TALES OF PHOSPHOINOSITIDE-BINDING ENZYMES: UNRAVELING THE PHYSIOLOGICAL ROLE OF HUMAN SAC2 AND LEGIONELLA EFFECTORS, SIDF & SIDC

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
FoSheng Hsu
May 2015
Phosphoinositides (PIs), a group of key signaling and structural lipids, are involved in a myriad of biological processes of cell physiology. Their intracellular concentration and localization must be tightly regulated for proper cell function and survival. This is achieved by specific PI phosphatases and kinases, which are responsible for the spatial and temporal conversion of distinct phosphorylated species. Mutations in several of these metabolic enzymes have been found in a range of neurological diseases, and human cancers, further underscoring the crucial role of PIs. In recent years, it has also become evident that many intracellular pathogens have exploited host PIs for their pathogenesis by secreting eukaryotic-like factors such as PI phosphatases or PI-binding effector proteins. Thus, elucidating the role of mammalian/bacterial PI phosphatases and PI-binding effectors is paramount to the understanding of the host and host-bacterial signaling mechanism.

The first part of the dissertation focuses on the structure, function and disease-associated mutations of mammalian PI phosphatases with emphasis on a family of enzymes containing a conserved enzymatic domain called Sac. We determined that one of the Sac domain-containing proteins, Sac2, is a PI 4-phosphatase that is highly expressed in the brain. Using overexpressed and CRISPR knockout Sac2 neuronal cell lines, we found that Sac2 is an important regulator in the endocytic recycling of surface receptors such as transferrin and integrin receptors, and plays an essential role in maintaining cell migration in neurons. This work uncovered a previously unidentified role of Sac2 and implicated Sac2 in neuronal functions and brain development.
The second part of this dissertation discusses the role of bacterial PI phosphatases in infection, and investigates in detail the role of an effector called SidF from the intracellular pathogen *Legionella pneumophila*. We found that SidF is a PI 3-phosphatase that specifically hydrolyzes PI(3,4)P₂ and PI(3,4,5)P₃ to PI(4)P and PI(4,5)P₂, respectively. We solved the crystal structure of SidF in complex with its substrate, PI(3,4)P₂ and revealed the structural basis for substrate recognition. Finally, we demonstrated that the catalytic activity of SidF is important for sculpting the lipid composition of the *Legionella*-containing vacuole (LCV) during infection. Our work revealed the first PI phosphatase encoded by *Legionella pneumophila* and elucidated its function during infection.

In addition, the second half of Part II addresses specifically the function of the *Legionella* effector SidC in vitro and in vivo, and how the PI(4)P-binding property of SidC facilitates the LCV maturation process. Unexpectedly, we discovered that SidC is an E3 ubiquitin ligase that uses a cysteine in the catalytic triad to generate polyubiquitin chains. The enzymatic activity is necessary for recruiting ubiquitin conjugates and host ER vesicles to the LCV. Our findings illustrated a powerful strategy employed by the *Legionella*, and potentially other intracellular pathogens, to simultaneously utilize both the host PI and ubiquitination system for their pathogenesis.

The final chapter summarizes our current understanding of both the mammalian/bacterial PI phosphatases and PI-binding effectors in *Legionella*, and explores future directions in further expanding our knowledge on host signaling and bacterial infection mechanisms.
BIOGRAPHICAL SKETCH

FoSheng Hsu obtained his Bachelor of Science in Microbiology from the University of Washington in 2009. In 2009, he joined the doctoral program in the graduate field of Biochemistry, Molecular and Cell Biology (BMCB) under the Department of Molecular Biology and Genetics (MBG) at Cornell University.

Dr. Hsu has been the recipient of numerous awards including the “Rita and Joe Calvo Graduate Student Teaching Award” in 2011, and “Harry & Samuel Mann Outstanding Graduate Award” in 2012. Aside from these research and teaching endeavors, Dr. Hsu enjoys expressing scientific ideas through creative art forms. He was the winner of “Dance Your PhD” sponsored by Science for best dance in Chemistry category in 2011. He was also the recipient for the finalist award from LabTV Tribeca Video Contest in 2014 for his outstanding role model film for future scientists.

While pursing his degree, Dr. Hsu was also involved in the Cornell’s Center for Teaching Excellence Graduate Teaching program and held a 2-year position as 1 of 6 Graduate Resident Fellows at Hans Bethe House, where he mentored over 300+ undergraduate residents.

Dr. Hsu has presented his research at domestic and international conferences including the Gordon Conference on Molecular Membrane Biology as well as IRIC International Symposium. Under the guidance and supervision of Dr. Yuxin Mao, his research works have been published in a number of journals including the Proceeding of the National Academy of Sciences and the Journal of Cell Biology.
This dissertation is dedicated to all the people who I have encountered in my life.
It is because of you that I am here today.

I would like to extend a special dedication to my loving and supportive family. My parents, Ting-Chi Hsu and Yuan-Chiao Hung, for their unconditional love. My two brothers, Chao-Wen and Hsiu-Chi Hsu, for their words of encouragement and advice.

Finally, to my loving, kind, and compassionate wife, Sonia Tien, this work is dedicated to you for your unconditional support throughout my graduate career.
ACKNOWLEDGEMENTS

My research in the last 6 years has been a most fruitful and enjoyable experience. This dissertation is an assembly of time, work, and effort of many people who contributed vastly to my scientific education. It would not have been possible without my extraordinary mentor, Yuxin Mao, who was generous with his time, constructive feedback, and guidance. I feel extremely privileged to have the opportunity to work with a mentor who fostered a learning environment in the lab on a daily basis and as a result, helped me to become a more well-rounded scientist. I am also deeply appreciative of his encouraging words during difficult times and of his unrelenting support for my pursuit in the scientific field.

One of the highlights during my PhD career was having the privilege to meet many mentors. Among whom are my committee members, Scott Emr, Jerry Feigenson, and Rick Cerione. I am grateful for their advice at our annual Monday seminars and committee meetings which always felt natural, honest, and encouraging.

It was a great honor to conduct scientific experiments at Weill Institute for Cell and Molecular Biology, a state-of-the-art research facility that opened in 2008. The architecture and culture of the building enhanced collaboration with other labs, contributing to a rich and lively graduate experience. This could not have been possible without the support of all the faculty members, especially Scott Emr, Tony Bretcher, Chris Fromme, Fenghua Hu, Marcus Smolka, and Haiyuan Yu. I am also grateful for the time that was shared with all past and current members in the labs. I want to thank them all for their valuable friendship.

Outside of lab, I am privileged to live with a group of talented, smart, and caring individuals that is composed of undergraduates, graduates, Cornell staff, and professors. Life in Hans Bethe House for the last two years of my graduate career has been a mind-opening, horizon-broadening, and character-deepening experience. I want to thank all of the past and current Graduate Resident Fellows and House Fellows who I have worked with, especially the House Assistant Dean, Erica Ostermann, and House Professor, Scott MacDonald, for giving me the
opportunity to be part of such a meaningful community at Cornell.

Finally, I would like to express my heartfelt thanks to the department of Biochemistry, Molecular and Cell Biology (BMCB), for accepting me to the program and for providing me with the graduate assistantship which allowed me to thrive during my PhD career.
TABLE OF CONTENTS

LIST OF FIGURES ................................................................. xi
LIST OF TABLES .................................................................... xiii
LIST OF ABBREVIATIONS ........................................................ xiv

PART I: Phosphoinositide phosphatases in eukaryotes

Chapter 1: The mammalian phosphoinositide phosphatases:

Insights into substrate specificity and catalysis ........................................... 1
1.1 Introduction ........................................................................... 1
1.2 5-Phosphatases ..................................................................... 2
  1.2.1 The PI-5-phosphatase family ........................................... 2
  1.2.2 Crystal structure of the PI 5-phosphatase catalytic domain .......... 6
  1.2.3 Membrane binding .......................................................... 7
  1.2.4 Substrate specificity ........................................................ 7
  1.2.5 Catalytic mechanism ....................................................... 8
  1.2.6 Human disease mutations associated with the PI-5-phosphatase domain ... 9
1.3 The Sac1 family of phosphatases ........................................... 10
  1.3.1 The Sac1 family ............................................................. 10
  1.3.2 Crystal structure of Sac1 ................................................... 11
1.4 PTEN ................................................................................ 13
  1.4.1 The PTEN phosphatase family ....................................... 13
  1.4.2 Crystal structure of PTEN ............................................... 14
  1.4.3 Substrate specificity and membrane binding ......................... 15
1.5 Myotubularins ................................................................... 16
  1.5.1 The myotubularin family of phosphatases .......................... 16
  1.5.2 Crystal structure of myotubularin phosphatases .................. 18
  1.5.3 Substrate specificity and catalytic mechanism ...................... 18
1.6 4-phosphatases .................................................................. 20
1.7 Discussion ......................................................................... 21

Chapter 2: The Sac domain-containing phosphoinositide phosphatases ............... 23
2.1 Introduction ........................................................................ 23
2.2 Sac1 .................................................................................. 24
  2.2.1 Yeast and Mammalian Sac1 ............................................. 24
  2.2.2 Sac1 in other species ...................................................... 25
2.3 Sac2/INPP5f ...................................................................... 26
2.4 Sac3/Fig4p ......................................................................... 26
2.4.1 Yeast Sac3/Fig4p ................................................................. 26
2.4.2 Mammalian Sac3/Fig4p .................................................... 26
2.4.3 Sac3/Fig4p disease mutations ............................................. 27
2.5 Synaptojanins ................................................................. 28
  2.5.1 Enzyme activity and domain organization ......................... 28
  2.5.2 Yeast synaptojanins function and regulation ....................... 29
  2.5.3 Mammalian synaptojanins function and regulation ............... 30
  2.5.4 Synaptojanins in other species ........................................ 30
2.6 Plant Sac phosphatase family ........................................... 31
2.7 Discussion .................................................................. 33

Chapter 3: Spatiotemporal control of phosphatidylinositol 4-phosphate by Sac2 regulates endocytic recycling ................................................................. 36
3.1 Introduction ................................................................ 36
3.2 Results ..................................................................... 37
  3.2.1 Sac2 is highly expressed in the brain ......................... 37
  3.2.2 Sac2 is a phosphoinositide 4-phosphatase ............... 39
  3.2.3 The catalytically inactive mutant of Sac2 is localized to punctuate structures ............................................. 39
  3.2.4 Sac2 is localized to early and recycling endosomal structures ................................................................. 41
  3.2.5 Sac2CS mutant alters the cellular distribution of TfnR and perturbs its recycling ............................................. 44
  3.2.6 CRISPR-mediated knock-out of Sac2 delays TfnR recycling ................................................................. 44
  3.2.7 CRISPR-mediated knock-out of Sac2 perturbs integrin distribution and recycling ................................................................. 46
  3.2.8 Tfn-containing compartment is positive for PI(4)P in Sac2 null cells ................................................................. 48
  3.2.9 Sac2 null cells are defective in cell migration ................................................................. 49
  3.2.10 Structure of the homology Sac2 domain (hSac2) ................................................................. 49
3.3 Discussion .................................................................. 53
3.4 Materials and Methods .................................................. 56
  3.4.1 Cloning, mutagenesis, and plasmids ................................................................. 56
  3.4.2 Cell culture and transfection ................................................................. 56
  3.4.3 CRISPR .................................................................. 56
  3.4.4 Immunofluorescence ................................................................. 57
  3.4.5 Quantitative analysis of colocalization and total cell fluorescence ................................................................. 57
  3.4.6 Time-lapse microscopy ................................................................. 57
  3.4.7 Immunoprecipitation ................................................................. 58
PART II: PI-phosphatase and PI-binding effectors encoded by *Legionella pneumophila*

