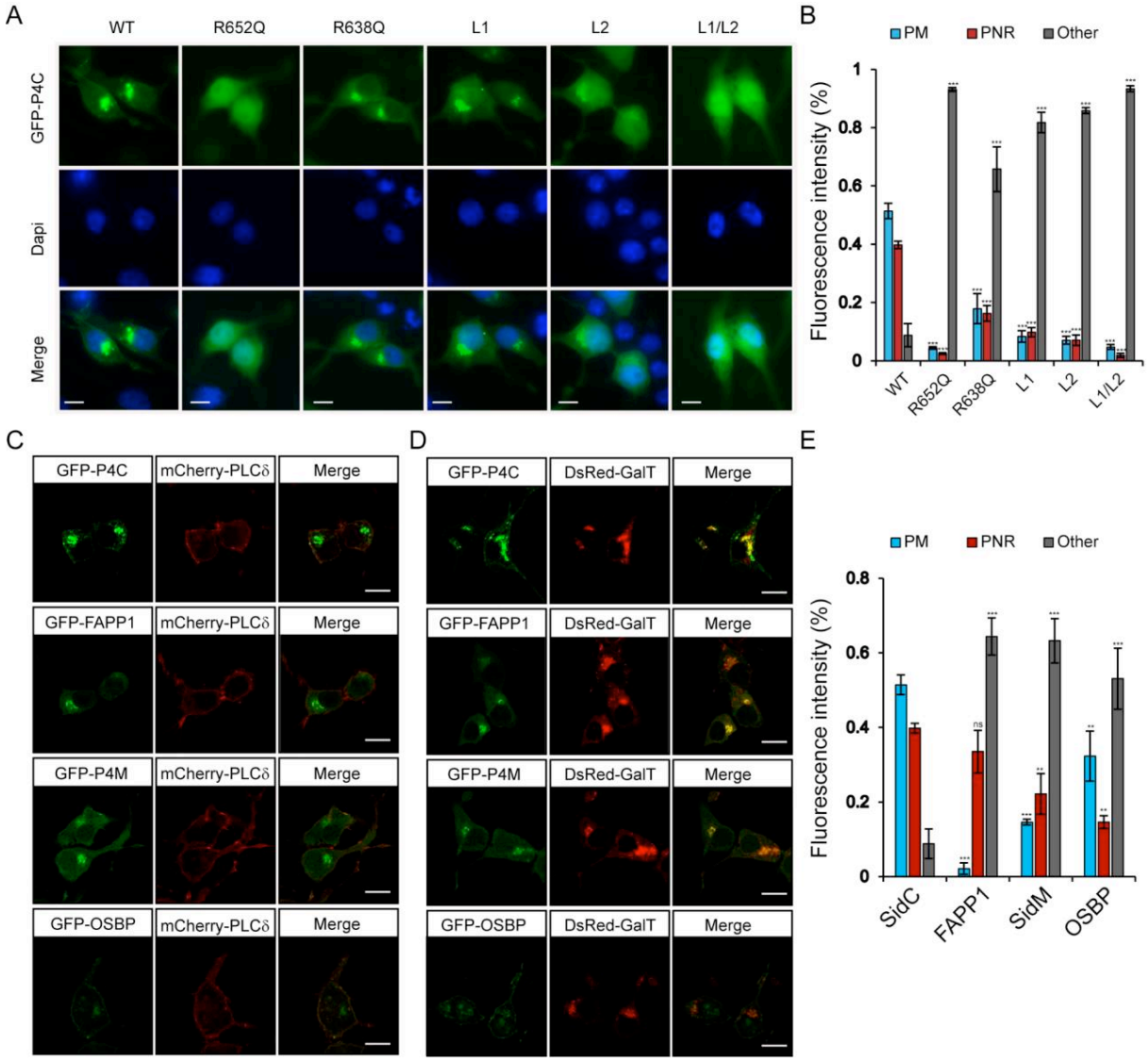

A
WT
R652Q
R638Q
L1
L2
L1/L2
GFP-P4C
Dapi
Merge
B
PM
PNR
Other
Fluorescence intensity (%)
WT
R652Q
R638Q
L1
L2
L1/L2
C
GFP-P4C
mCherry-PLCδ
Merge
GFP-FAPP1
GFP-P4M
GFP-OSBP
D
DsRed-GalT
E
SidC
FAPP1
SidM
OSBP

Figure 3.9. Intracellular localization of fluorescent protein fusions of the P4C domain from SidC (A) GFP-tagged wild type P4C domain localized to the perinuclear region and plasma membrane, while this localization is altered in PI(4)P-binding defective P4C mutants in N2A cells. The nucleus was stained with DAPI. Wild type P4C showed both plasma membrane and perinuclear localization. The R638Q, L1, and L2 mutants had a more diffuse localization while the R652Q and the L1/L2 mutants were completely cytosolic. (B) Quantification of the intracellular localization of GFP-P4C represented by the percentage of the fluorescence intensities at the plasma membrane (PM), perinuclear region (PNR), and other areas of the cell. Error bars represent standard deviation. The measurements were averaged by three randomly selected cells. (C) Confocal images of localizations of the plasma membrane marker mCherry-PLCδ-PH with GFP-tagged PI(4)P probes in N2A cells. (D) Confocal images of colocalizations of the Golgi marker DsRed-GalT with GFP-tagged PI(4)P probes in N2A cells. Scale bar = 10 μm in all images. (E) Quantification of the intracellular localization of PI(4)P probes. Error bars represent standard deviation. ** $P < 0.01$; *** $P < 0.001$.

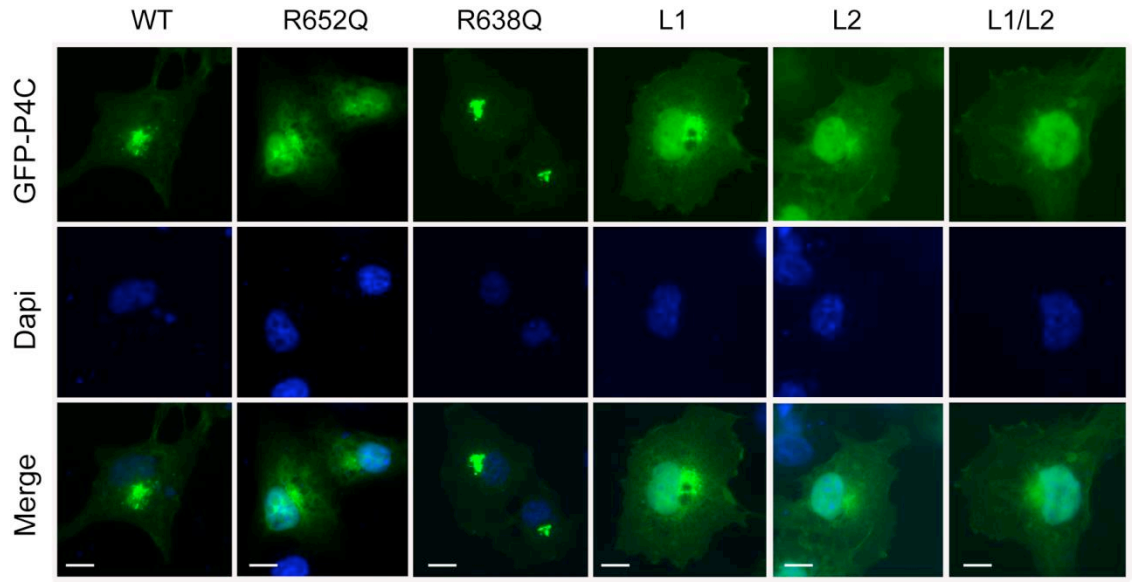


Figure 3.10. Intracellular localization of GFP-P4C and its PI(4)P-binding defective mutants in Cos7 cells. Epi-fluorescent images of Cos7 cells transfected with GFP-tagged P4C or PI(4)P-binding defective mutants. The nucleus was stained with DAPI. P4C showed both plasma membrane and perinuclear localization. The R638Q, L1, and L2 mutants had a more diffuse localization, while the R652Q and the L1/L2 mutants were completely cytosolic.

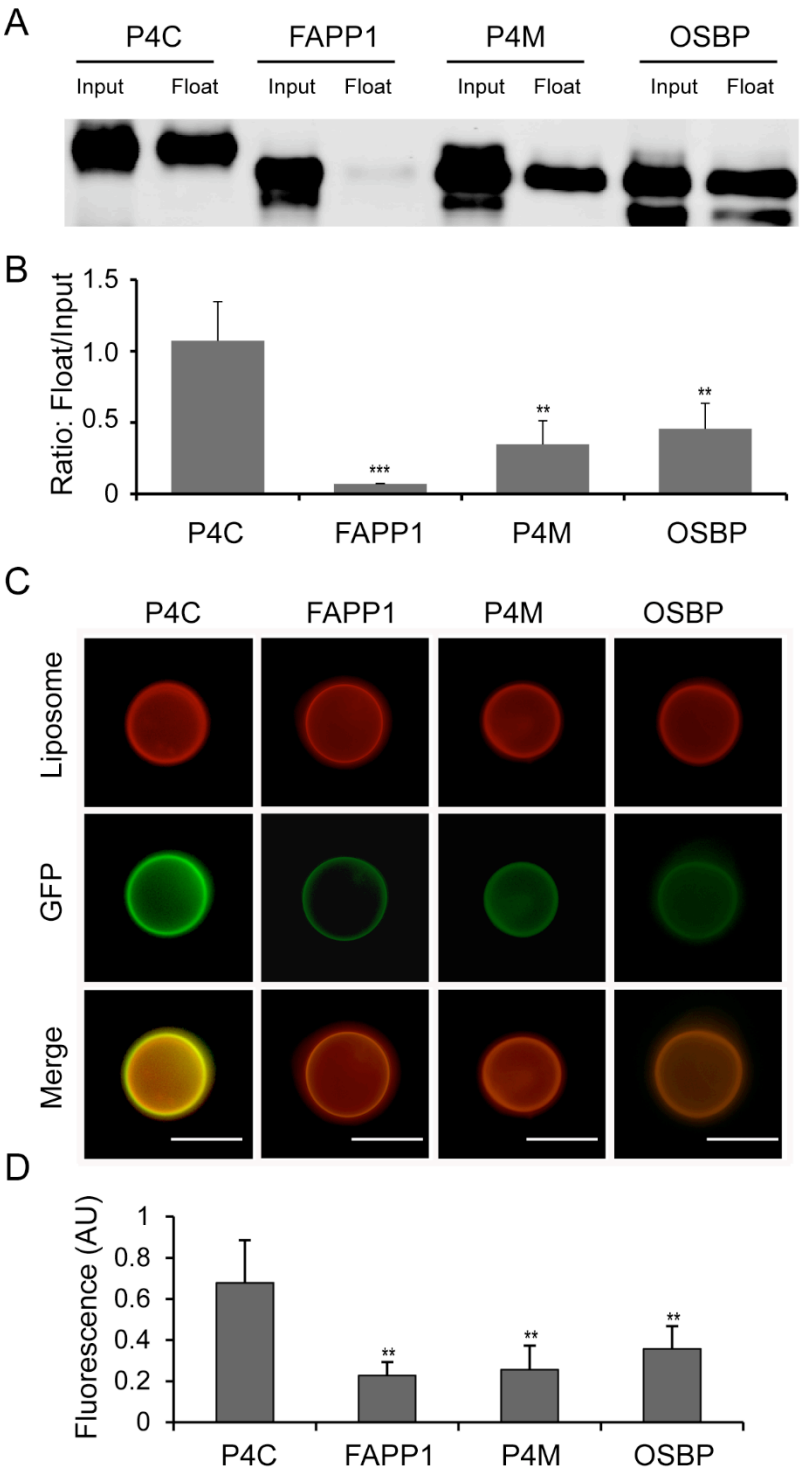

A
P4C
FAPP1
P4M
OSBP
Input
Float
Input
Float
Input
Float
Input
Float
B
Ratio: Float/Input
1.5
1.0
0.5
0
***
**
**
P4C
FAPP1
P4M
OSBP
C
P4C
FAPP1
P4M
OSBP
Liposome
GFP
Merge
D
Fluorescence (AU)
1
0.8
0.6
0.4
0.2
0
**
**
**
P4C
FAPP1
P4M
OSBP