Chapter 4: Structural basis for substrate recognition

by a unique *Legionella* phosphoinositide phosphatase

4.1 Introduction .................................................62
4.2 Results .........................................................64
  4.2.1 *Legionella* effector SidF is a phosphoinositide 3-phosphatase ...........................64
  4.2.2 SidF facilities the anchoring of effector proteins to bacterial phagosomes ...............65
  4.2.3 Crystal structure of SidF ..................................67
  4.2.4 Structure of SidF-PI(3,4)P2 complex ................................70
  4.2.5 Structure comparison of SidF to other phosphatases ......................................73

4.3 Discussion ....................................................75

4.4 Materials and Methods ........................................76
  4.4.1 Cloning and mutagenesis ................................76
  4.4.2 Protein expression and purification ........................................78
  4.4.3 Bioinformatic search ........................................78
  4.4.4 Crystallization ..............................................78
  4.4.5 Data collection and processing ................................79
  4.4.6 Structure determination and refinement ..................................79
  4.4.7 Enzymatic assays ...........................................79
  4.4.8 Thin-layered chromatography ................................80
  4.4.9 Cell culture and transfection and fluorescence microscopy ...............................80
  4.4.10 Bacterial infection ..........................................81
  4.4.11 Immunostaining ............................................81

Chapter 5: The *Legionella* effector SidC defines a unique family

of ubiquitin ligases important for bacterial phagosomal remodeling ................................ 82

5.1 Introduction ..................................................82
5.2 Results ........................................................85
  5.2.1 Crystal structure of the N-terminal domain of SidC
                          reveals a canonical catalytic triad ................................................85
LIST OF FIGURES

PART I

Fig. 1. Schematic domain organization of 5-phosphatases. ................................................. 5
Fig. 2. 5-phosphatase structure and catalytic mechanism. ................................................ 6
Fig. 3. Schematic domain organization of the phosphatases in the Sac1 family. ............ 10
Fig. 4. Representative structures of CX5R motif-containing PI phosphatases. .............. 12
Fig. 5. Schematic domain organization of the phosphatases in the PTEN family. ......... 15
Fig. 6. Schematic domain organization of the phosphatases in the myotubularin family. 17
Fig. 7. Crystal structure of MTMR2, a representative member of the myotubularin family. 19
Fig. 8. Crystal structure of the bacterial CX5R motif-based PI phosphatase SidF. ....... 21
Fig. 9. Sac domain-containing phosphatase family. ........................................................... 25
Fig. 10. Disease mutations mapping
don the modeled structure of the Sac domain of human Sac3/Fig4. .......................... 29
Fig. 11. Sac domain structure of yeast Sac1. ................................................................. 34
Fig. 12. Sac2 is a PI 4-phosphatase that specifically hydrolyzes PI(4)P. ......................... 38
Fig. 13. Sac2 in vitro phosphatase assay. ................................................................. 40
Fig. 14. Sac2CS mutant localizes to punctate structures. ............................................. 41
Fig. 15. Sac2 localizes to early and recycling endosomes. ............................................ 42
Fig. 16. Colocalization of Sac2CS with different endosomal
and organelle markers in N2A and HeLa cells ............................................................. 43
Fig. 17. Expression of Sac2CS mutant perturbs TfnR distribution and Tfn recycling. .... 44
Fig. 18. Defects of Tfn recycling in Sac2 null cells. .................................................... 45
Fig. 19. Defects of integrin recycling in Sac2 null cells. .............................................. 47
Fig. 20. The cellular distribution of CI-M6PR is unaffected in Sac2 null cells. ............. 48
Fig. 21. Tfn-containing vesicles are positive for PI(4)P in Sac2 null cells. .................... 50
Fig. 22. Sac2 deletion delays cell migration. ................................................................. 51
Fig. 23. Crystal structure of the hSac2 domain of Sac2. ............................................. 52
Fig. 24. Model for Sac2 function during endosomal recycling. ....................................... 54
PART II

Fig. 1. *Legionella* effector SidF is a phosphoinositide phosphatase. ................................. 66
Fig. 2. Intracellular localization of transfected SidF in mammalian cells. ......................... 67
Fig. 3. SidF is required for anchoring SidC to the bacterial phagosomes. ......................... 68
Fig. 4. Multiple sequence alignment of SidF ......................................................... 69
Fig. 5. Crystal structure of SidF ........................................................................ 71
Fig. 6. Substrate recognition by SidF ...................................................................... 72
Fig. 7. Two-dimension depiction of SidF–diC4-PI(3,4)P2 complex ................................. 73
Fig. 8. Structure comparison of SidF with other phosphoinositide phosphatases .......... 74
Fig. 9. Functional model of SidF ........................................................................ 77
Fig. 10. Crystal structure of the N-terminal SNL domain of SidC (amino acids 1–542) .. 85
Fig. 11. Ectopic expression of SidC altered the intracellular ubiquitination pattern ....... 88
Fig. 12. The SNL domain of SidC has E3 ubiquitin ligase activity .............................. 90
Fig. 13. Ubiquitin ligase activity assay of the SNL domain and the full length SidC ....... 91
Fig. 14. In vitro ubiquitination assays with full length SdcA (1-908) .............................. 92
Fig. 15. The ubiquitin linkage preference by the SNL domain of SidC ...................... 93
Fig. 16. The E3 ubiquitin ligase activity is required for the ER marker recruitment
to the bacterial phagosome .................................................................................. 94
Fig. 17. The E3 ubiquitin ligase activity is required for the recruitment of ubiquitin conjugates
to the bacterial phagosome .................................................................................. 95
LIST OF TABLES

PART I
Table 1. Sac domain enzymatic activities and associated phenotypes. .................................. 35
Table 2. hSac2: Data collection and structural refinement statistics. ................................. 52

PART II
Table 1. SidF: Data collection, phasing, and structural refinement statistics. ....................... 70
Table 2. SidC: Data collection and structural refinement statistics. ................................. 87
Table 3. SILAC hits of proteins with increased ubiquitination. ......................................... 89
LIST OF ABBREVIATIONS

Sac: Suppressor of yeast actin mutations
TGN: Trans-golgi network
CCPs: Clathrin-coated pits
CCV: Clathrin-coated vesicle
ERC: Endocytic recycling compartment
FAPP1: Four-phosphate adaptor protein
hSac2: homology Sac2 domain
PCC: Pearson's correlation coefficient
PH: Pleckstrin homology
PI(4)P: phosphatidylinositol 4-phosphate
PI: Phosphoinositide
CI-M6PR: Cation-Independent Mannose-6-phosphate receptor
Tfn: Transferrin
TfnR: Transferrin receptor
EGF: Epidermal growth factor
EGFR: Epidermal growth factor receptor
PAO: Phenylarsine oxide
BFA: Brefeldin A
LCV: Legionella-containing vacuole
SYNJ: Synaptojanin
MTM: Myotubularin
PTEN: phosphatase and tension homolog located on chromosome TEN
OCRL: Oculo-cerebro-renal syndrome
CMT: Charcot-Marie Tooth disorder
ALS: Amyotrophic lateral sclerosis
PRD: Proline-rich domain
MIM: Membrane interaction motif
CRISPR: Clustered regularly interspaced short palindromic repeats
SNL: SidC N-terminal ubiquitin ligase domain
SILAC: Stable Isotope Labeling by Amino acids in Cell culture
Ub: Ubiquitin
PART I: Phosphoinositide phosphatases in eukaryotes

Chapter 1: The mammalian phosphoinositide phosphatases: Insights into substrate specificity and catalysis

Abstract: Phosphoinositides (PIs) are specialized lipids that play many essential biological processes in eukaryotes. PI phosphatases, together with PI kinases, are responsible for the conversion of PIs between distinctive phosphorylation states. PI phosphatases are a large collection of enzymes that are evolved from at least two disparate ancestors. One group is distantly related to endonucleases, which applies divalent metal ions for phosphoryl transfer. The other group is related to protein tyrosine phosphatases, which contains a highly conserved active site motif Cys-X$_5$-Arg (CX$_5$R). In this chapter, we focus on structural insights to illustrate current understanding of the molecular mechanisms of each PI phosphatase family, with emphasis on their structural basis for substrate specificity determinants and catalytic mechanisms.

1.1 Introduction

Phosphoinositides (PIs) are pivotal cellular regulators and play essential roles in a broad spectrum of cellular processes including defining intracellular organelle identity, cell signaling, proliferation, cytoskeleton organization, and membrane trafficking [1-4]. PIs are a collection of seven lipid derivatives that differ with regard to the presence or absence of phosphate groups at the 3', 4', and 5' positions. The reversible phosphorylation of PIs is under tight spatial temporal control by a large number of PI kinases and phosphatases that are present throughout eukaryotic species.

In the past thirty years, significant advances have been made in the PI phosphatase field. With the completion of human genome and whole genome sequencing of many other species, genes that encode PI phosphatases have been well documented. Based on catalytic mechanisms,
PI phosphatases can be categorized into two major classes.

One class is the inositol polyphosphate 5-phosphatases, which are Mg\(^{2+}\)-dependent phosphatases comprising of 10 mammalian [5] and 4 yeast members [6]. 5-phosphatases have a characteristic central catalytic domain that exhibits sequence homology to the apurinic/apyrimidinic family of endonucleases. Members of this class hydrolyze the D5 phosphate of the inositol ring of both soluble inositol polyphosphates and membrane-bound phosphotidylinositides [7].

The other class is characterized by the presence of a highly conserved CX\(_5\)R active site motif, of which the cysteine residue functions as the nucleophile while the arginine residue positions the scissile phosphate group. Based on their primary sequences, enzymes in this class can be further divided into four sub-families that include the Sac1 domain containing phosphatases, PTEN, myotubularins, and 4-phosphatases. In addition to the eukaryotic PI phosphatases, several bacterial pathogens also encode CX\(_5\)R-based PI phosphatases, which will be discussed in detail in Part II.

This chapter highlights structural insights into each family of PI phosphatases to illustrate the mechanism of substrate specificity, membrane targeting, and catalysis.

1.2 5-phosphatases

1.2.1 The PI-5-phosphatase family

The inositol polyphosphate 5-phosphatases are a group of Mg\(^{2+}\)-dependent phosphatases first characterized in the studies on the metabolism of the inositol phosphate second messenger Ins(1,4,5)P\(_3\) [8, 9]. The first 5-phosphatase gene cloned in the early 90’s [10, 11] is named INPP5A or 5-phosphatase type I. It has a molecular weight of 43-kD and hydrolyses only the soluble inositol phosphates Ins(1,4,5)P\(_3\) and Ins(1,3,4,5)P\(_4\). Following these earlier studies, other 5-phosphatases were discovered. For example, the 75-kD INPP5B purified from human platelet was shown to have activity against PI(4,5)P\(_2\), and PI(3,4,5)P\(_3\) besides Ins(1,4,5)P\(_3\) and Ins(1,3,4,5)P\(_4\) [12, 13]. Subsequently, the human cDNA encoding INPP5B was cloned [14, 15]. Approximately the same time, the gene responsible for the human X-linked disease Oculo-
Cerebro-Renal Syndrome of Lowe (OCRL) was mapped [8, 9] and shown to be highly homologous to INPP5B [16]. Like INPP5B, OCRL is a multidomain-containing phosphatase with a PH domain at the N-terminus [17] followed by a central 5-phosphatase catalytic domain, an ASH domain [18], and a C-terminal RhoGAP-like domain. Unlike INPP5B, which has a CAAX motif at its C-terminal [14], OCRL has two clathrin-binding motifs within its N-terminal PH domain and C-terminal RhoGAP-like domain, respectively (Fig. 1) [17, 19]. OCRL hydrolyzes PI(4,5)P₂, PI(3,4,5)P₃, and inositol polyphosphate Ins(1,4,5)P₃ and Ins(1,3,4,5)P₄, with a preference for PI(4,5)P₂ [20, 21].

The repertoire of 5-phosphatases continued to grow when synaptojanin 1 (SYNJ1) was discovered to be responsible for the hydrolysis of PI(4,5)P₂ on synaptic vesicles at the nerve terminal [22]. SYNJ1 deletion mice develop neurological defects with accumulated clathrin-coated vesicles in nerve endings [23]. These findings provided strong evidence for the critical role of PIs in membrane trafficking in addition to their classical roles as second messengers in signal transduction [24]. Along this line, a ubiquitously expressed SYNJ1 homologue, SYNJ2 was identified [25]. Besides the 5-phosphatase domain, both of these two phosphatases share an N-terminal Sac1 domain, which was later found to be a member of a new family of PI phosphatases [26]. Human SYNJ1 and SYNJ2, but not yeast synaptojanin homologues, also contain an RNA binding domain (protein data bank ID: 2DNR; 1UFW) after their central 5-phosphatase domain (Fig. 1). The large C-terminal portion in both SYNJs is a divergent proline-rich region (PRD), which contains a string of peptide motifs that mediate bindings with a variety of endocytic proteins such as amphiphysin and endophilin [27, 28]. Biochemical characterization showed that the N-terminal Sac1 domain of synaptojanins predominantly dephosphorylates PI(3)P and PI(4)P, whereas the 5-phosphatase domain dephosphorylates PI(4,5)P₂ and PI(3,4,5)P₃ at the 5’ position of the inositol ring [26, 29]. Abnormalities of these two phosphatases have been linked to several neuronal diseases. For example, overexpression of SYNJ1 resulting from trisomic duplication may contribute to brain dysfunction in mouse models of Down’s syndrome [30, 31] and bipolar disorder [32, 33]. In addition, a mutation in the Sac1 domain is linked with early-onset progressive Parkinson-
ism [34]. In a mouse mutant strain called Mozart, a mutation in SYNJ2 is implicated in progressive hearing loss [35].

The fact that synaptojanins possess both the Sac1 and the 5-phosphatase domains leads to a fascinating speculation that the two catalytic domains may function in a concerted way. In the concerted model, the 5-phosphatase domain converts PI(4,5)P₂ to PI(4)P, which is directly “channeled” to the Sac1 domain for further dephosphorylation to phosphatidylinositol without accumulation of the intermediate en route. This may permit the rapid inactivation of PI(4,5)P₂ following endocytosis, thereby avoiding mis-targeting of endocytic transport factors to intracellular membranes.

The other 5-phosphatase family member SHIP1 plays an important role in immune system. It was first identified as a 145-kD tyrosine phosphorylated protein from haematopoietic cells upon cytokine stimulation [36]. SHIP1 was then cloned and shown to hydrolyze PI(3,4,5)P₃ and Ins(1,3,4,5)P₄ in vitro [37-39]. The close homologue of SHIP1, SHIP2, plays a role in the negative regulation of insulin signaling [40] and serves as a potential drug target for obesity. Interestingly, SHIP2 prefers to hydrolyze Ins(1,3,4,5)P₄, PI(3,5)P₂, and PI(3,4,5)P₃, but not PI(4,5)P₂ [41]. Both SHIP1 and SHIP2 have an SH2 domain at their N-terminus, a central 5-phosphatase catalytic domain, and a C-terminal PRD. SHIP2, but not SHIP1, also has a SAM (sterile alpha motif) domain at its C-terminus (Fig. 1).

Other 5-phosphatases, including PIPP (INPP5J), SKIP (INPP5K), and Pharbin (INPP5E), are less well studied. INPP5E has an N-terminal PRD, a central 5-phosphatase domain, and a C-terminal CAAX motif and has been shown to preferentially hydrolyze PI(3,4,5)P₃ and PI(4,5)P₂ [42, 43]. INPP5E is linked to PI signaling in the primary cilium and mutations in INPP5E cause cillum signaling defects, which lead to a human genetic disease, called Joubert syndrome [44-46]. SKIP (INPP5K) is a skeletal muscle and kidney enriched 5-phosphatase, which contains an N-terminal 5-phosphatase domain and a unique SKICH domain at the C-terminus [47, 48]. SKIP decreases PI(3,4,5)P₃ levels upon insulin stimulation and SKIP knockout mice show increased insulin sensitivity [49]. The last human 5-phosphatase is PIPP, which has two PRD domains at its
N- and C-terminus and a central catalytic domain followed by a SKICH domain [50]. PIPP can hydrolyze Ins(1,4,5)P$_3$, Ins(1,3,4,5)P$_4$, PI(4,5)P$_2$, and PI(3,4,5)P$_3$ in vitro [50]. However, it is shown to be a novel negative regulator of PI3-kinase pathway by hydrolyzing PI(3,4,5)P$_3$ in vivo [51].

![Fig. 1. Schematic domain organization of 5-phosphatases. Abbreviations: 5-ptase, 5-phosphatase catalytic domain; PH, pleckstrin homology; ASH, ASPM-SPD2-Hydin; CB, clathrin binding; RBD, RNA binding domain; SH2, Src homology 2; SKICH, SKIP carboxyl homology.](image)

With the completion of genome projects of human and of other organisms, ten 5-phosphatases have been identified in mammals and five members in yeast [52]. In addition to their catalytic domain of approximately 300 amino acids, all 5-phosphatases contain a variety of domains or motifs that determine their cellular localization and/or regulate their activity (Fig. 1). Crystal structures of 5-phosphatase catalytic domains from several members have been determined [53-55]. These structures provide insights into the molecular basis for catalysis, substrate recognition, membrane interaction, as well as human genetic diseases.
1.2.2 Crystal structure of the PI-5-phosphatase catalytic domain

The first crystal structure of the 5-phosphatase domain was determined from Schizosaccharomyces pombe synaptojanin [55]. The structure reveals that the core of the 5-phosphatase domain consists of two layered β-sheets with one containing 5 β-strands and 6 for the other. These two sheets stack on each other in a parallel way forming a stable core by burying a large number of hydrophobic residues in between. The α-helices form two α-helical layers on each side of the central β-sheets core. The catalytic site created by loops and several α-helices lies at the top edge.

Fig. 2. 5-phosphatase structure and catalytic mechanism. (A) Ribbon diagram of the catalytic domain of INPP5B with bound diC8 PI(4)P (in sticks) and a Mg\(^{2+}\) ion (in green sphere) (PDB ID: 3MTC). (B) Zoom-in view of the membrane insertion motif (MIM) of INPP5B. MIM is colored in gold and enriched with hydrophobic and cationic residues. (C) Zoom-in view of the interactions between INPP5B with the head-group of PI(4)P. Residues Y502, K503, K516, and R518 form a cationic pocket P4 interacting motif (P4IM) for the recognition of the D4 phosphate group. The D1 phosphate group is bonded by N379 and K380. The Mg\(^{2+}\) ion is coordinated by N275 and E303. Residues directly involved in hydrolysis (D447, N449, D524, and H549) are labeled and shown in sticks. (D) Schematic model of 5-phosphatase catalytic mechanism. The reaction starts with the activation of a water molecule by the general base D447. The nucleophilic attack by this water molecule results in a pentavalent transition state of the scissile phosphate. The negative charge on the scissile phosphate is stabilized by H549 and the Mg\(^{2+}\) ion. Finally, the 5’ hydroxyl group of the lipid is protonated by a proton relayed from D447.
of the two layered β-sheets (Fig. 2A). Although it has no detectable sequence homology, 5-phosphatase shows an extensive structure similarity to endonucleases, such as the well-studied apurinic/apyrimidic (AP) endonuclease APE1 [56]. Structural comparison with APE1, together with recent crystal structures of 5-phosphatase/lipid complexes [54], has shed light on the molecular basis for the catalytic function of 5-phosphatases.

1.2.3 Membrane binding

Most of the 5-phosphatases prefer to act on membrane-embedded PI substrates, thus, hydrophobic membrane interacting motifs (MIM) on the catalytic domain are often required to mediate the interaction between the protein and membrane bilayer. The 5-phosphatase domain of INPP5B in complex with diC8 PI(4)P reveals exemplary structural features for membrane docking (Fig. 2B) [54]. In the complex structure, the aliphatic chains of diC8 PI(4)P pack against two α-helices enriched with hydrophobic residues (colored gold in Fig. 2B). These hydrophobic residues are likely to penetrate into the hydrophobic lipid bilayer to access substrates. In addition to hydrophobic residues, several cationic residues, including His314, Lys308, Arg376, and Arg410, may contribute to the electrostatic interactions with negatively charged lipid head groups. Therefore, both hydrophobic and electrostatic interactions may facilitate the anchoring for the 5-phosphatase domain on membrane interface and may enhance the processivity of PI hydrolysis. It is interesting to note that although the primary sequence of these two α-helices is variable in other 5-phosphatases, the hydrophobic nature of these residues is conserved. The differences in the membrane binding region imply a differential preference for membrane with unique physical properties. This may explain the observed enzyme sensitivities to fatty acid composition and membrane curvature [20, 57].

1.2.4 Substrate specificity

Most of the 5-phosphatases hydrolyze the D5 phosphate from either PI(4,5)P2 or PI(3,4,5)P3. This preference suggests that the D4 phosphate is the key group for enzyme recognition.
The complex structure of the diC8 PI(4)P bound INPP5B catalytic domain reveals four highly conserved residues within a long loop, forming a so-called P4IM (P4-interacting-motif) to accommodate the D4 phosphate (Fig. 2C). Thus, these four residues function as the key specificity determinants for D4-phosphate binding. As a deviation from this basic structural feature for substrate selectivity, the P4IM loop in SHIP2 and SHIP1 is seven residues longer than in INPP5B. Upon binding of a competitive inhibitor, the P4IM loop folds over the inhibitor and an arginine residue shifts to a position allowing direct binding with the D3 phosphate on the inositol ring [53]. This new structural feature in SHIP2 and SHIP1 may explain the preferred specificity of these two enzymes towards PI(3,4,5)P3 vs. PI(4,5)P2. The complex structure of INPP5B with diC8 PI(4)P also highlights the binding site for D1 phosphate group, which is recognized by Asn379 and Lys380 through direct hydrogen bonding. Specific recognition by the P4IM and the D1 phosphate binding site aligns the substrate in a way that allows the D5 phosphate to be positioned to the catalytic site (Fig. 2C).

1.2.5 Catalytic mechanism

Structural similarity between 5-phosphatases and endonucleases and their similar chemical nature of the scissile bond suggest a conserved catalytic mechanism between these two families of enzymes [7, 55]. Recent comprehensive studies combining structural, biochemical, and computational analyses have led to a unified mechanism for AP nucleases [58]. A similar mechanism can be applied to PI-5-phosphatases. Key residues, as well as a magnesium ion, required for phosphoryl cleavage have been identified based on the diC8 PI(4)P bound INPP5B catalytic domain structure [54] (Fig. 2D). Invariable residues N449 and D447 coordinate and activate a water molecule with D447 acting as a general base. The activated water molecule attacks the D5 phosphate to form a pentavalent phosphate intermediate. The highly negatively charged intermediate is stabilized by the Mg\(^{2+}\), which is coordinated by invariable residues E303 and N275, as well as several first hydration shell water molecules. The protonation of the departing PI product is not well understood, however, one possible mechanism is that the cleaved phosphate group may relay
the proton from the protonated D524 (Fig. 2D).