Figure 3.11 Comparison of PI(4)P-containing liposome binding by PI(4)P probes.

(A) Western blot of samples from liposome floatation assays. Recombinant proteins of GFP-P4C, GFP-FAPP1-PH, GFP-P4M, and GFP-OSBP-PH were incubated with PI(4)P-containing liposomes. Proteins that floated with liposomes were analyzed by Western blot using poly-clonal antibodies against GFP. (B) Quantification of liposome floatation assays from three independent assays. Error bars represent standard deviation. (C) Fluorescent images of liposome binding by GFP-tagged probes. The same amount of GFP-fusion proteins were incubated with PI(4)P-containing liposomes. After a 20 min incubation at room temperature, mixtures were applied to glass cover-slips for imaging. The GFP-P4C demonstrated strongest binding affinity with PI(4)P-containing liposomes. Scale bar = 10 μm. (D) Quantification of liposome binding of GFP-P4C. GFP fluorescent signals were normalized to red Dil dye signals on the same liposome and averaged on three randomly picked liposomes. Error bars represent standard deviation. ** $P < 0.01$; *** $P < 0.001$.

The interface between the P4C and SNL domain

Our crystal structure revealed that the P4C packs against the SNL domain and blocks the accessibility of the E3 catalytic site (Figure 3.2). The interaction between the P4C and SNL domains is mainly mediated by extensive hydrophobic interactions between residues L629 and I633 of the P4C domain and a hydrophobic patch consisting of residues V50, I53, A60, I140, and P141 near the ligase catalytic site on the SNL domain (Figure 3.12A and B). To characterize the nature of this intramolecular interaction, we created the L629R mutant with the speculation that the bulky and charged arginine residue would destroy both surface complementarity and the hydrophobic nature at the interface. Thus, this mutation would displace the P4C domain and keep SidC in an open conformation. Indeed, SidC743 C46A/L629R, but not SidC743 C46A, formed a stable complex with UbcH7~Ub as shown by size exclusion chromatography analysis (Figure 3.12C and 3.12D). The formation of the ternary complex is likely due to the open conformation caused by this mutation, which allows for the binding of the ubiquitin moiety to the catalytic site. These data suggest that the intramolecular interaction between the P4C and the SNL domain is not due to crystal packing. Instead, it likely represents an intramolecular regulatory mechanism in which the displacement of the P4C domain from the catalytic site switches the enzyme into an open conformation and releases the occlusion of the ligase active site to activate the ubiquitin ligase activity of SidC.

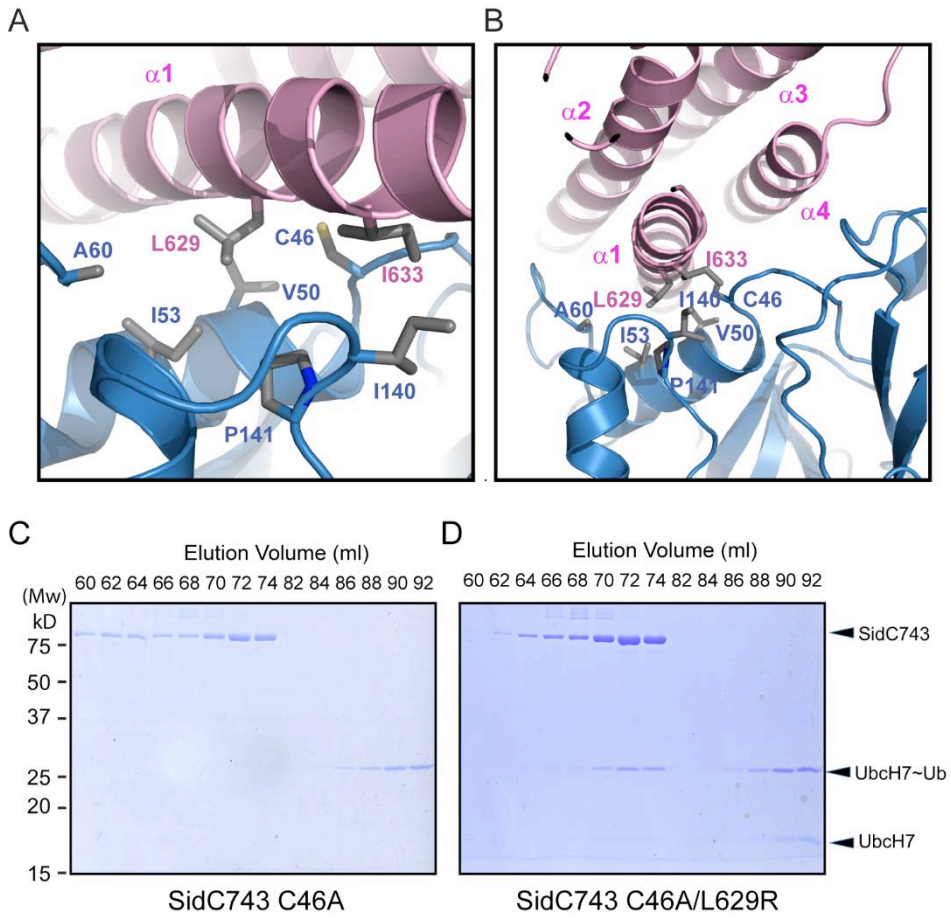


Figure 3.12. The interface between the P4C domain and the SNL domain.

(A) and (B) Two orthogonal views of the interface between the P4C and SNL domains. Hydrophobic residues at the interface are shown in sticks. The P4C domain is colored in pink and the SNL domain in blue. (C) SidC743 C46A does not form a stable complex with ubiquitin-charged UbcH7 as indicated by SDS-PAGE gel analysis of fractions from size exclusion chromatography experiment. (D) SidC743 C46A/L629R forms stable complex with UbcH7~Ub as demonstrated by the co-migration of SidC743 C46A/L629R with UbcH7~Ub on size exclusion column.

Binding of PI(4)P by the P4C domain stimulates the ubiquitin E3 ligase activity of SidC

The intramolecular interaction between the P4C domain and the SNL domain led us to hypothesize that the P4C domain plays a role in regulating the E3 ligase activity. To test this model, we performed ubiquitin ligase activity assays and quantified the reaction by measuring the rate of ubiquitin consumption. SidC542, which lacks the C-terminal P4C domain, showed a similar ubiquitin ligase activity in the absence of liposomes or in the presence of liposomes containing PC/PS or PC/PS/PI(4)P (Figure 3.13A and B). However, SidC743, which contains the P4C domain, exhibited enhanced ligase activity in the presence of PI(4)P containing liposomes (Figure 3.13C and D). These data suggest that the binding of PI(4)P may induce a conformational switch of the enzyme to an open form that exposes the catalytic site and stimulates the ligase activity. This conclusion is further supported by the observation that SidC743 L629R, which prefers an open conformation as shown in above (Figure 3.12), is more active even in the absence of PI(4)P-containing liposomes compared to SidC743 (Figure 3.13C-F). However, the stimulating effect caused by PI(4)P is abolished in both SidC743 R652Q and SidC743 L1/L2 mutants, which are defective in PI(4)P binding (Figure 3.13G-J). In conclusion, PI(4)P-binding stimulates the ubiquitin ligase activity of SidC presumably through an induced conformational switch from a closed to an open form of the enzyme.