1.2.6 Human disease mutations associated with the PI-5-phosphatase domain

Structural studies on these enzymes have shed light on many missense mutations in 5-phosphatases that are associated with several human hereditary diseases [59]. Human genetic studies have documented a large number of missense mutations within the 5-phosphatase domain of OCRL. These are responsible for the Lowe syndrome, as well as the Dent 2 disease, which has a less severe symptom compared to the Lowe syndrome [60]. Most of the missense mutations found in Lowe syndrome patients are clustered either in the hydrophobic core of the catalytic domain or at the catalytic site. These mutations may cause the instability of the protein or directly impair the enzymatic activity. Mutations found in the Dent 2 patients tend to be on the surface around the catalytic site, which may have a less impact on the enzymatic activity [59].

1.3 The Sac1 family of phosphatases

1.3.1 The Sac1 family

The founding member of this family Sac1 was originally identified in yeast by two independent genetic screens searching for modifiers of actin cytoskeleton defects and of trans-Golgi network exocytic failure caused by inactivation of Sec14p, respectively [61-63]. Subsequently, the Sac1 domain was found to share considerable homology with mammalian synaptojanin [22]. The Sac1 domain comprises of approximately 500 amino acids with a highly conserved CX_{5}R motif [64] and was demonstrated to possess a PI phosphatase activity [26]. With the completed genomes of several species, all genes that contain the Sac1 domain have been identified. There are five Sac1 domain-containing proteins in both human and yeast (Fig. 3).

Sac1 is a 67kD type II membrane protein that localizes to both ER and Golgi apparatus [65, 66]. In yeast, loss of function of Sac1 causes a broad range of cellular defects [6] such as disorganization of the actin cytoskeleton [63], inositol auxotrophy [66], cold sensitivity for growth
[63], multiple drug sensitivities [67], vacuolar function [68, 69], cell wall maintenance [70], and ATP uptake into the ER [71]. Sac1 mutants in Drosophila die as embryos with defects in dorsal closure [72]. Mouse strains deficient in Sac1 are embryonic lethal [73, 74]. Knock-down of Sac1 expression in mammalian cell lines results in disorganization of Golgi membranes and mitotic spindles [73, 74]. These findings suggest an essential role of Sac1 in multicellular organisms. In vitro studies have demonstrated that Sac1 dephosphorylates a number of PIs, including PI(3)P, PI(4)P, and PI(3,5)P2 [26]. In vivo, genetic ablation of Sac1 activity in yeast results in a nearly 10-fold increase in the steady state levels of PI(4)P with little effect on PI(4,5)P2 [26, 75], suggesting that yeast Sac1p is a major enzyme for PI(4)P degradation in vivo.

![Fig. 3. Schematic domain organization of the phosphatases in the Sac1 family. Abbreviations: TM, transmembrane motif; hSac2, homology of Sac2.](image)

Sac2 is also named INPP5F. In addition to its N-terminal Sac1 domain, Sac2 has a unique domain named hSac2 (homology of Sac2) and a C-terminal proline-rich domain (Fig. 3). We recently determined the crystal structure of the hSac2 domain, which resembles a PH domain (see Chapter 3). Initial cloning and characterization show that this enzyme exhibits 5-phosphatase activity specific for PI(4,5)P2 and PI(3,4,5)P3 [76]. Along this line, Sac2 is shown to hydrolyze PI(3,4,5)P3 and thereby inactivates Akt signaling, which eventually leads to the attenuation of heart hypotrophy [77]. Moreover, Sac2 deficient mice display abnormal fetal gene reactivation and increased susceptibility to stress induced heart hypertrophy. On the other hand,
Sac2 transgenic mice prevent the animal from developing these symptoms [78]. However, our recent data demonstrated that Sac2 is a PI-4-phosphatase that specifically hydrolyzes PI(4)P in vitro and plays a role in the endocytic recycling pathway (see Chapter 3).

Sac3, also known as Fig4p, was originally identified in a screen for genes induced by mating pheromone in *Saccharomyces cerevisiae* [79]. Sac3 is a phosphoinositide 5-phosphatase that specifically hydrolyzes PI(3,5)P$_2$ to generate PI(3)P both in vitro and in vivo [80-82]. Interestingly, Sac3/Fig4p forms a complex with a PI(3)P-5-kinase Fab1p and a scaffold protein Vac14p. This complex is also conserved in mammals and is responsible for the acute regulation of subcellular levels of PI(3,5)P$_2$ [83-85]. Genetic mutations affecting the function of this complex lead to a number of diseases, including Charcot-Marie-Tooth (CMT) type 4J, a subset of amyotrophic lateral sclerosis (ALS) in human, and neurodegeneration in the pale tremor mouse [86-88].

### 1.3.2 Crystal structure of Sac1

The Sac1 domain of yeast Sac1 is the only crystal structure solved in this family to this day [89]. The structure reveals that the Sac1 domain is composed of two closely packed sub-domains: a novel N-terminal sub-domain and the catalytic sub-domain. The N-domain has a unique fold comprised of three layers of β-sheets, and one long and three short α-helices. The function of the N-domain is still unclear, but it may mediate protein-protein interactions that may direct Sac1-mediated dephosphorylation of PI(4)P on specific membrane compartment [90].

The catalytic domain consists of a nine-stranded β-sheet flanked by five α-helices with two and three on each side of the β-sheet (Fig. 4A). Conserved among all CX$_5$R motif-based phosphatases, Sac1 also has a central structural core formed by four parallel β-strands and one α-helix. The loop (P-loop) connecting this α-helix and one of the β-strands harbors the catalytic CX$_5$R motif (Fig. 4B). The other conserved structural feature among all CX$_5$R motif-based phosphatases is the overall surface charge distribution. The surface hosting the catalytic CX$_5$R motif is largely positively charged (Fig. 4C). On this surface, several flexible loops enriched with hydrophobic residues surround the catalytic site and likely facilitate the association of the enzyme with the
anionic membrane bilayer.

Compared with the structure of PTEN [91] or myotubularins [92, 93], one difference is evident. The catalytic P-loop of Sac1 assumes a unique conformation in that the catalytic cysteine (C392) is oriented away from the conserved R398 in the CX5R motif. This difference suggests that a conformational change of the catalytic P-loop may be required to achieve full enzymatic activity (Figs. 4B and E). In agreement with the structural observation, Sac1 is found to be an allosteric enzyme and its activity can be stimulated by anionic phospholipids [94]. However, comprehensive understanding of the catalytic mechanism requires co-crystal structures of Sac1 with bound substrate or activator.

**Fig. 4. Representative structures of CX5R motif-containing PI phosphatases.** (A) Ribbon diagram of the crystal structure of Sac1 (PDB ID: 3LWT). Sac1 has two tightly packed domains: the SacN (cyan) and the catalytic domain (labeled as Ptase in grey). The conserved structural core is colored in pink and the catalytic P-loop in red. (B) Zoom-in view of the structural core (in pink), which is conserved in all CX5R-based PI phosphatases. (C) Molecular surface of the catalytic site of Sac1. The surface is colored based on electrostatic potential with positively charged region in blue (+4 kcal/electron) and negatively charged surface in red (-4 kcal/electron). Sac1 is in the same orientation in (A) and (C). (D)–(F) Crystal structure of PTEN (PDB ID: 1D5R). The C2 domain is shown in yellow. (G)–(I) Crystal structure of MTMR2 (PDB ID: 1ZVR). The PH-GRAM domain is shown in green. (J)–(L) Crystal structure of SidF (PDB ID: 4FYG; discussed in Chapter 4).
Sac1 is localized on the endoplasmic reticulum (ER) and Golgi apparatus via its two C-terminal transmembrane motifs. An intriguing question remains in the field as to whether Sac1 hydrolyzes PI(4)P on the same membrane that it anchors (cis model) or the substrates are located on a different membrane compartment (trans model). Genetic data have shown that inactivation of Sac1p leads to a specific increase in the cellular level of PI(4)P and the bulk of accumulated PI(4)P is generated by the phosphoinositol-4 kinase Stt4 at the plasma membrane [68, 69]. Sac1 has been proposed to control the PI(4)P levels at the plasma membrane (PM) at the ER-PM contact sites in trans [95]. The presence of a long linker between the catalytic domain and the first transmembrane motif allows the enzyme to overcome the gap between ER and PM at the contact sites and hydrolyze substrates in trans. Recent studies also suggest that Sac1 can hydrolyze PI(4)P in cis on the ER membrane to facilitate the sterol/PI(4)P exchange between ER and Golgi apparatus [96]. The exact mode of how Sac1 functions is still under debate. More delicate experiments are needed to dissect the mechanism. However, it is possible that Sac1 can hydrolyze its substrates both in cis and in trans depending on the presence of its associated regulatory proteins. Refer to Chapter 2 for details on mammalian Sac1.

1.4 PTEN

1.4.1 The PTEN phosphatase family

PTEN (phosphatase and tension homolog located on chromosome TEN) was first identified as a tumor suppressor gene located on chromosome 10q23 [97, 98]. PTEN is one of the most frequently mutated genes found in diverse human cancers [97-101] and in hereditary cancer predisposition syndromes, such as Cowden [102, 103]. PTEN contains the CX5R phosphatase catalytic signature motif and is initially reported to function as a protein tyrosine phosphatase [100]. It was later demonstrated that the principal catalytic function of PTEN is to dephosphorylate the D3 phosphate of PI(3,4,5)P₃ [104], thereby antagonizing the PI 3-kinase signaling pathway [105]. Following its initial identification and characterization, extensive research on its physiological
roles has been done in various cell cultures and animal models [106]. Results from these studies have implicated PTEN in variety of cellular processes such as cell polarity, adhesion, migration, survival, metabolism, immune response, cell cycle progression, and DNA repair [107].

PTEN is composed of an N-terminal phosphatase domain, followed by a C2 domain, which contributes to membrane binding and phosphatase activation [108]. PTEN also has a C-terminal tail region that contains multiple phosphorylation sites and a PDZ domain-binding sequence (Fig. 5). Bioinformatic analysis revealed several distinct PTEN-related genes. The testis specific TPTE (transmembrane phosphatase with tensin homology) is localized to the PM but lacks PI phosphatase activity [109, 110]. TPIP (TPTE and PTEN homologous inositol lipid phosphatase), which comprises several splice isoforms and displays similar 3-phosphatase activity compared with PTEN [110]. Interestingly, another TPIP ortholog in the marine invertebrate, Ciona intestinale, is a voltage-sensitive phosphatase (VSP) [111]. VSP consists of a transmembrane voltage sensing domain (VSD) and a cytoplasmic PTEN-related phosphatase domain (Fig. 5), which couples the phosphatase action directly to membrane potential changes. However, unlike PTEN, VSP hydrolyzes the D5 phosphate from PI(4,5)P₂ and PI(3,4,5)P₃ [112, 113]. Structural studies of PTEN [91] and several VSPs [114, 115] have provided significant insights into their substrate specificity and mutations in the context of various human diseases.