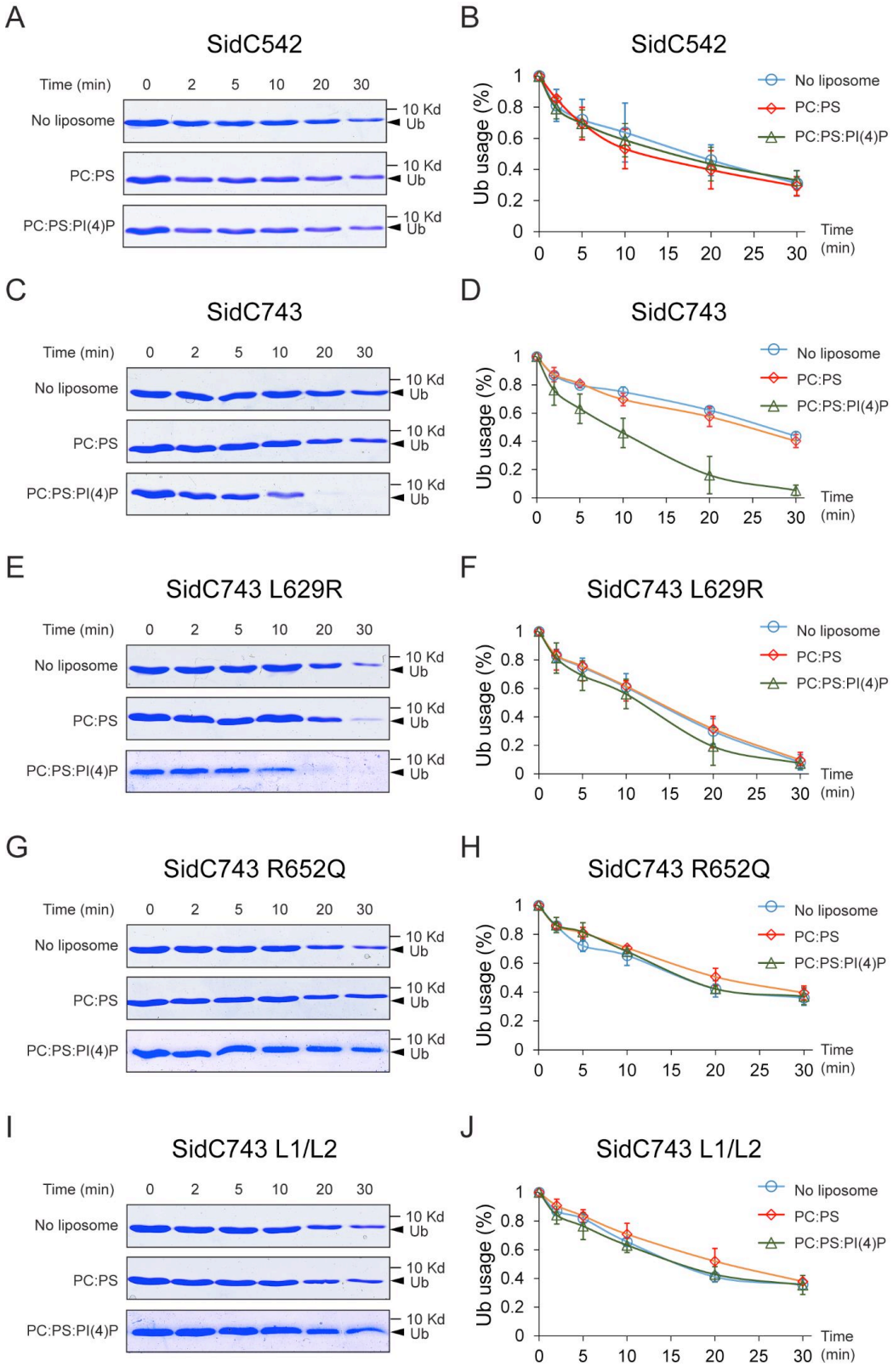

A
SidC542
Time (min) 0 2 5 10 20 30
No liposome
PC:PS
PC:PS:PI(4)P
10 Kd
Ub
B
SidC542
Ub usage (%)
Time (min)
No liposome
PC:PS
PC:PS:PI(4)P
C
SidC743
D
SidC743
E
SidC743 L629R
F
SidC743 L629R
G
SidC743 R652Q
H
SidC743 R652Q
I
SidC743 L1/L2
J
SidC743 L1/L2

Figure 3.13. PI(4)P stimulates the ubiquitin E3 ligase activity of SidC

(A) In vitro ubiquitin ligase activity assay with SidC542 in absence of liposomes and in the presence of liposomes containing PC/PS or PC/PS/PI(4)P. The reactions were stopped at the indicated time points and the samples were analyzed by SDS-PAGE. The decreased intensity of ubiquitin bands indicates the consumption of free ubiquitin during the ligase reaction. (B) Percentage of free ubiquitin left in the reaction at each indicated time points averaged from three independent experiments. (C) and (D) In vitro ubiquitin ligase activity assay with SidC743. The ubiquitin ligase activity is enhanced in the presence of PI(4)P. (E) and (F) In vitro ubiquitin ligase activity assay with SidC743 L629R. This mutant has higher ubiquitin ligase activity even in the absence of PI(4)P, presumably due to the open conformation caused by this mutation. (G)-(H) and (I)-(J) In vitro ubiquitin ligase activity assay with SidC743 R652Q andSidC743 L1/L2. No stimulation of the ubiquitin ligase activity by PI(4)P was observed with these two PI(4)P-binding defective mutants.

PI(4)P-binding by the P4C domain is required for the localization of SidC to the *Legionella* containing vacuole (LCV)

Our structural and biochemical studies revealed the key determinants of the P4C domain for PI(4)P binding. We next asked whether these key residues are also responsible for the association of SidC with the LCV during bacterial infection. SidC and three SidC mutants (R652Q; L1/L2, which carries mutations of the MIM motif, W642S/W643S/F644S and W704S/W705S; and L1/L2/R652Q, which bears both R652Q and the MIM mutations) were constructed in the SidC expressing plasmid pZL199 (VanRheenen et al., 2006) and were transformed into the *L. pneumophila* strain Lp02 lacking both endogenous *sidC* and *sdcA*(Δ*sidC-sdcA*) (Hsu & Luo et al., 2014; VanRheenen et al., 2006). All mutant proteins were expressed at a level comparable to wild type (Figure 3.14A) and had no distinguishable difference in Dot/Icm-mediated translocation by the bacteria (Figure 3.14B). In agreement with *in vitro* results, the number of SidC positive LCVs was reduced from more than 90% for wild type SidC to about 20% for both the R652Q and L1/L2 mutants (Figures 3.15A and 3.15B). Moreover, the average intensity of the immunofluorescent signals observed from the LCVs positive for PI(4)P binding-defective SidC mutants were about 10 times weaker than those from the wild type (Figure 3.16A and 3.16B). These data suggest that both the headgroup binding mediated by R652 and the membrane insertion mediated by the L1 and L2 loops are indispensable for anchoring of SidC to the bacterial phagosome.

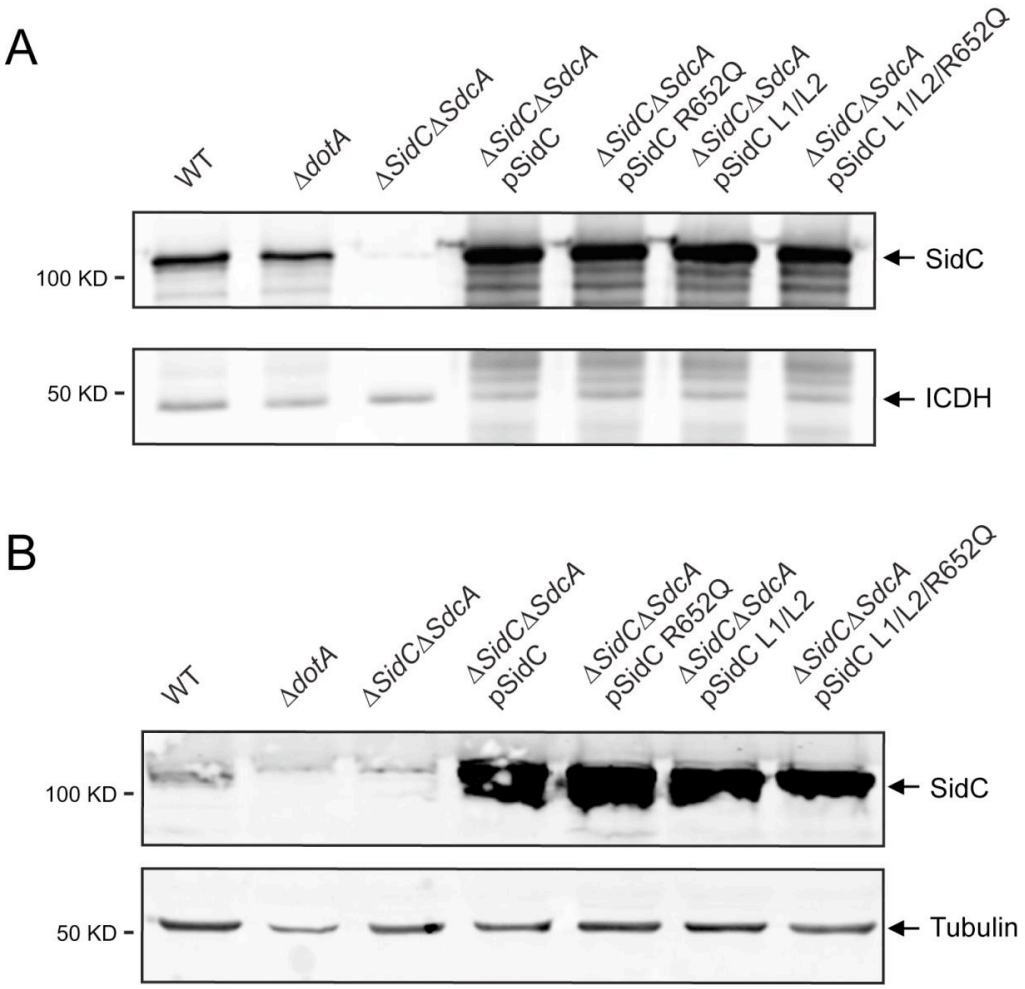


Figure 3.14. SidC expression and translocation after Legionella infection.

(A) Expression of SidC and its mutants in *L. pneumophila*. Bacteria were grown in ACES-buffered medium to OD600=3.5 and collected cells were lysed with SDS-PAGE sample buffer. SidC was detected with a specific antibody and the metabolic enzyme isocitrate dehydrolase (ICDH) was probed as a loading control. *Legionella* strains are the same as used in Figure 3.17. (B) Translocation of SidC mutants by the Dot/Icm transporter. U937 cells were infected with *L. pneumophila* strains at an MOI of 2 for 2hr. Infected cells were lysed with 0.2% saponin and the soluble fractions were probed for SidC after SDS-PAGE. The host protein tubulin was detected as a loading control. *Legionella* strains are the same as used in (A).

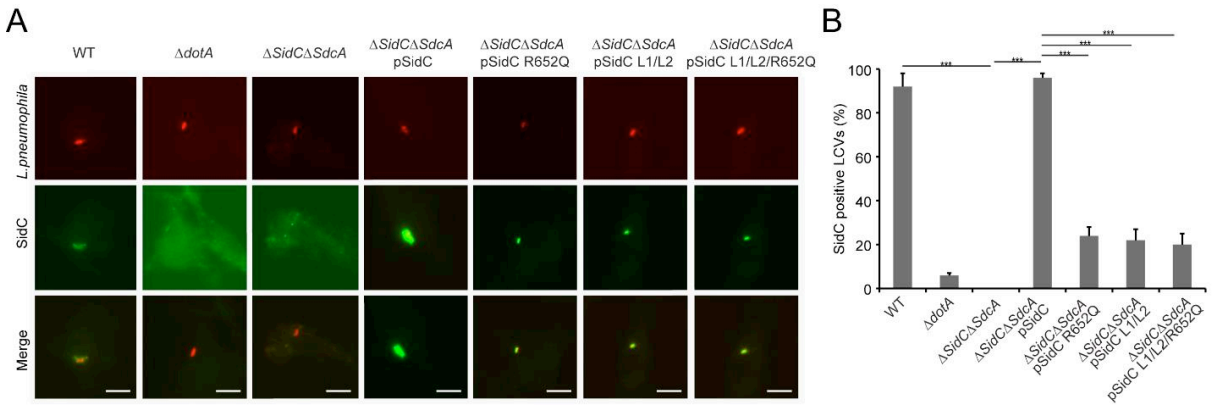


Figure 3.15. The anchoring of SidC to the LCV is dependent on PI(4)P binding by the P4C domain. (A) Immuno-fluorescent staining of SidC on the LCV. Bone marrow-derived macrophages were infected with indicated *L. pneumophila* strains at an MOI of 1 for 2hrs. Samples were first stained for extracellular and intracellular bacteria before being stained for SidC with specific antibody. Scale bars, 2 μm. (B) Percentage of cells containing SidC positive LCVs counted from three independent experiments (at least 150 vacuoles were scored in each experiment).

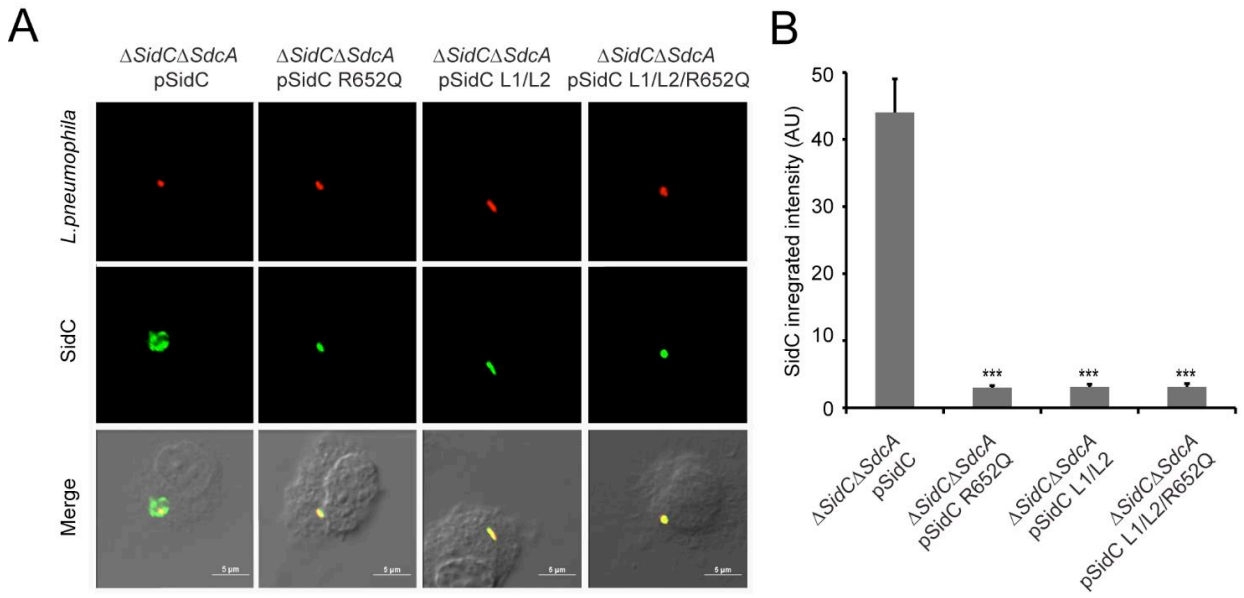


Figure 3.16. The association of the PI(4)P-binging defective SidC mutants with the LCV are significantly reduced.

(A) SidC immuno-fluorescent signal intensity analysis on SidC-positive vacuoles. U937 cells were infected with L. pneumophila strains at an MOI of 2 for 2hr. Samples were stained for bacterium (red) and SidC (green). Images were acquired and SidC signals were processed and analyzed as described in materials and methods. (B) SidC immuno-fluorescent signal intensities from wild type and PI(4)P-binding defective mutants were plotted. In addition to the significant reduction of SidC-positive LCVs (Figure S8), the intensity of SidC immuno-fluorescent signals on SidC-positive vacuoles are also decreased by about 10 fold when infected by Legionella strains expressing PI(4)P-binding defective SidC mutants. Experiments were done in triplicate and at least 150 vacuoles were scored for each treatment.

PI(4)P binding by the P4C domain is essential for the recruitment of ubiquitin and ER proteins to the LCV

It has been shown that shortly after *L. pneumophila* infection, polyubiquitin conjugates are recruited to the bacterial phagosome (Dorer et al., 2006). We recently have shown that the ubiquitin E3 ligase activity of SidC, which is mediated by its N-terminal SNL domain, is required for this recruitment (Hsu & Luo et al., 2014). Here we asked whether the recruitment of ubiquitinated species to the LCV is dependent on the P4C domain-mediated anchoring of SidC to the LCV. U937 macrophages were infected with *Legionella* strains expressing either wild type or SidC mutants. Ubiquitin signals were detected by immunostaining with the FK1 antibody. This recruitment was almost abolished in infections with the Δ*sidC-sdcA* mutant strain, and wild type SidC completely restored the recruitment. However, none of the mutants defective in PI(4)P binding retained this function (Figure. 3.17A and 3.17B). Thus, the association of SidC with the LCV via the P4C domain is necessary for the recruitment of ubiquitinated species.

SidC and its paralog SdcA are also important for efficient recruitment of ER proteins to the bacterial phagosome (Ragaz et al., 2008). Using a *Dictyostelium discoideum* strain stably expressing GFP-HDEL (Singer et al., 2008) (a GFP fusion with the ER retention marker HDEL), we have shown that the ubiquitin E3 ligase activity of SidC is critical for its role in the recruitment of ER components to the bacterial phagosome (Hsu & Luo et al., 2014). To address whether the recruitment of ER components is also dependent on the P4C domain-mediated anchoring of SidC to the LCV, *D. discoideum* cells stably expressing GFP-HDEL were infected with relevant *Legionella* strains. In *D. discoideum* infected with wild type *L. pneumophila*, about 40% of the LCVs were positive for GFP-HDEL 2 hrs after infection (Figure 3.18A and 3.18B). In contrast, nearly no GFP-HDEL signals were detected on LCVs in infections with a Dot/Icm

deficient strain or the *ΔsidC-sdcA* mutant. While this defect was restored by a plasmid expressing SidC, all the PI(4)P binding mutants failed to complement the phenotype (Figure 3.18). Together, these results demonstrate that P4C-mediated anchoring of SidC to the LCV is critical for the recruitment of ER components to bacterial phagosomes.

Discussion

In this study, we report the structure of the nearly full-length SidC (SidC871), which includes the C-terminal P4C domain. Our structure revealed a novel PI(4)P binding domain with a four-alpha-helix-bundle fold. The PI(4)P binding site resides at a cationic pocket at one end of the bundle. The P4C domain is packed against the SNL domain and covers the ubiquitin E3 ligase active site (colored in red in Figure 3.2C). Hence, this conformation may represent an inactive closed form of the enzyme. We also show that the L629R mutant switches the enzyme from the closed form to an open conformation and activates the enzyme. Interestingly, we demonstrate that the ubiquitin ligase activity is regulated by PI(4)P. These observations lead us to propose a model for SidC activation on the LCV surface (Figure 3.19). In this model, SidC/SdcA is translocated into the host in an inactive conformation. The presence of PI(4)P on the LCV allows for high affinity binding via the P4C domain, thus anchoring of SidC to the LCV. The binding of PI(4)P triggers a conformational change that extends the P4C domain, which is connected to the SNL domain through a long linker peptide, away from the SNL domain. The separation of the P4C domain from the SNL domain causes the exposure of the ubiquitin ligase active site. This model further implies a mechanism for the tight control of the ubiquitin E3 ligase activity of SidC. It is likely that SidC is active only when it is properly anchored on the surface of the LCV, which allows specific ubiquitination of proteins in the vicinity of SidC.

However, an interesting question remains as what are the specific substrates ubiquitinated by SidC/SdcA. Since SidC/SdcA is involved in the recruitment of ER-derived vesicles to the LCV, factors that regulate host ER-related membrane trafficking are likely candidates for ubiquitination by SidC/SdcA. Clearly, the identification of substrates of SidC/SdcA will not only aid in understanding the functional role of bacterial virulent effectors, but will also open a new path for the study of host membrane trafficking.

Compared to the previously reported N-terminal structure of SidC (Hsu & Luo et al., 2014), our new structure also revealed a hinge bending motion between the INS domain and the SNL domain. This hinge bending motion is reminiscent of a similar flexible motion between the N- and the C-lobes of the HECT family of ubiquitin ligases (Lin et al., 2012; Verdecia et al., 2003). We hypothesized that the INS domain may directly mediate the binding of E2. Indeed, in our size exclusion chromatography analysis, stable ubiquitin-charged UbcH7 co-fractionated with the wild type N-terminal part of SidC (1-542), but not with its INS domain deleted counterpart. These results suggest that the INS domain mediates the interaction of SidC with E2s. This argument also explains our previous observations that SidC and SdcA prefer different E2s for their ligase activity (Hsu & Luo et al., 2014). SidC and SdcA are highly homologous at the primary sequence level, except in the INS domain region. The sequence variability of the INS domain is suggestive of the discriminative binding of E2s by SidC and SdcA. However, how SidC/SdcA bind to their preferred E2s remains to be investigated. The E3 ligase IpaH from *S. flexneri* has been shown to bind to E2s at a unique site that is not used by mammalian E3s (Singer et al., 2008). As a future direction, a structural complex of SidC with ubiquitin-charged UbcH7 will be pursued to dissect the molecular mechanism for the interaction between SidC/SdcA with E2s.

Our *Legionella* infection experiments show that the binding of PI(4)P via the P4C domain is required for the anchoring of SidC to, and its function on, the LCV. Although the mutants defective for PI(4)P binding still contain an intact ubiquitin E3 ligase domain, they lose the ability to facilitate the recruitment of ubiquitinated species and ER proteins to the LCV. These results indicate that proper spatial localization of some effectors is crucial for their function during infection. In agreement with this notion, reducing the levels of PI(4)P on the LCV, which likely interferes with proper effector anchoring, has been shown to impair intracellular bacterial replication (Hubber et al., 2014). Mutations in either the cationic pocket or the MIM that totally abrogate PI(4)P binding in vitro do not completely abolish the association of SidC with the LCV. Moreover, a mutant lacking all PI(4)P-binding determinants (R652Q/L1/L2) shows no further reduction of LCV association compared to single mutations (Figure 3.16). This residual binding of SidC to the LCV can be explained by a possible scenario in which SidC interacts with other LCV associated proteins. In fact, it is common to find that the specific membrane targeting of PI binding proteins is governed by a so-called coincidence detection mechanism (Di Paolo et al., 2006). Nevertheless, the residual binding of SidC mutants to the LCV do not appear to play an important role in infection, as it restores neither the ubiquitin signals nor the ER proteins recruited to the bacterial phagosome governed by SidC/SdcA.