1.4.2 Crystal structure of PTEN

The crystal structure of PTEN was solved with the phosphatase domain and the C2 domain in tandem [91]. The phosphatase assembles in an overall fold similar to protein tyrosine phosphatases with a central five-stranded β-sheet sandwiched by 6 α-helices (Fig. 4D). Compared with other proteins or PI phosphatases, the core of the phosphatase domain, which contains four parallel β-strands and one α-helix with the catalytic P-loop connecting one of the β-strands and the α-helix, is conserved (Fig. 4E). The catalytic P-loop containing the CX₅S motif resides in a deep and wide pocket at the center of the phosphatase domain. This pocket is surrounded by several functional loops including a so-called “TI” loop (named after the conserved the threonine
and isoleucine residues) and a “WPD” loop, where the Asp residue functions as both a general acid and base during catalysis. The catalytic pocket is highly positively charged (Fig. 4F), which is a critical determinant for the binding of heavily negatively charged substrates. The structure also reveals the presence of a C2 domain. The C2 domain consists of a core β-sandwich made up of two antiparallel β-sheets, and two short α-helices near the interface. The phosphatase domain and the C2 domain are tightly packed against each other sharing an extensive interface enriched with hydrophobic and aromatic residues, and an interdomain hydrogen-bonding network. Mutations that affect this interface are frequently observed in cancer [116].

1.4.3 Substrate specificity and membrane binding

The structure of PTEN, as well as recently determined VSP structures, unraveled several structural determinants for the remarkable substrate specificity of PTEN family phosphatases. PTEN prefers to hydrolyze the D3 phosphate of PI(3,4,5)P₃. To accommodate the bulky head-group of PI(3,4,5)P₃, PTEN has evolved to possess a wide and deep catalytic pocket. This enlarged pocket is created by a conserved 4-residue insertion within the “TI” loop. Moreover, the catalytic site of PTEN is more positively charged than that of Sac1 or MTM as demonstrated by the calculated electrostatic potentials (Fig. 4F), which is an agreement with its preference for PI(3,4,5)P₃, the mostly negatively charged PIs species. Contrary to PTEN, the other family member, VSP prefers to hydrolyze the D5 phosphate on PI(4,5)P₂ and PI(3,4,5)P₃ [112, 113]. This discrepancy is at least partially explained by a residue substitution within the catalytic CX₅R motif.
(a G365 in VSP, while A126 in PTEN) since the G365A mutation abolishes the activity against PI(4,5)P$_2$ [112]. Structures of the PTEN family enzymes also revealed the role of the C2 domain in membrane association. The C2 domain of PTEN has two protruding solvent exposed loops clustered with 9 basic and two hydrophobic residues on the same face as and in close proximity to the active site. These two loops are likely inserted into the lipid bilayer, thereby, contributing to the activation of the phosphatase domain [91].

In spite of the intense biochemical and structural studies of PTEN family phosphatases, the exact molecular mechanisms for substrate determination, particularly the distinct preference for the scissile phosphate group by PTEN and VSP, remain to be established. Future work with complex structures of enzymes with their cognate lipid substrates or products is necessary to answer these intriguing questions.

1.5 Myotubularins

1.5.1 The myotubularin family of phosphatases

The myotubularin family consists of 15 members that are named MTM1 and MTMRs 1-14. Myotubularins are only found in eukaryotes and are highly conserved from yeast to human. The first member of this family, $MTM1$, was identified as gene mutated in patients with a congenital muscle disorder called X-linked myotubular myopathy [117]. The protein sequence of MTM1 shares similarity with protein tyrosine phosphatases (PTPs), but exhibits poor activity towards phosphoprotein substrates [118]. The enzymatic activity of MTM1 was further characterized by two independent studies which demonstrated that MTM1 is a PI-3-phosphatase that hydrolyzes PI(3)P and PI(3,5)P$_2$ [119, 120]. With complete genome sequencing, 14 myotubularin-related (MTMR) genes were identified to date. MTM1 and MTMRs constitute the largest family of CX$_5$R dependent PI phosphatases.

Members of the myotubularin family share a common functional core, which encompasses the PI-binding PH-GRAM (Pleckstrin homology-glucosyltransferases, rab-like GTPase
activators and myotubularin) domain [121-123] and the catalytic PTP (Protein tyrosine phosphatase) domain (Fig. 6). Most myotubularins also contain a CC (coiled-coil) region, which is important for homodimerization and/or heterodimerization [124]. In addition, MTM1 and MTMR1/2 have a PDZ-binding domain at the C-terminal end to mediate protein-protein interactions. Other individual family members, MTMR3/4 have a C-terminal FYVE domain [125], while MTMR5/13 contain a DENN [126] and a PH domain at the N- and C- terminus, respectively (Fig. 6). Interestingly, the PTP phosphatase domains of some members are catalytically inactive. Hence, the MTM family of phosphatases can be divided into both catalytically active and inactive members. Similar to MTM1, the catalytically active members (MTMR2-4, MTMR7-8, and MTMR14) contain the intact canonic CX\textsubscript{5}R motif and preferentially hydrolyze PI(3)P and PI(3,5)P\textsubscript{2}. However, the rest of members in the family (MTMR5 and MTMR9-13) contain inherent missense substitutions at the conserved Cys and Arg residues within the CX\textsubscript{5}R motif and thus are catalytic inactive [127, 128]. Despite lacking enzymatic activity, these catalytic-dead phosphatases can heterodimerize with the active homologs thus may act in conjunction with active MTMRs to regulate, perhaps, the localization and activity of the active MTMR [124, 129, 130].

Fig. 6. Schematic domain organization of the phosphatases in the myotubularin family. Abbreviations: PH-GRAM, Pleckstrin homology-glucosyltransferases, Rab-like GTPase activators and myotubularin; CC, coiled-coil; FYVE: Fab1, YOTB, Vac1, and EEA1; Dead Ptase, catalytically-inactive phosphatase; DENN, Differentially expressed in neoplastic vs. normal cells.
Crystal structures of human MTMR2 in its apo form or in complex with its substrates have been determined. The structural information has significantly advanced our understanding of the catalytic mechanism as well as the pathological mutations associated with human disease for this family of phosphatases.

1.5.2 Crystal structure of myotubularin phosphatases

The crystal structure of human MTMR2 [93] reveals that the PH-GRAM domain adopts the PH domain architectural fold of 7 β-strands sealed on one side by a C-terminal α-helix. It is connected to the phosphatase domain via a 20-residue linker and is directly apposed to the phosphatase domain to form a compact globular structure (Fig. 4G). However, a recent structure of MTMR6 (PDB ID: 2YF0; Moche et al., unpublished) reveals that the PH-GRAM domain is completely displaced from the phosphatase domain. This difference suggests that at least in some myotubularins, the PH-GRAM domain has an independent role to the phosphatase domain. The function of the PH-GRAM is likely to facilitate membrane association via the binding with PIs by flexible solvent accessible loops (loops connecting β1/β2, β3/β4, and β6/β7). The phosphatase domain (~375 residues) is composed of a flat 7-stranded β-sheet surrounded by 13 α-helices. Similar to other CX5R motif-based phosphatases, the catalytic core consists of 4 parallel β-strands, one α-helix, and a loop containing the catalytic CX5R motif connecting between one of the β-strands and the α-helix (Fig. 4H). Conserved in all phosphatases acting at the membrane interface, the structure also reveals a highly positively charged surface encompassing the active site pocket (Fig. 4I).

1.5.3 Substrate specificity and catalytic mechanism

The complex structures of MTMR2 with its substrates PI(3)P and PI(3,5)P2 revealed the structural component for membrane interaction [92, 131] (Fig. 7A). The aliphatic lipid acyl chains of the substrates are packed against helix α6, which contains a number of surface exposed hydrophobic residues and two lysine residues (Fig. 7B). This α-helix is highly solvent exposed and
projects above the substrate binding pocket. The structure and chemical nature of α6 may allow it to partially insert into membrane bilayer while membrane-embedded substrates are being hydrolyzed.

The complex structures also highlighted structural features for substrate selectivity [92, 131]. First, MTMR2 has a unique D5 phosphate recognition site, where R459 and R463 engage with the D5 phosphate of PI(3,5)P₂ through extensive salt bridges as well as hydrogen bonding interactions (Fig. 7C). However, the D5 phosphate recognition site is not involved in the binding of PI(3)P. The recognition of PI(3)P is through two other key structural features. First, residues N330, N355, S418, and R423 form another site to accommodate the D1 phosphate of the substrate by making hydrogen bonds with the D1 phosphate. Second, the side chains of R463 and particularly the bulky W421 are close to and form hydrogen bonds with the 4’ hydroxyl group of the inositol ring. These two residues prohibit the binding of any PIs phosphorylated at the D4.
position due to a potential steric clash with the D4 phosphate. Together, these structural features provide not only the selectivity but also the proper positioning of the D3 phosphate of the substrate. The D3 phosphate makes seven hydrogen bonds with the amides of the P-loop and the guanidinium side chain of R423 and is readily attacked by the catalytic cysteine. Due to the conservation of the catalytic motif, it is plausible to propose a similar catalytic mechanism for all CX₅R-based phosphatase including myotubularins. The detailed mechanism is illustrated in Fig. 8.

1.6 4-phosphatases

The first member of this unique CX₅R motif-based phosphatase family was demonstrated to hydrolyze specifically the D4 phosphate of PI(3,4)P₂ with proteins purified from rat brain [132]. Subsequently, two mammalian genes encoding PI(3,4)P₂ 4-phosphatases, INPP4A and INPP4B were cloned [133, 134]. INPP4A and INPP4B share an overall 37% sequence identity and a similar domain structure of an N-terminal C2 domain, large unknown central domain, and a C-terminal catalytic domain [135]. INPP4A is shown to be a negative regulator of Akt signaling [136]. Mutations in this gene in two different mouse models lead to neurodegeneration, suggesting a critical role of INPP4A in neuronal function [137, 138]. INPP4B is also shown to reduce Akt activation and accumulating evidence suggests INPP4B is a tumor suppressor in human cancers [135]. The 4-phosphatase catalytic domain is found to share sequence homology with two other enzymes, P-Rex1 and P-Rex2 [139]. Both P-Rex1 and P-Rex2 have a RacGEF activity mediated by their N-terminal DH-PH domain. However, neither of the 4-phosphatase homology domains exhibits 4-phosphatase activity [135, 139].

Unfortunately, crystal structure of any members in this family is still unavailable. It is intriguing to address how this group of enzymes achieve their unique substrate specificity by hydrolyzing the D4 phosphate only from PI(3,4)P₂. Due to their essential role in mammals and links to human diseases, elucidation of the protein structure of this family is in dire need.
Fig. 8. Schematic model of the general catalytic mechanism for all CX2R containing PI phosphatases. The reaction has two steps. In the first step, the catalytic cysteine attacks the scissile phosphate to break the bond between the phosphate and hydroxyl group of the lipid. D650 functions as the general acid to donate a proton to the leaving hydroxyl, resulting in a phospho-cysteine intermediate. In the second step, D650 now functions as the general base to activate a water molecule. The activated water molecule hydrolyzes the phospho-cysteine intermediate to restore the enzyme.