PI(4)P has emerged as a direct regulator of several cellular processes, including anterograde trafficking from the Golgi complex (Graham et al., 2011). The use of fluorescent protein tags fused to protein domains that recognize PI(4)P in vivo has allowed the mapping of this lipid to multiple cellular organelles, such as the Golgi apparatus, the endosomal system, and the plasma membrane (Balla et al., 2005; D'Angelo et al., 2008; Roy et al., 2004).A sensitive

and unbiased PI(4)P probe is essential for the studies of PI(4)P dynamics in living cells. However, current PI(4)P probes have their limitations. For example, the pleckstrin homology (PH) domain of Four-phosphate-adaptor protein 1 (FAPP1) also interacts with ARF1, a Golgi-localized member of the Ras family of small GTPases, thus biasing its localization towards the Golgi complex (Balla et al., 2005; Godi et al., 2004; He et al., 2011). Similarly, this major caveat is also true for the PH domain of Oxysterol-binding protein (OSBP) (Levine et al., 2002). Besides interacting with ARF1, the PH domains of FAPP1 and OSBP also recognize PI(4,5)P2 in vitro (Levine et al., 2002; Rameh et al., 1997). These drawbacks render these PH domain-based PI(4)P sensors less faithful probes. Recently, a new PI(4)P probe derived from the Legionella effector protein SidM (DrrA) was reported (Hammond et al., 2014). The C-terminal P4M domain has a unique structure and has been shown to bind PI(4)P with strong affinity and specificity (Schoebel et al., 2010; Del Campo et al., 2014; Zhu et al., 2010). The in vivo application of this novel probe revealed pools of PI(4)P at Rab7-positive late endosomes/lysosomes beyond the Golgi (Hammond et al., 2014). Here we provide preliminary evidence that the P4C domain of SidC has a high affinity and specificity for PI(4)P binding. Compared with other commonly used PI(4)P probes, the P4C domain appears to have a stronger affinity for PI(4)P-positive liposomes (Figure 3.11). Furthermore, the P4C domain displays high selectivity for PI(4)P (Figure 3.6). The GFP-tagged P4C domain exhibits high preference for both the Golgi complex and the plasma membrane, where PI(4)P are enriched. Although more quantitative analysis is warranted to demonstrate the superiority of this novel PI(4)P probe, our data suggest that the P4C domain of SidC has a promising potential to be developed into anfaithful accurate and sensitive in vivo PI(4)P probe.

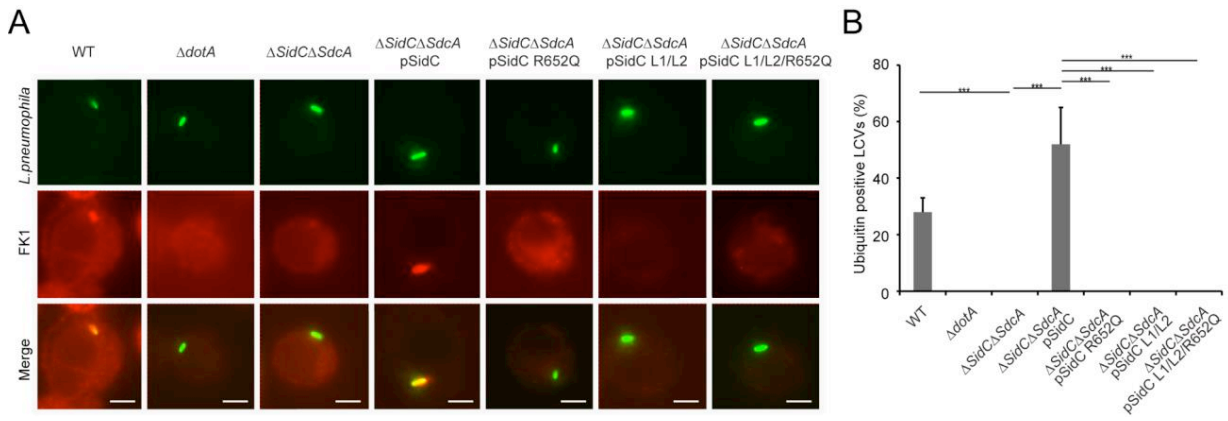


Figure 3.17. PI(4)P-binding by the P4C domain is essential for the recruitment of ubiquitinated species to the LCV.

(A) Immuno-fluorescent staining of ubiquitinated species on the LCV. U937 Cells were infected with indicated *L. pneumophila* strains at an MOI of 1 for 2hrs and samples were fixed prior to immunostaining with antibodies against *Legionella* or m-ubiquitin (FK1). Strains: WT: *L. pneumophila* Philadelphia-1 strain Lp02; *dotA*: the type IV secretion system defective strain Lp03; Δ*sidC-sdcA*: the *SidC* and *SdcA* double deletion mutant of the Lp02 strain; Δ*sidC-sdcA*(pSidC), Δ*sidC-sdcA*(pSidC R652Q), Δ*sidC-sdcA*(pSidC L1/L2), and Δ*sidC-sdcA*(pSidC L1/L2/R652Q): Δ*sidC-sdcA* strain complemented with a plasmid expressing either wild type or PI(4)P binding defective mutants. (B) Percentage of cells containing ubiquitin positive LCVs counted from three independent experiments (at least 150 vacuoles were scored in each experiment). ** $P < 0.01$; *** $P < 0.001$.

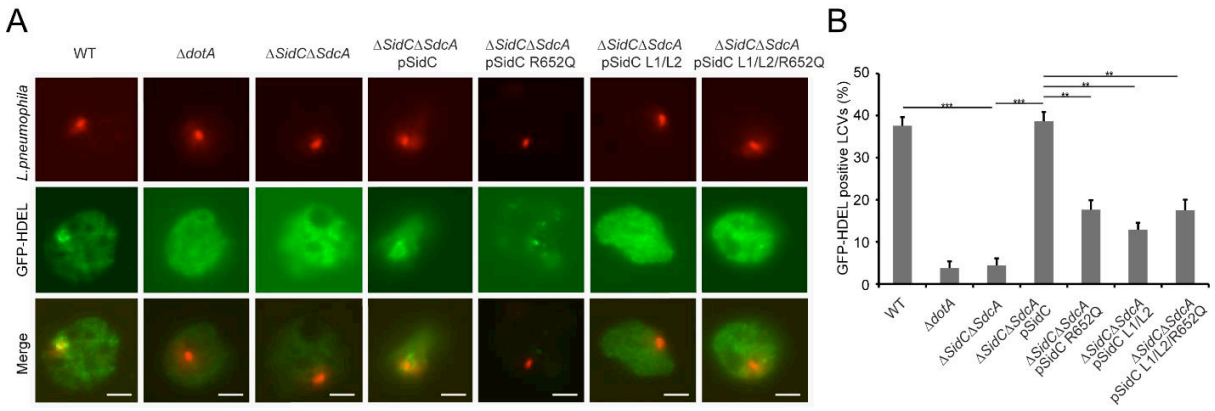


Figure 3.18. PI(4)P-binding by the P4C domain is involved for the ER marker recruitment to the bacterial phagosome.

(A) Images show the recruitment of the ER marker GFP-HDEL (green) to the LCVs in *D. discoideum* cells infected with the indicated Legionella strains (red). Scale bars, 2 μm. Legionella strains are the same as used in Figures S7 and S8. (B) Percentage of cells containing GFP-HDEL positive LCVs counted from three independent assays under the conditions infected with the indicated Legionella strains.

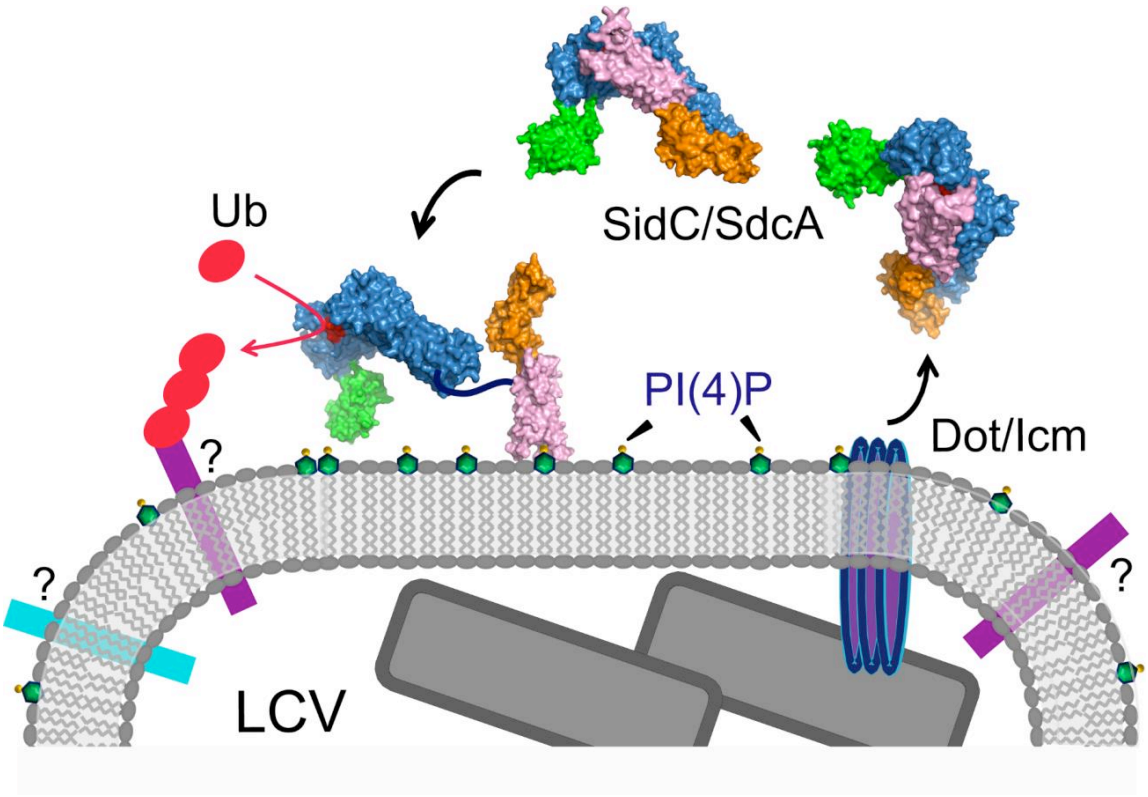


Figure 3.19. A schematic model of SidC functions at the LCV surface.

SidC/SdcA is translocated into the cytosol through the Dot/Icm apparatus. SidC anchors to the LCV membrane through the binding to PI(4)P by its P4C domain. The binding of PI(4)P allows the P4C domain to move away from the SNL domain and exposes the ubiquitin ligase active site. The activated SidC/SdcA ubiquitinates unknown factors, presumably proteins residing on the LCV that determine the identity of the bacterial phagosome.

Materials and Methods

Cloning and mutagenesis. PCR products for SidC (aa. 1-871) and P4C of SidC (aa. 614-743) amplified from full length SidC gene were digested with BamHI and XhoI restriction enzymes and inserted into a pET28a-based vector in frame with an N-terminal His-SUMO tag (Hsu & Luo et al., 2014). To generate bacteria expressing GFP tagged proteins, the same vector was modified by inserting a GFP tag between the His-SUMO tag and the downstream MCS region. Single or multiple amino acid substitutions of SidC were introduced by in vitro site-directed mutagenesis using oligonucleotide primer pairs containing the appropriate base substitutions. SidC542ΔINS was generated by ligating two PCR fragments corresponding to residues 1-224 and 324-542. For mammalian expression, wild type and mutant SidC P4C, the PH domain of OSBP, the PH domain of FAPP1 and the P4M domain of SidM were subcloned into the pEGFP-N1 vector or the modified pEGFP-N1 vector with the *egfp* replaced by the mCherry gene. All constructs were confirmed by DNA sequencing.

Protein Expression and Purification. *E. coli* Rosetta strains harboring the expression plasmids were grown in Luria-Bertani medium supplemented with 50 μg/ml kanamycin to mid-log phase. Protein expression was induced for overnight at 18 °C with 0.1 mM isopropyl-B-D-thiogalactopyranoside (IPTG). Harvested cells were resuspended in a buffer containing 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, and were lysed by sonication. Soluble fractions were collected by centrifugation at 18,000 rpm for 30 min at 4 °C and incubated with cobalt resins (Golden-Bio) for 1 h at 4 °C. Protein bound resins were extensively washed with the lysis buffer. The His-SUMO tag was cleaved by the SUMO-specific protease Ulp1 to release SidC

from the resin. Eluted protein samples were further purified by FPLC size exclusion chromatography. The peak corresponding to the expressed protein was pooled and concentrated to 10 mg/ml in a buffer containing 20 mM Tris, pH 7.5, and 50 mM NaCl. SidC542, SidC542 ΔINS, SidC743, SidC743 L629R, SidC743 R652Q, SidC743 L1/L2 were purified with a similar protocol as described above. Ubiquitin, ubiquitin activating enzyme E1, and hUbcH7C85S were expressed and purified as previously described (Hsu & Luo et al., 2014).

SidC-Ub~UbcH7 ternary complex formation. Catalytically inactive ubiquitin conjugating enzyme E2, hUbcH7C85S was charged with ubiquitin at 37 °C for 4 h in the presence of 0.5 μM E1, 20 μM E2, and 25 μM ubiquitin in the reaction buffer containing 50 mM Tris-HCl (pH 8.5), 5 mM MgCl2, 0.5 mM DTT, 50 mM creatine phosphate (Sigma P7396), 3 U/ml of pyrophosphatase (Sigma I1643), 3 U/ml of creatine phosphokinase, 2.8 mM ATP. After the charging, SidC542, SidC542 ΔINS, SidC743, or SidC743 L629R was added to the reaction and incubated at 4 °C for 3 h. The final mixture was loaded onto a Superdex 200 size exclusion column (GE Life Sciences). Eluted fractions were analyzed by SDS-PAGE.

In vitro ubiquitin E3 ligase assay. Time course ubiquitination assays were performed at 37 °C in the presence of 80 nM E1, 2.8 μM E2, 1 μM SidC truncations or mutants, and 10 μM ubiquitin in a reaction buffer containing 50 mM Tris-HCl (pH 8.0), 5 mM MgCl2, 0.5 mM DTT, 50 mM creatine phosphate (Sigma P7396), 3 U/ml of pyrophosphatase (Sigma I1643), 3 U/ml of creatine phosphokinase, 2.8 mM ATP. For ubiquintin comsumption assays, SidC proteins were pre-incubated with 15 μl 1mM liposome suspensions containing PC/PS or PC/PS/PI(4)P for 30 min at room temperature before the ubiquitination reactions. All reactions were stopped by the

addition of 5X SDS-PAGE loading buffer containing 250 mM BME and seperated by SDS-PAGE. All the SDS gels were stained by Coomassie blue dye and the intensity of the ubiquitin protein bands was quantified using an Odyssey Infrared imaging system (LI-COR Biosciences). These assays were repeated in three independent experiments. The percentage of ubiquitin usage were averaged and plotted.

Crystallization, data collection, and processing. Crystals were grown at room temperature by the hanging-drop vapor diffusion method by mixing 1 μl of protein (10 mg/ml) with an equal volume of reservoir solution containing 20 % PEG 8,000, pH 7.3, 0.1 M HEPES. Sheet-shaped crystals were formed within 2-3 days. The crystals were further optimized by microseeding. Crystals were soaked in a cryo-protection solution containing the reservoir solution with 20% glycerol. Soaked crystals were flash frozen in liquid nitrogen before data collection. Diffraction data sets were collected at the Cornell synchrotron light source, MacCHESS beam line A1. The data sets were indexed, integrated, and scaled with HKL-2000 (Otwinowski et al., 1997). The crystal belongs to the space group C2 with cell parameters as: a = 228.16 Å; b = 83.934 Å; c = 129.4 Å; $\alpha = 90^0$, $\beta = 108.82^0$, $\gamma = 90^0$ (Table 3.1). The calculated Matthews coefficient Vm= 2.99 and with 58.92% of solvent in the crystal and two protein molecules in an asymmetric unit (Matthews et al., 1968).

Structure determination and refinement. The crystal structure of SidC871 was determined by molecular replacement using the SidC542 structure (PDB ID: 4TRH) as the search model with the AMoRe program (Trapani et al., 2008) in the CCP4 suite (Collaborative Computational Project N, 1994). Iterative cycles of model building and refinement were carried out with the

program COOT (Emsley et al., 2004) and the refmac5 program (Murshudov et al., 1997) in the CCP4 suite (Collaborative Computational Project N, 1994). The final model is quality checked with the procheck program (Laskowski et al., 1993).

Liposome preparation, flotation assay, and liposome imaging. Phosphatidic acid (PA), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine (POPS) and di-C16-phophatidylinositol polyphosphates were purchased from Avanti. Unilamellar liposomes were generated from a mixture of 64 μg of POPC (80%), 8 μg PS (10%), and 10 μg (10%) of other phospholipids [PI(3)P, PI(4)P, PI(4,5)P_2, or PA] with addition of 0.08 μg the near-infrared Dil dye (invitrogen) to aid the visualization and quantitation of lipids. Following air drying, lipid films were hydrated in 50 mM Tris (pH 7.0), 150 mM NaCl followed by 1h incubation at 37 °C for the spontaneous formation of liposomes.

Liposome flotation assays were performed as previously described (Matsuoka et al., 2000). Briefly, 10 μL of liposome suspension was mixed with 2 μg proteins and the mixture was brought to 80 μL with HK buffer (20 mM HEPES PH 7.4, 150 mM KOAc) for liposome binding. After 30 min of incubation, 50 μL of 2.5 M sucrose in HK buffer was added to the binding reaction and mixed thoroughly. 30 μL of the sample was taken out as inputs and 100 μL of the sample was transferred to a Beckman 7x20 mm PC ultracentrifuge tube to form the bottom heavy layer. On top of this layer, 100 μL of 0.75 M sucrose-HK buffer was gently overlayed as the middle layer, followed by 20 μL HK buffer as the top layer. Liposomes were separated from unbound protein by ultracentrifugation at 100,000 rpm for 20 minutes at 20ºC in a TLA-100 ultracentrifuge rotor. 30 μL of sample from the top layer was collected and bound proteins were analyzed by SDS-PAGE. Sample loading was normalized based on lipids recovery as measured

by the fluorescence of the Dil dye. Western blots were then performed with anti-GFP antibody (1:5,000) to detect the GFP-tagged proteins. Fluorescence intensity of the protein bands was quantified using an Odyssey Infrared imaging system (LI-COR Biosciences). Floatation assays were repeated in three independent experiments. The protein recovery ratios were averaged and plotted.

For liposome imaging, 20 μL of ten-fold diluted liposome suspension was mixed with 20 μL protein solutions containing 0.4 μg proteins for liposome binding. After a 20 min incubation at RT, 5μL of the reaction mixture was added to the chamber created between a cover slip and a glass slide. Fluorescence microscopy images were acquired using a Zeiss Observer inverted microscope (Carl Zeiss) and were analyzed using the Zen software (Carl Zeiss). Fluorescence intensity of the liposomes was quantified using ImageJ. The quantification of the fluorescence intensities were calculated and averaged from three randomly selected liposomes.

Cell culture, transfection, and fluorescence microscopy. N2A (ATCC) or Cos7 (ATCC) cells were maintained and grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS (Cellgro) and 0.1% Pen/Strep (Cellgro) at 37 °C in 5% CO2 atmosphere. Mouse bond marrow-derived macrophages were prepared as described (Kagan et al., 1995). Cells were transfected with 0.5μg of each plasmid by using polyethyleneimine (PEI) reagent at a ratio of 1:5. Cells transfected overnight were fixed with 4% formaldehyde. Fluorescence microscopy images were acquired either using a Zeiss Observer inverted microscope (Carl Zeiss) for epi-fluorescence images or a Zeiss LSM 700 confocal microscope (Carl Zeiss) equipped with a ×63 oil immersion objective (Carl Zeiss). Images were then analyzed using the Zen software (Carl Zeiss). Fluorescence intensity of cell images was quantified using the ImageJ software. The

fluorescence intensities at the cell periphery (for PM) and the perinuclear region (PNR) were integrated. The fluorescence intensity of other area of the cell was calculated by subtracting the intensities of PM and PNR from the total integrated intensity of the whole cell. The final fluorescence intensities were averaged and plotted from three randomly selected cells.

***Legionella* strains and infection.** *Legionella pneumophila* strains used were Lp02 (Berger et al., 1993), Lp03 (Berger et al., 1993) defective for the Dot/Icm secretion system and Lp02ΔSidC/SdcA, which lacks both SidC and SdcA (Hsu & Luo et al., 2014). Bacteria were grown on N-(2-acetamido)-2-aminoethanesulfonic acid (ACES)-buffered charcoal-yeast extract agar (CYE) or on CYE supplemented with 100 μg of thymidine/ml (CYET) when necessary (Conover et al., 2003). The mutations were introduced into *sidC* on plasmid pZL199 (VanRheenen et al., 2006). All mutants were verified by DNA sequencing analysis. The plasmids were transformed into strain Lp02ΔSidC/SdcA. For infection, bacteria were grown to post-exponential phase as judged by motility and by growth phase (OD600=3.4-3.80). An MOI of 1 was used for infections with the U937 human macrophages and for mouse bone marrow-derived macrophages. The *Dictyostelium discoideum* strain AX4 stably expressing GFP-HDEL for assessing the ER membrane conversion phenotype was grown and maintained as previously described (Liu et al., 2007). *D. discoideum* seeded on glass coverslips in MLB medium (Liu et al., 2007) was infected with properly grown bacteria at an MOI of 10. One hour after infections, infected cells were washed three times with PBS to remove extracellular bacteria and samples fixed after incubation for an additional hour. *L. pneumophila* was labeled with specific antibodies (Xu et al., 2010) before immunostaining with the FK1 antibody (BIOMOL International/Affiniti, Exeter, United Kingdom) (Dorer et al., 2006) or the SidC specific antibody

(Luo et al., 2004). Prepared samples were mounted onto slide and inspected under a fluorescence microscope (Olympus IX-81) and images were acquired with a CCD camera.