1.7 Discussion

Over the past years, our understanding of the structural basis for the function of each family of PI phosphatases has expanded considerably. Although the fundamental structural fold and the catalytic mechanism are apparently different between the two major classes of PI phosphatases, some structural features are common among them. First, PI phosphatases usually have a membrane interaction motif near the catalytic site, which not only plays a role in anchoring the enzyme to the membrane interface but also in discriminating specific membrane bilayers in terms of membrane curvature or lipid composition. Second, for phosphatases that hydrolyze PIs with di- or tri- phosphate group, a cationic pocket is usually present adjacent to the catalytic site to provide substrate selectivity by the recognition with the non-scissile phosphate group on the inositol ring. Despite the substantial progress, paramount questions still remain unaddressed. The 4-phosphatase family still lacks a representative structure to explain their unique D4 phosphatase activity against PI(3,4)P2. How the substrate specificity is achieved by members in the Sac1 and PTEN families is still uncertain due to the lack of complex structures of the enzymes with their cognate substrates. Moreover, many PI phosphatases contain regulatory domains besides the catalytic domain. How these regulatory domains function in cellular targeting or direct control of the
enzyme activity require more investigations. Further structural works pursuing these questions will certainly enhance our knowledge about the molecular basis of PI phosphatases and provide a framework for the understanding of human genetic diseases and the development of potential therapeutics.
Chapter 2: The Sac domain-containing phosphoinositide phosphatases

Abstract: The presence of Sac domain-containing proteins spans across different kingdoms. Although the Sac domain is homologous among different Sac domain-containing proteins, all appear to exhibit varied substrate specificity and subcellular localization. Dysfunctions in several members of this family are implicated in a range of human diseases such as cardiac hypertrophy, bipolar disorder, Down’s syndrome, Charcot-Marie-Tooth disease (CMT) and Amyotrophic Lateral Sclerosis (ALS). In plants, several Sac domain-containing proteins have been implicated in the stress response, chloroplast function and polarized secretion. In this chapter, we focus on recent findings in the family of Sac domain-containing PI phosphatases in yeast, mammal and plant, including the structural analysis into the mechanism of enzymatic activity, cellular functions, and their roles in disease pathophysiology.

2.1 Introduction

In the past 30 years, the roles of PIs have been characterized extensively in both uni- and multi-cellular organisms. PtdIns comprise ~10% of total cellular lipids and PI constitute less than 1%, yet they control vital cellular processes such as cell signaling, proliferation, organization of cytoskeleton, and membrane trafficking [2-4]. It has also been shown that PIs play a role in regulating ion channel activity [140], transcription, mRNA trafficking, RNA splicing [141], chromatin structure [142], nuclear export [143], and bacterial infection [144, 145].

The mechanism for controlling such divergent processes is achieved mainly by three non-mutually exclusive schemes. First, each PI has its own unique subcellular localization to allow them to selectively recruit proteins containing PI recognition modules (e.g. ANTH, ENTH, FYVE, PX, PH, PDZ, PHD, PTB, C2, GRAM, PROPPINs, Tuby, and FERM) to specific organelle membranes [146-149]. Second, the resulting heterogeneous anionic charge (produced by the
number of phosphate groups) at each organelle membrane can differentially recruit soluble polycationic proteins [150]. Third, PI in combination with specific Rab GTPase can constitute coincidence detection via recruitment of effectors with dual recognition domains [3, 151]. Coincidence detection increases the affinity and therefore the specificity of effector localization.

The heterogeneous distribution of PIs imparts an “organelle identity” to the cell. The maintenance of the selective concentration of specific PI species, as well as the dynamic control of PI composition in response to acute signaling inputs is achieved by a large number of PI kinases and phosphatases that are present throughout eukaryotic species. This chapter will focus on a family of these enzymes that contains a conserved PI phosphatase domain termed Sac (Suppressor of yeast actin mutations).

2.2 Sac1

2.2.1 Yeast and Mammalian Sac1

The structure and function of yeast Sac1 is described in Chapter 1.3. The role of mammalian homolog of Sac1 has also been studied extensively. In mammals, Sac1 is ubiquitously expressed in most mouse tissues with a particularly high level in cerebellum, hippocampus, and heart [65]. Similar to yeast Sac1, mammalian Sac1 shuttles between ER and Golgi apparatus. In quiescent cells, mammalian Sac1 accumulates at the Golgi to downregulate anterograde trafficking by depleting PI(4)P, thus slow secretion from the Golgi apparatus [152-154]. The Golgi localization has been shown to be dependent on the leucine zipper region of mammalian Sac1 to form oligomer and the recruitment of the coat protein COPII complex. When quiescent cells are stimulated by mitogens, Sac1 dissociates from its oligomeric state and rapidly shuttles back to the endoplasmic reticulum (ER). The retrograde trafficking of Sac1 is mediated by the coatomer COPI complex through the interaction with a canonical KXXXX motif at C-terminus of mammalian Sac1 [155] (Fig. 9A).

Unlike the case in yeast, genetic ablation of the single murine Sac1 gene results in progeny
failing to progress past E3.5 [74] (Table 1, page 35). In mammalian cells, knockdown of Sac1 expression results in disorganization of Golgi membranes, and mitotic spindle, which leads to a decrease in cell viability in cells due to a blockade in progression through the G2-M cell cycle [74, 156].

2.2.2 Sac1 in other species

In Drosophila, Sac1 has been shown to regulate axon guidance in the embryonic CNS midline and Sac1 mutants die as embryos with defects in dorsal closure [72, 157]. Loss of
Drosophila Sac1 leads to improper activation of several key events during development: cell shape change in the amnioserosa, increase in JNK signaling [72] and activation of Hedgehog signaling [158]. These collective data suggest an evolutionarily conserved function and point to an essential housekeeping role of Sac1 in multicellular organisms.

2.3 Sac2/INPP5f

Sac2 is the only Sac phosphatase that lacks an ortholog in S. cerevisiae, but is found in other Fungi genus such as Aspergillus spp. and Yarrowia lipolytica. Initial cloning and characterization show that this enzyme exhibits 5-phosphatase activity specific for PI(4,5)P2 and PI(3,4,5)P3, generating PI(4)P and PI(3,4)P2, respectively [76]. However, our recent data revealed that Sac2 is a PI(4)P phosphatase that regulates endocytic recycling (see Chapter 2 for details).

2.4 Sac3/Fig4p

2.4.1 Yeast Sac3/Fig4p

The function of yeast Sac3/Fig4p is briefly described in Chapter 1.3. Unlike other Sac domain proteins, Sac3/Fig4p dephosphorylation reaction is Mg²⁺ dependent [82]. Mutations in Fig4 result in abnormal actin distribution at the shmoo tip, a failure to establish mating cell polarity leading to enlarged cells, and reduced mating efficiency [79]. Fig4p-GFP is localized to the limiting membrane of yeast vacuole, and this localization is dependent on a scaffold protein Vac14 [82]. Interestingly, Vac14 also positively regulates the Fab1 kinase that is responsible for generating PI (3,5)P2 from PI(3)P, suggesting the formation of a multi-protein complex in regulating PI(3,5)P3 levels [82]. In this Vac14-Fig4-Fab1 complex, Fig4 is unexpectedly required for the activation of Fab1 to synthesize PI(3,5)P2 [159]. For this reason, deletion of Fig4 elicits reduced, rather than elevated levels of PI(3,5)P2. In agreement with this notion, Fig4 is found to be responsible for the hyperosmotic shock-induced increase as well as the turnover in PI(3,5)P2 levels [80].
2.4.2 Mammalian Sac3/Fig4p

In mouse, Sac3/Fig4 RNA transcript is ubiquitously detected in all tissues, with highest level in the testes, spleen and heart [87], whereas Sac3/Fig4 protein level is highest in the brain [85, 160]. Genetic mutations of Sac3/Fig4 lead to a number of diseases in human and mouse, including an autosomal recessive Charcot-Marie-Tooth disorder (CMT4J) and a subset of ALS in human [86, 87, 161] as well as neurodegeneration in the “pale tremor” mouse that exhibits severe tremor, abnormal gait and diluted pigmentation [87, 162] (Table 1, page 35). Fig4 null mice reveal a dramatic reduction in myelin in the brain, spongiform degeneration, gliosis, and juvenile lethal lethality [163-165]. These neurological defects have been proposed to be caused by enlarged vacuoles in neurons, which may physically interfere with normal vesicular trafficking. In support of this hypothesis, time-lapse imaging of fibroblasts from CMT4J patients displays impaired trafficking of intracellular organelles due to enlarged vacuoles [161]. Furthermore, Fig4-/- derived fibroblasts and neurons exhibit enlarged endosomal and lysosomal compartments, and impaired organelle trafficking [166].

2.4.3 Sac3/Fig4p disease mutations

Sac3 is a large protein with an N-terminal Sac homology domain (Fig. 9A). Several missense mutations in the gene are found to be responsible for the CMT4J (I41T and E302K) [87, 162] and ALS disorder (D48G, D53Y, R388G, and I411V) [86]. Based on homology structure modeling of the Sac phosphatase domain of human Sac3 to the crystal structure of yeast Sac1, I41T, which is a recessive mutation found in patients with CMT4J, is mapped to the Sac N-terminal sub-domain and 40 Å away from the catalytic site (Fig. 10). The I41T substitution may not affect the catalytic activity of Fig4, but the protein stability and/or protein binding ability. Consistent with this, I41T mutant shows no significant difference in its PI(3,5)P2 phosphatase activity [87]. In contrast to WT Fig4, I41T mutation leads to a rapid decline in protein level even in the presence of ArPIKfyve (human homolog of Vac14) [167] and in yeast two-hybrid system, interaction between Fig4 I41T and Vac14 is significantly impaired [162]. Another CMT4J mutation
identified at high frequency is E302K. This residue is located on a β-strand of the catalytic domain and is at the interface with the N-domain where it makes several hydrogen bonds with the surrounding residues (Fig. 10). Mutation of E302K is predicted to destabilize the protein and render the protein inactive. This is supported by the fact that Fig4 E302K construct fails to rescue the enlarged vacuole observed in Fig4 mutant in yeast [164].

D48G and D53Y, which have been predicted to be responsible for a subset of ALS disease [86], are mapped to the exposed surface area on the N-domain. These two mutations may affect the surface property and consequently may interfere with protein–protein interactions mediated through the N-domain. The R388G mutation is located in a loop close to the catalytic P-loop (Fig. 10). This arginine residue likely contributes to positive electrostatic potentials at the catalytic site. R388G mutation may directly affect substrate binding and thus the catalytic activity of the enzyme. In fact, R388G mutant fails to rescue enlarged vacuole phenotype in ΔFig4 yeast strain [86]. The other mutation, I411V is found in a hydrophobic core on the opposite site of the P-loop (Fig. 10). The deleterious effect of this substitution by a valine residue of similar hydrophobic property is not clear as the I411V mutant behaves similar to wild type in yeast rescue assays [86].

2.5 Synaptojanins

2.5.1 Enzyme activity and domain organization

Synaptojanins contain an N-terminal Sac domain, a central 5- phosphatase domain, a RNA binding domain (RBD) that is conserved in multi-cellular organisms, but not in yeast, and a C-terminal PRD (Fig. 9A). The Sac phosphatase domain is ~30% sequence identical to yeast Sac1. Biochemical characterization shows that the N-terminal Sac domain of synaptojanins predominantly dephosphorylates PI(3)P and PI(4)P, whereas the 5-phosphatase domain dephosphorylates PI(4,5)P2 and PI(3,4,5)P3 at the 5’ position of the inositol ring [26, 168] (Table 1). It is speculated that the Sac domain and the 5-phosphatase domains may function in a concerted way. In the concerted model, the 5-phosphatase domain converts PI(4,5)P2 to PI(4)P, which is directly channeled to Sac domain for further dephosphorylation to PI(3,4,5)P3 without accumulation of the intermediate.
Figure 10. Disease mutations mapping on the modeled structure of the Sac domain of human Sac3/Fig4. The Sac domain of human Sac3 is modeled based on the crystal structure of yeast Sac1 with the MODELER program. The local structural environment of six known CMT4J and ALS associated mutations are shown in zoomed-in panels.