Analysis of SidC immuno-fluorescence signals. Images were acquired with an upright Nikon Eclipse 90i laser-scanning confocal microscope equipped with a Nikon 60x oil Plan Apo vc objective. The 488 and 561 nm laser lines (Sapphire Coherent) were used to acquire images of SidC and *Legionella* bacteria, respectively. Only cells that contained bacterial vacuoles with overlapping SidC signals were chosen for imaging. At least 50 fields for each of three coverslips (at least 150 cells) were imaged for each condition. To analyze the signal intensity, integrated intensities of the SidC signal within each LCV region were collected from each field and were averaged per coverslip. The three mean values from each coverslip were then averaged to generate a mean of means +/- SEM. Due to the presence of unequal variances between the strains being compared, raw integrated intensities for each vacuole were log-transformed (log[x+1]), and then re-averaged. Data was analyzed with GraphPad Prism 5 software, using one-way ANOVA followed by Dunnett's post-hoc test to compare all conditions.

Acknowledgements

The authors would like to thank Dr. Peter Hollenbeck and Ms. Elisabeth Garland-Kuntz (Purdue University, West Lafayette, IN) for microscopy analysis.

REFERENCES

Asrat S, de Jesus DA, Hempstead AD, Ramabhadran V, Isberg RR (2014) Bacterial Pathogen Manipulation of Host Membrane Trafficking. Annu Rev Cell Dev Biol.

Berger KH, Isberg RR (1993) Two distinct defects in intracellular growth complemented by a single genetic locus in Legionella pneumophila. Mol Microbiol 7: 7-19.

Balla A, Tuymetova G, Tsiomenko A, Varnai P, Balla T (2005) A plasma membrane pool of phosphatidylinositol 4-phosphate is generated by phosphatidylinositol 4-kinase type-III alpha: studies with the PH domains of the oxysterol binding protein and FAPP1. Mol Biol Cell 16: 1282-1295.

Brombacher E, Urwyler S, Ragaz C, Weber SS, Kami K, et al. (2009) Rab1 guanine nucleotide exchange factor SidM is a major phosphatidylinositol 4-phosphate-binding effector protein of Legionella pneumophila. J Biol Chem 284: 4846-4856.

Cho W, Stahelin RV (2005) Membrane-protein interactions in cell signaling and membrane trafficking. Annu Rev Biophys Biomol Struct 34: 119-151.

Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. Acta Cryst D: 760-763.

Conover GM, Derre I, Vogel JP, Isberg RR (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.

D'Angelo G, Vicinanza M, Di Campli A, De Matteis MA (2008) The multiple roles of PtdIns(4)P -- not just the precursor of PtdIns(4,5)P2. J Cell Sci 121: 1955-1963.

Del Campo CM, Mishra AK, Wang YH, Roy CR, Janmey PA, et al. (2014) Structural basis for PI(4)P-specific membrane recruitment of the Legionella pneumophila effector DrrA/SidM. Structure 22: 397-408.

Di Paolo G, De Camilli P (2006) Phosphoinositides in cell regulation and membrane dynamics. Nature 443: 651-657.

Dorer MS, Kirton D, Bader JS, Isberg RR (2006) RNA interference analysis of Legionella in Drosophila cells: exploitation of early secretory apparatus dynamics. PLoS Pathog 2: e34.

Dowler S, Currie RA, Campbell DG, Deak M, Kular G, et al. (2000) Identification of pleckstrin-homology-domain-containing proteins with novel phosphoinositide-binding specificities. Biochem J 351: 19-31.

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

Fields BS, Benson RF, Besser RE (2002) Legionella and Legionnaires' disease: 25 years of investigation. Clin Microbiol Rev 15: 506-526.

Gazdag EM, Schobel S, Shkumatov AV, Goody RS, Itzen A (2014) The structure of the N-terminal domain of the Legionella protein SidC. J Struct Biol.

Godi A, Di Campli A, Konstantakopoulos A, Di Tullio G, Alessi DR, et al. (2004) FAPPs control Golgi-to-cell-surface membrane traffic by binding to ARF and PtdIns(4)P. Nat Cell Biol 6: 393-404.

Graham TR, Burd CG (2011) Coordination of Golgi functions by phosphatidylinositol 4-kinases. Trends Cell Biol 21: 113-121.

Hammond GR, Machner MP, Balla T (2014) A novel probe for phosphatidylinositol 4-phosphate reveals multiple pools beyond the Golgi. J Cell Biol 205: 113-126.

He J, Scott JL, Heroux A, Roy S, Lenoir M, et al. (2011) Molecular basis of phosphatidylinositol 4-phosphate and ARF1 GTPase recognition by the FAPP1 pleckstrin homology (PH) domain. J Biol Chem 286: 18650-18657.

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

Hsu F, Zhu W, Brennan L, Tao L, Luo ZQ, et al. (2012) Structural basis for substrate recognition by a unique Legionella phosphoinositide phosphatase. Proc Natl Acad Sci U S A 109: 13567-13572.

Hsu F, Luo X, Qiu J, Teng YB, Jin J, et al. (2014) The Legionella effector SidC defines a unique family of ubiquitin ligases important for bacterial phagosomal remodeling. Proc Natl Acad Sci U S A 111: 10538-10543.

Horenkamp FA, Mukherjee S, Alix E, Schauder CM, Hubber AM, et al. (2014) Legionella pneumophila Subversion of Host Vesicular Transport by SidC Effector Proteins. Traffic.

Hubber A, Arasaki K, Nakatsu F, Hardiman C, Lambright D, et al. (2014) The machinery at endoplasmic reticulum-plasma membrane contact sites contributes to spatial regulation of multiple Legionella effector proteins. PLoS Pathog 10: e1004222.

Isberg RR, O'Connor TJ, Heidtman M (2009) The Legionella pneumophila replication vacuole: making a cosy niche inside host cells. Nat Rev Microbiol 7: 13-24.

Ingmundson A, Delprato A, Lambright DG, Roy CR (2007) Legionella pneumophila proteins that regulate Rab1 membrane cycling. Nature 450: 365-369.

Kagan JC, Roy CR (2002) Legionella phagosomes intercept vesicular traffic from endoplasmic reticulum exit sites. Nat Cell Biol 4: 945-954.

Kamadurai HB, Souphron J, Scott DC, Duda DM, Miller DJ, et al. (2009) Insights into ubiquitin transfer cascades from a structure of a UbcH5B approximately ubiquitin-HECT(NEDD4L) complex. Mol Cell 36: 1095-1102.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26: 283-291.

Levine TP, Munro S (2002) Targeting of Golgi-specific pleckstrin homology domains involves both PtdIns 4-kinase-dependent and -independent components. Curr Biol 12: 695-704.

Lin DY, Diao J, Chen J (2012) Crystal structures of two bacterial HECT-like E3 ligases in complex with a human E2 reveal atomic details of pathogen-host interactions. Proc Natl Acad Sci U S A 109: 1925-1930.

Liu Y, Luo ZQ (2007) The Legionella pneumophila effector SidJ is required for efficient recruitment of endoplasmic reticulum proteins to the bacterial phagosome. Infect Immun 75: 592-603.

Luo ZQ, Isberg RR (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.

Machner MP, Isberg RR (2006) Targeting of host Rab GTPase function by the intravacuolar pathogen Legionella pneumophila. Dev Cell 11: 47-56.

Matsuoka K, Schekman R (2000) The use of liposomes to study COPII- and COPI-coated vesicle formation and membrane protein sorting. Methods 20: 417-428.

Matthews BW (1968) Solvent content of protein crystals. J Mol Biol 33: 491-497.

Moravcevic K, Oxley CL, Lemmon MA (2012) Conditional peripheral membrane proteins: facing up to limited specificity. Structure 20: 15-27.

Muller MP, Peters H, Blumer J, Blankenfeldt W, Goody RS, et al. (2010) The Legionella effector protein DrrA AMPylates the membrane traffic regulator Rab1b. Science 329: 946-949.

Murata T, Delprato A, Ingmundson A, Toomre DK, Lambright DG, et al. (2006) The Legionella pneumophila effector protein DrrA is a Rab1 guanine nucleotide-exchange factor. Nat Cell Biol 8: 971-977.

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53: 240-255.

Nash TW, Libby DM, Horwitz MA (1984) Interaction between the legionnaires' disease bacterium (Legionella pneumophila) and human alveolar macrophages. Influence of antibody, lymphokines, and hydrocortisone. J Clin Invest 74: 771-782.

Neunuebel MR, Chen Y, Gaspar AH, Backlund PS, Jr., Yergey A, et al. (2011) De-AMPylation of the small GTPase Rab1 by the pathogen Legionella pneumophila. Science 333: 453-456.

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

Ragaz C, Pietsch H, Urwyler S, Tiaden A, Weber SS, et al. (2008) The Legionella pneumophila phosphatidylinositol-4 phosphate-binding type IV substrate SidC recruits endoplasmic reticulum vesicles to a replication-permissive vacuole. Cell Microbiol 10: 2416-2433.

Rameh LE, Arvidsson A, Carraway KL, 3rd, Couvillon AD, Rathbun G, et al. (1997) A comparative analysis of the phosphoinositide binding specificity of pleckstrin homology domains. J Biol Chem 272: 22059-22066.

Roy A, Levine TP (2004) Multiple pools of phosphatidylinositol 4-phosphate detected using the pleckstrin homology domain of Osh2p. J Biol Chem 279: 44683-44689.

Schoebel S, Blankenfeldt W, Goody RS, Itzen A (2010) High-affinity binding of phosphatidylinositol 4-phosphate by Legionella pneumophila DrrA. EMBO Rep 11: 598-604.

Segal G, Purcell M, Shuman HA (1998) Host cell killing and bacterial conjugation require overlapping sets of genes within a 22-kb region of the Legionella pneumophila genome. Proc Natl Acad Sci U S A 95: 1669-1674.

Sherwood RK, Roy CR (2013) A Rab-centric perspective of bacterial pathogen-occupied vacuoles. Cell Host Microbe 14: 256-268.

Singer AU, Rohde JR, Lam R, Skarina T, Kagan O, et al. (2008) Structure of the Shigella T3SS effector IpaH defines a new class of E3 ubiquitin ligases. Nat Struct Mol Biol 15: 1293-1301.

Stenmark H (2009) Rab GTPases as coordinators of vesicle traffic. Nat Rev Mol Cell Biol 10: 513-525.

Swanson MS, Isberg RR (1995) Association of Legionella pneumophila with the macrophage endoplasmic reticulum. Infect Immun 63: 3609-3620.

Tan Y, Luo ZQ (2011) Legionella pneumophila SidD is a deAMPylase that modifies Rab1. Nature 475: 506-509.

Tan Y, Arnold RJ, Luo ZQ (2011) Legionella pneumophila regulates the small GTPase Rab1 activity by reversible phosphorylcholination. Proc Natl Acad Sci U S A 108: 21212-21217.

Tilney LG, Harb OS, Connelly PS, Robinson CG, Roy CR (2001) How the parasitic bacterium Legionella pneumophila modifies its phagosome and transforms it into rough ER: implications for conversion of plasma membrane to the ER membrane. J Cell Sci 114: 4637-4650.

Trapani S, Navaza J (2008) AMoRe: classical and modern. Acta Crystallogr D Biol Crystallogr 64: 11-16.

Toulabi L, Wu X, Cheng Y, Mao Y (2013) Identification and structural characterization of a Legionella phosphoinositide phosphatase. J Biol Chem 288: 24518-24527.

Vogel JP, Andrews HL, Wong SK, Isberg RR (1998) Conjugative transfer by the virulence system of Legionella pneumophila. Science 279: 873-876.

Watzele G, Bachofner R, Berger EG (1991) Immunocytochemical localization of the Golgi apparatus using protein-specific antibodies to galactosyltransferase. Eur J Cell Biol 56: 451-458.

Weber SS, Ragaz C, Hilbi H (2009) The inositol polyphosphate 5-phosphatase OCRL1 restricts intracellular growth of Legionella, localizes to the replicative vacuole and binds to the bacterial effector LpnE. Cell Microbiol 11: 442-460.

Weber S, Wagner M, Hilbi H (2014) Live-cell imaging of phosphoinositide dynamics and membrane architecture during Legionella infection. MBio 5: e00839-00813.

Weber SS, Ragaz C, Reus K, Nyfeler Y, Hilbi H (2006) Legionella pneumophila exploits PI(4)P to anchor secreted effector proteins to the replicative vacuole. PLoS Pathog 2: e46.

VanRheenen SM, Luo ZQ, O'Connor T, Isberg RR (2006) Members of a Legionella pneumophila family of proteins with ExoU (phospholipase A) active sites are translocated to target cells. Infect Immun 74: 3597-3606.

Verdecia MA, Joazeiro CA, Wells NJ, Ferrer JL, Bowman ME, et al. (2003) Conformational flexibility underlies ubiquitin ligation mediated by the WWP1 HECT domain E3 ligase. Mol Cell 11: 249-259.

Xu L, Shen X, Bryan A, Banga S, Swanson MS, et al. (2010) Inhibition of host vacuolar H+-ATPase activity by a Legionella pneumophila effector. PLoS Pathog 6: e1000822.

Yagisawa H, Hirata M, Kanematsu T, Watanabe Y, Ozaki S, et al. (1994) Expression and characterization of an inositol 1,4,5-trisphosphate binding domain of phosphatidylinositol-specific phospholipase C-delta 1. J Biol Chem 269: 20179-20188.

Zhu Y, Hu L, Zhou Y, Yao Q, Liu L, et al. (2010) Structural mechanism of host Rab1 activation by the bifunctional Legionella type IV effector SidM/DrrA. Proc Natl Acad Sci U S A 107: 4699-4704.

Zhu W, Banga S, Tan Y, Zheng C, Stephenson R, et al. (2011) Comprehensive identification of protein substrates of the Dot/Icm type IV transporter of Legionella pneumophila. PLoS One 6: e17638.

CHAPTER IV

Conclusions and future direction

The opportunistic intracellular pathogen *Legionella pneumophila* is the causative agent of Legionnaires' disease. *L. pneumophila* delivers nearly 300 effector proteins into the host cells for the establishment of a replication permissive compartment known as the *Legionella*-containing vacuole (LCV). SidC and its paralogue SdcA are two effectors that have been shown to anchor on the LCV via binding to phosphatidylinositol-4-phosphate [PI(4)P] and facilitate the recruitment of endoplasmic reticulum (ER)-derived vesicles to the LCV. We determined the crystal structure of SidC (1-871). The N-terminal SNL (SidC N-terminal E3 Ligase; 1-542) domain of SidC is a ubiquitin E3 ligase and required for the recruitment of ER proteins, as well as ubiquitinated species to the LCV. We also revealed the P4C (PI(4)P binding of SidC) domain of SidC as a novel PI(4)P specific binding module. Biochemical and cell biological studies highlighted key residues involved in PI(4)P binding and membrane insertion. Upon PI(4)P binding, the P4C domain of SidC shifts away from the SNL domain and releases its inhibitory function. SidC FL therefore switches its conformation from closed to open and is activated. In addition, we found that proper spatial localization of SidC to the cytoplasmic surface of the bacterial phagosome by binding to PI(4)P was crucial for its function. Challenges remain to identify the biological substrate of SidC/SdcA to understand the molecular mechanisms of how SidC interrupts host cellular processes. Another exciting future direction is further structural studies of SidC/E2 and SidC/PI(4)P interactions, as well as applications of the novel P4C probe.

The first exciting future direction is to dissect the molecular mechanism for the interaction between SidC/SdcA with E2s by pursuing the structural complex of SidC with stably ubiquitin-charged UbcH7 and determining the E2 binding interface. We discovered that the N-terminal SNL domain of SidC possesses a cysteine based ubiquitin ligase activity. Compared to other cysteine based ubiquitin ligases, the SNL domain family has some unique features. First, the SNL domain has no detectable primary and tertiary structural similarity to any known protein, suggesting a novel fold of this domain. Second, the SNL domain has a conserved catalytic C-H-D triad. The aspartic and histidine residues within the triad likely render the catalytic cysteine a stronger nucleophile than just a single catalytic cysteine found in other ubiquitin ligases. This feature may enhance the kinetics of Ub transfer from E2s to SidC and may also be advantageous during the competition with host Ub ligases for the pool of activated E2~Ub. Third, in HECT E3s, the catalytic cysteine is located at the C-terminal small lobe, which can move a large distance towards the N-terminal lobe through a hinge-like motion. This conformational flexibility is believed to allow the transfer of Ub from E2 to the E3 catalytic cysteine. However, the catalytic cysteine of the SNL domain is localized at the center of a surface of its main sub-domain. This structural organization suggests that the Ub transfer by the SNL domain may not involve large movements of the catalytic cysteine, which further indicates that the binding site for E2~Ub is in close proximity to the catalytic cysteine. To address this hypothesis, future structural and biochemical experiments are warranted to identify the E2 binding site on the SNL domain.

Although SidC and its paralogue SdcA share 72% sequence identity, they exhibit differential preference for ubiquitin E2 conjugating enzymes. This discrimination of E2s suggests that these two proteins have evolved to maximally exploit components in the host

ubiquitin pathway. Structural and biochemical analysis of SdcA will be preformed to find the molecular mechanism for their different E2 preference and to explain why *L. pneumophila* has maintained two such highly similar proteins in evolution.

Although our structure of the SidC P4C domain defines the fold and locates the PI(4)P binding site, it contains no ligand in the binding pocket. Thus, the structural basis for stereospecific recognition of PI(4)P remains unclear. The crystal structure of SidC P4C domain in complex with dibutyl PI(4)P needs to be determined. It will shed light on explaining how the headgroup of PI(4)P tightly coordinated in the deep binding pocket of SidC P4C and what residues and interactions contribute to the high affinity and specificity of binding.

One significance of our research is that the P4C probe from SidC is a more sensitive and unbiased biosensor for detecting PI(4)P in living cells comparing with other reported PI(4)P probes. People could apply the P4C probe to study a broad range of PI(4)P related intracellular events. One possible direction is the study of PI(4)P metabolisms in yeast. PI(4)P is synthesized by specific PI-4-kinases at the Golgi and the PM, yet its phosphatase, Sac1, is an integral ER membrane protein (Kim et al., 2013). Two recent publications proposed opposite mechanisms to solve this paradox: that Sac1p hydrolyses PI(4)P in either *trans* (Stefan et al., 2011) or *in cis* (Mesmin et al., 2013) Golgi complex. The intracellular distribution of P4C probe in yeast is also consistent with the PI(4)P enriched regions mainly in Golgi complex and PM, whereas it relocalize to ER in Sac1 deletion strain. It would be interesting to demonstrate where these ER localized PI(4)P come from, either from Golgi or PM, and further investigate the PI(4)P trafficking events in yeast.

We have revealed the biochemical function of SidC as an E3 ubiquitin ligase. But the detailed mechanism of how the E3 ligase activity of SidC is related to the recruitment of host ER proteins remains unclear. It is plausible that some host factors may be involved as specific SidC substrates which are uniquitinated by SidC and so trigger ER vesicles recruitment. These factors most likely function in cellular signaling and membrane trafficking pathways. Therefore, the exciting and challenging future direction of my thesis project is to identify these specific SidC associated unknown factors. We are also very interested in looking for the substrate of SdcA. The identification of different substrates for SidC and SdcA will shed light on how *L. pneumophila* affects host vesicle trafficking pathway through two highly similar but not functional redundant proteins. The identification of unknown factors that interact with SidC/SdcA would help to explain the molecular mechanisms of how SidC affects host cellular processes during the *Legionella* pathogenesis.

REFERENCES

Mesmin, B., J. Bigay, J. Moser von Filseck, S. Lacas-Gervais, G. Drin, and B. Antonny (2013). A four-step cycle driven by PI(4)P hydrolysis directs sterol/PI(4)P exchange by the ER-Golgi tether OSBP. *Cell*. 155:830–843. doi:10.1016/j.cell.2013.09.056

Y.J. Kim, M.L. Hernandez, T. Balla (2013) Inositol lipid regulation of lipid transfer in specialized membrane domains. *Trends Cell Biol.,* 23 pp. 270–278

C.J. Stefan, A.G. Manford, D. Baird, J. Yamada-Hanff, Y. Mao, S.D. Emr (2011) Osh proteins regulate phosphoinositide metabolism at ER-plasma membrane contact sites. *Cell*, 144, pp. 389–401