### 2.5.2 Yeast synaptojanins function and regulation

*S. cerevisiae* encodes in its genome three synaptojanin-like genes, sjl1, sjl2, and sjl3, which are involved in endocytic membrane trafficking pathway [169, 170]. It is interesting to note that the Sac domain in sjl1 lacks the conserved CX5R motif and consequently does not exhibit phosphatase activity [26]. Although neither of these genes is essential for viability, a triple mutant of Δsjl1-3 is inviable under normal growth conditions [171] and a Δsjl1Δsjl2 double mutant displays a pronounced increase in PI(4,5)P2 level. The hydrolysis of PI(4,5)P2 is critical for clathrin and actin-mediated endocytosis, including vesicle scission [172]. Consistent with this, sjl2 is shown to localize to cortical actin patches via recruitment of actin patch component Abp1. The physical interaction is critical for the regulation of vesicle formation and fission during endocytosis [173]. In addition, sjl1 and sjl2 have been implicated in eisosome receptor-mediated endocytosis based en route.
on studies using eisosome marker Pil1 [174]. Besides regulating the levels of PI(4,5)P_2, sjl2 and sjl3 are also involved in mediating PI(3)P levels. In Δymr1Δsjl3 double mutant (ymr1 is the ortholog of MTM family of PI-3-phosphatase), cells display more than two fold increase in PI(3)P level compare to either single mutant [175]. The role of sjl3 in PI(3)P level likely contributes to the AP-1/clathrin transport from TGN (trans Golgi network) to endosome [176].

2.5.3 Mammalian synaptojanins function and regulation

There are two members that are present in mammals: synaptojanin 1 and 2. The founding member of this family, the mammalian synaptojanin 1 was first cloned and characterized in the search for presynaptic factors involved in synaptic vesicle recycling [22]. Synaptojanin 1 can be alternatively spliced to form a 145 kDa and 170 kDa isoforms. The 145 kDa isoform is highly expressed in adult neurons and at nerve terminals where it is involved in clathrin-dependent synaptic vesicle recycling [23]. On the other hand, the 170 kDa isoform, containing an additional C-terminal fragment involved in protein–protein interactions, is ubiquitously expressed in a variety of tissues [177]. These two splicing isoforms have been shown to be differentially targeted to and regulate the maturation of clathrin-coated pits through the interactions with an array of proteins at sites of clathrin-mediated endocytosis [178]. It has been shown that both isoforms interact with the BAR domain protein endophilin to preferentially remove PI(4,5)P_2 from curved membranes as opposed to flat ones [57]. Other interacting partners include amphiphysin [179], intersectin [180, 181], Grb2 (growth-factor- receptor-bound protein 2) [28], Snx9 [182], and Myosin 1E [183]. All these interactions are critical for the proper localization at clathrin-mediated endocytic sites and/or catalytic activity of the enzyme. However, only the 170 kDa isoform directly interacts with clathrin and its adaptor protein AP-2 [184] via its unique C-terminal portion. These studies suggest a role of synaptojanin in endocytosis. The hydrolysis of PI(4,5)P_2 by synaptojanins may facilitate the disassembly of clathrin from endocytic vesicles [185].

The in vivo functions of synaptojanin are also well studied in knockout mice. Deletion of this gene in mice causes neurological defects and early postnatal lethality [23]. In neurons of
synaptojanin mutant mouse, abnormal amount of clathrin-coated vesicles are observed to accumulate at the nerve termini and a delay in re-entry of recycling vesicles, which correlates with increased PI(4,5)P$_2$ levels [23, 186, 187]. These mutant mice are further shown to have impaired constitutive and induced endocytosis of post synaptic AMPA receptors [188]. In contrast to synatojanin 1 deletion, overexpression of synaptojanin 1 in transgenic mice or in neuroblastoma cell lines leads to enlarged endosomes [189]. Interestingly, synaptojanin 1 has been shown to be a trisomic gene located at chromosome 21 that is over-expressed in Down’s syndrome patients. The increased levels of synaptojanin 1 may account for the observed enlarged endosomes in neurons and brain dysfunction in Down’s syndrome patients [30, 31]. Mutations of synaptojanin 1 have also been linked to bipolar disorder [32, 33]. However, the mechanism for the association of synaptojanin 1 with bipolar disorder is not clear.

Similar to synaptojanin 1, synaptojanin 2 also consists of several splice isoforms that are referred to as 2A and 2B, in which the 2B variant can undergo further alternative splicing to generate 2B1 and 2B2 [190]. The 2A isoform is broadly expressed in many tissues, while the 2B isoform is predominantly expressed at nerve terminals in the brain as well as at spermatid manchette in adult testis [168, 191]. Unlike synaptojanin 1, the function of synaptojanin 2 is less well understood. It has been shown that synaptojanin 2 can be recruited to plasma membrane via the direct interaction with active Rac1 [192]. The interaction of synaptojanin 2 with Rac1 has been suggested to play a role in human glioma cell invasion and migration [193]. In mouse mutant strain called Mozart, a mutation in synaptojanin 2 that abolishes its enzymatic activity has been shown to be responsible for the progressive hearing loss [35], suggesting a link between phosphoinositide metabolism to hair cell survival and hearing.

2.5.4 Synaptojanins in other species

The phenotypes observed in both mice and human are further corroborated by studies in Danio rerio, Drosophila melanogaster, and Caenorhabditis elegans. In D. rerio, synaptojanin 1 mutant displays a defect in cone photoreceptor ribbon anchoring, abnormal synaptic transmission at
synapses [194] as well as a role in maintaining the quantity, fusion and release of synaptic vesicles at hair cell ribbon synapses [195]. In *D. melanogaster*, synaptojanin is recruited by endophilin to promote synaptic vesicle uncoating [196]. Overexpression of synaptojanin in *Drosophila* leads to abnormal synaptic morphology [197]. In *C. elegans*, mutations in synaptojanin (unc-26) not only disrupt synaptic vesicle recycling [198], but also enhance polyglutamine toxicity which is a phenotype observed in Huntington-interacting protein 1 mutant implicated in neuronal function in Huntington's disease [199]. These data point to a role of synaptojanins in clathrin-mediated endocytic processes particularly in synaptic vesicle recycling at the nerve terminus.

2.6 Plant Sac phosphatase family

Phosphoinositides in plants regulate many cellular activities such as vesicle trafficking [200], stomatal movement [201], polar tip growth of pollen tube and root hairs [202], and responses to stress and hormonal treatments [203, 204]. In *Arabidopsis thaliana*, a genome wide analysis reveals the presence of nine Sac domain-containing phosphatases with high sequence similarity (55%–69%) to Sac domain of yeast Sac1, which are subsequently named AtSac1 to AtSac9 [205]. These nine genes can be sub-divided into three groups based on sequence similarity: 1) Group I contains C-terminal sequences with a range of 252–338 amino acids. 2) Group II contains a C-terminal transmembrane motif. 3) Group III consists of only one member, AtSac9, with a long stretch of C-terminal sequences of ~1100 amino acids [205] (Fig. 9B). Although other Sac homologs are found in other plant species, the following section will mainly focus on the Sac domain phosphatases in the context of *Arabidopsis*.

The group I enzymes consist of AtSac1-AtSac5, which are considered as Fig4 homologs based on sequence identity. Their gene expression is widely distributed in most plant organs, with relatively lower level in mature leaves. Among this group, AtSac1 has been well characterized. AtSac1 is a PI-phosphatase specific for PI(3,5)P2 and mutations in AtSac1 cause defects in cell morphogenesis, such as a decrease in cell wall synthesis, a reduction in cell elongation, defect in actin cytoskeleton and a global change in plant architecture [206].
The group II, which consists of AtSac6–AtSac8, is most closely related to yeast Sac1, which contains two transmembrane motifs at its C-terminal end. AtSac7 and AtSac8 are widely expressed in most plant organs, whereas AtSac6 is only present in flowers [205]. Among the three enzymes, only AtSac7 has been characterized so far. AtSac7 has been shown to be enriched at the TGN-like compartment at the tips of growing root hairs [207]. Mutation of this gene is responsible for the root hair defective 4-1 (rhd4-1) phenotypes that display aberrant root hair and polarized tip growth. This phenotype is associated with an accumulation of PI(4)P in the tip root hairs, consistent with the in vitro assay showing its preferred PI phosphatase activity toward PI(4)P [207, 208]. AtSac9 is the only enzyme belonging in Group III due to its distinct domain features that include a long C-terminal domain and the presence of a WW domain within the Sac domain (Fig. 9B). Microarray analysis shows highest expression in the roots and lower expression in the leaves and flowers [205]. AtSac9 mutant displays elevated PI(4,5)P_3 and water soluble Ins(1,4,5)P_3 levels in the roots compared to wild type plants [209]. The AtSac9 mutants are hypersensitive to external stress and constitutively overexpress stress-induced genes and over-accumulate reactive-oxygen species [209].

2.7 Discussion

It has only been a little more than 20 years since the discovery of the first Sac domain-containing protein, the yeast Sac1. With the advance of whole-genome sequencing, all the Sac homology domain-containing PI phosphatases have been annotated in quite a few species, including yeast, mammal, and plant. Extensive functional studies on this family of PI phosphatases have been documented and many genetic diseases have been linked to mutations in genes encoding PI phosphatases in this family. Meanwhile, the crystal structure for the Sac phosphatase domain provides an excellent model for the molecular understanding of genetic mutations occurred on Sac domain-containing phosphatases. Despite the current progress in this field, there are still several key questions that remain to be addressed. Given that the catalytic pocket residues are highly conserved across species (Fig. 11), the mechanistic details for how substrate recognition and
specificity are achieved would require detail structural information of the Sac domain with its cognate substrate. Moreover, how the activity of individual Sac domain containing proteins is precisely controlled by sophisticated signaling networks under physiological or disease state remains to be explored in the future. With the availability of disease models that have deletions or mutations of the Sac domain family members from a variety of organisms, future studies to screen for small-molecules that interfere with phosphoinositide metabolism may lead to potential therapeutic agents for the treatment of human diseases.

**Fig. 11. Sac domain structure of yeast Sac1.** (A) The Sac phosphatase domain is comprised of two subdomains: a novel N-terminal domain (blue) and the catalytic domain (yellow). The loop colored red represents the CX₅R motif. (B) Space filling model of the Sac domain with residues colored based on evolutionary conservation, which is calculated by an evolution trace method. The least and highest conserved residues are colored from white to yellow respectively.
Table 1. Sac domain enzymatic activities and associated phenotypes.

<table>
<thead>
<tr>
<th>Protein name</th>
<th>In vivo substrate</th>
<th>Major in vivo substrate</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeast Sac1</td>
<td>PK(3)P, PI(4)P, PI(5)P, PK(3,5)P₂</td>
<td>PI(4)P</td>
<td>Cold-sensitive growth, actin reorganization, inositol auxotrophy, etc.</td>
</tr>
<tr>
<td>Aspergillus spp.</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Mammalian Sac2</td>
<td>PK(4,5)P₂, PI(3,4,5)P₃</td>
<td>PI(3,4,5)P₇</td>
<td>Stress-induced cardiac hypertrophy</td>
</tr>
<tr>
<td>Yeast Fig4</td>
<td>PK(3,5)P₂</td>
<td>PI(3,5)P₂</td>
<td>Vacuole morphology defects</td>
</tr>
<tr>
<td>Mammalian Sac3</td>
<td>PK(3)P, PI(4)P, PI(3,5)P₂</td>
<td>PI(3,5)P₂</td>
<td>CMT4L, ALS, neurodegeneration in the pale tremor mouse and juvenile lethality</td>
</tr>
<tr>
<td>Synaptojanin 1</td>
<td>PK(4)P, PI(4,5)P₂, PK(3,4,5)P₂</td>
<td>PI(4,5)P₂, PI(3,4,5)P₃</td>
<td>Bipolar disorder, Down's syndrome, early postnatal lethality in mouse</td>
</tr>
<tr>
<td>Synaptojanin 2</td>
<td>PK(4)P, PI(4,5)P₂, PK(3,4,5)P₂</td>
<td>PI(4,5)P₂</td>
<td>Hearing loss associated with cochlea hair loss</td>
</tr>
<tr>
<td>SjI</td>
<td>PI(4,5)P₂</td>
<td>PI(4,5)P₂</td>
<td>Endocytic defects, increase sensitivity to temperature and neomycin</td>
</tr>
<tr>
<td>SjII</td>
<td>PK(3)P, PI(4)P, PI(3,5)P₂, PI(4,5)P₂</td>
<td>PI(4,5)P₂</td>
<td>Endocytic defects, increase sensitivity to temperature and neomycin</td>
</tr>
<tr>
<td>AtSac1</td>
<td>PK(3,5)P₂</td>
<td>PI(3,5)P₂</td>
<td>Cell morphology defects</td>
</tr>
<tr>
<td>AtSac2</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>AtSac3</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>AtSac4</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>AtSac5</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>AtSac6</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>AtSac7</td>
<td>PK(4)P, PI(3,5)P₂</td>
<td>PI(4)P</td>
<td>Defective root hair and polarized growth</td>
</tr>
<tr>
<td>AtSac8</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>AtSac9</td>
<td>N/A</td>
<td>PI(4,5)P₂</td>
<td>Hypersensitive to stress</td>
</tr>
</tbody>
</table>
Chapter 3: Spatiotemporal control of phosphatidylinositol 4-phosphate by Sac2 regulates endocytic recycling


Abstract: It is well established that the spatial- and temporal-restricted generation and turnover of phosphoinositides (PIs) by a cascade of PI metabolizing enzymes is a key regulatory mechanism in the endocytic pathway. Here, we demonstrate that, in contrast to the previous report, the Sac1 domain-containing protein Sac2 is a PI 4-phosphatase that specifically hydrolyzes phosphatidylinositol-4-phosphate (PI(4)P) in vitro. We further show that Sac2 co-localizes with early endosomal markers and is recruited to transferrin (Tfn)-containing vesicles during endocytic recycling. Exogenous expression of the catalytically inactive mutant Sac2C458S results in altered cellular distribution of transferrin receptors (TfnR) and delayed Tfn recycling. Furthermore, genomic ablation of Sac2 results in a similar perturbation on Tfn and integrin recycling as well as defects in cell migration. Structural characterization of Sac2 reveals a unique pleckstrin-like homology Sac2 domain (hSac2) conserved in all Sac2 orthologs. Collectively, our findings provide evidence for the tight regulation of PIs by Sac2 in the endocytic recycling pathway.

3.1 Introduction

Phosphoinositides (PIs) have a heterogeneous distribution in different membranes thus allowing selective recruitment of proteins containing PI recognition modules to specific organelle membranes. The maintenance of the selective distribution of specific PI species, as well as the dynamic control of PI composition in response to acute signaling inputs is achieved by a large number of PI kinases and phosphatases [1].

The Sac1 domain-containing proteins constitute one essential family of the PI phosphatases. In vertebrates, five genes have been identified to contain the Sac1 homology domain, which
include Sac1, Sac2, Sac3/Fig4, and synaptojanin 1 and 2. Among the Sac1 domain-containing proteins, Sac2 remains the least well understood. Sac2 is a 128 kDa protein encoded by the gene INPP5F. Besides the conserved N-terminal Sac1 domain, Sac2 contains a unique homology domain named hSac2, followed by a proline-rich C-terminal portion with various lengths in different species. Sac2 is an evolutionarily conserved protein in multi-cellular organisms from nematode to human. Sac2 orthologs are also found in fungi genus such as Aspergillus spp. and Yarrowia lipolytica, but is absent in Saccharomyces cerevisiae. Initial cloning and characterization of human Sac2 showed that this enzyme exhibited 5-phosphatase activity specific for PI(4,5)P\textsubscript{2} and PI(3,4,5)P\textsubscript{3} to generate PI(4)P and PI(3,4)P\textsubscript{2}, respectively [76]. In line with this activity, Sac2 was shown to hydrolyze PI(3,4,5)P\textsubscript{3} and thereby inactivate Akt signaling, leading to the attenuation of heart hypotrophy [77]. Moreover, Sac2 deficient mice displayed abnormal fetal gene reactivation and increased susceptibility to stress induced heart hypertrophy. On the other hand, Sac2 transgenic mice prevented the animal from developing these symptoms [78].

Despite these findings, many functional aspects of Sac2 are still not well understood. In this study, we demonstrate that Sac2 proteins solely hydrolyze PI(4)P. We show that Sac2 is present at endosomal structures, which are positive for early endosomal markers Rab4 and Rab5. We further reveal that the exogenous expression of a catalytically inactive form of Sac2 or ablation of Sac2 by the CRISPR (clustered regularly interspaced short palindromic repeats) technique reduces transferrin receptor (TfnR) and \(\beta\)1 integrin receptor surface levels and delays their recycling. Furthermore, Sac2 knockout cells display a severe defect in cell migration, suggesting a link between recycling and motility. In addition, we solve the crystal structure of a unique domain hSac2. Overall, our data support a specialized role of Sac2 in the endocytic pathway through the spatiotemporal control of PI(4)P.

### 3.2 Results

#### 3.2.1 Sac2 is highly expressed in the brain

Sac2 mRNA transcripts are found ubiquitously expressed but more enriched in the brain,
heart, skeletal muscles, kidney, and placenta as revealed by Northern blot analyses [76]. Using specific poly-clonal antibodies generated against the unique hSac2 domain (aa. 593-760 of human Sac2), we were able to detect the tissue distribution of Sac2 at protein levels. We found that Sac2 has the highest expression levels in the whole adult mouse brain, as well as the hypothalamus tissue (Fig. 12A). Sac2 protein is also detected at high levels in the lung and the pancreas, and at lower levels in the heart. Contrary to the Northern blot data, there is no detectable Sac2 protein in the kidney and the skeletal muscle (Fig. 12A). Consistent with the high protein levels in the brain, Sac2 is highly expressed in several neuronal cell lines, such as Neuro-2A (N2A) and NSC-34 (Fig. 12B). These data suggest that Sac2 plays an important role in neuronal functions or in brain development. We thus chose the mouse neuroblastoma N2A cell line for subsequent experiments unless otherwise specified.

Fig. 12. Sac2 is a PI 4-phosphatase that specifically hydrolyzes PI(4)P. (A) Mouse tissue lysates were normalized and analyzed by Western blot with specific anti-Sac2 antibody. WBL: whole mouse brain lysate. (B) Different cell lines were harvested and analyzed by Western blot with anti-Sac2 and anti-Tubulin antibodies. (C) (Left) Phosphoinositide phosphatase assay with recombinant Sac2 proteins purified from insect cells. (Right) SDS-PAGE gel analysis of the purified Sac2 samples. (D) Quantification of phosphatase assays shown in (C). (E) (Left) Phosphatase assay with endogenous Sac2 immuno-purified by anti-Sac2 antibody. (Right) Western blot with anti-Sac2 of the purified protein samples. (F) Quantification of the assay shown in (E). Data are from three replicate experiments (mean ± SEM).
3.2.2  Sac2 is a phosphoinositide 4-phosphatase

We next reinvestigated the enzymatic activity of Sac2. It has been reported that Sac2 is a 5-phosphatase that removes the 5’ phosphate from the inositol ring of PI(4,5)P$_2$ and PI(3,4,5)P$_3$ [76]. However, in vitro phosphatase activity of the Sac1 domain from Sac1 or synaptojanins predominantly dephosphorylates PI(3)P and PI(4)P, and weakly against PI(3,5)P$_2$ [26]. Given the high primary sequence homology between Sac1 and Sac2 [89], it is surprising that Sac2 is a PI 5-phosphatase capable of hydrolyzing di- and tri- phosphorylated PI species exclusively. To resolve this apparent discrepancy, GST-tagged full-length human Sac2 protein was expressed and purified from insect cells. In contrast to the previous report [76], GST-Sac2 protein exhibited specific phosphatase activity towards PI(4)P but no activity on either PI(4,5)P$_2$ or PI(3,4,5)P$_3$. Furthermore, a single amino acid mutation of the catalytic cysteine (C458S) completely obliterated the enzymatic activity of Sac2 (Figs. 12C and D). To test whether the observed activity of Sac2 is conserved in other species, endogenous mouse Sac2 from N2A cells was immunoprecipitated using the Sac2 polyclonal antibody. Similarly, mouse Sac2 also exhibited enzymatic activity towards PI(4)P exclusively (Figs. 12E and F). The specific phosphatase activity is also observed with either GFP- or Flag-tagged Sac2 proteins expressed and affinity purified from HEK293T cells (Fig. 13). Taken together, these data demonstrate that Sac2 is a PI 4-phosphatase that specifically hydrolyzes PI(4)P to phosphatidylinositol.

3.2.3  The catalytically inactive mutant of Sac2 is localized to punctate structures

To detect the intracellular localization of Sac2, we first attempted to use our Sac2 antibodies to probe endogenous Sac2 in N2A cells. Although the antibody was able to recognize Sac2 from cell lysates by Western blots, we were unable to detect endogenous Sac2 signal by immunostaining using this antibody. To overcome this problem, N2A cells were transiently transfected with GFP-tagged Sac2 (GFP-Sac2) wild type or various Sac2 mutants (Fig. 14A). Wild type Sac2 showed a cytoplasmic localization (Fig. 14B). However, the catalytically inactive mutant GFP-Sac2CS displayed a punctate distribution in N2A cells (Fig. 14B). Both the C-terminal truncation
mutantSac2 1-768CS and ΔhSac2 (the hSac2 domain deleted) displayed a punctate distribution. However, both the catalytic domain alone (1-590) and the entire C-terminal region (590-1132) failed to localize to the punctate structures when expressed in N2A cells (Fig. 14B). These data suggest that the punctate localization of Sac2 requires both the catalytic domain and either the hSac2 domain or its C-terminal proline-rich domain. Since the main substrate of Sac2 is PI(4)P, we next determined the dependency of PI(4)P on Sac2 positive punctate structures. When N2A cells expressing GFP-Sac2CS were treated with the phosphatidylinositol-4-kinase inhibitor phenylarsine oxide (PAO) [210], GFP-Sac2CS became diffused and cytosolic while the proteins remained punctate localized in control cells treated with DMSO (Fig. 14C). These observations suggest that GFP-Sac2CS puncta are not caused by protein misfolding or aggregation. Instead, the GFP-Sac2CS proteins are likely recruited and trapped to its substrate PI(4)P on these punctate structures due to the loss of its catalytic activity.
3.2.4 Sac2 is localized to early and recycling endosomal structures

To examine the nature of Sac2CS puncta, we performed co-localization analyses with a series of membrane compartment markers. Sac2CS was detected at a small fraction of clathrin coated pits as indicated by the limited partial co-localization with RFP-tagged clathrin light chain (Figs. 15A and G). A significant co-localization of Sac2CS was observed with early endosomal marker Rab4 and Rab5 (Figs. 15B, C, and G). Sac2CS also displayed a partial co-localization with Rab11 recycling vesicles and the endocytic recycling compartment (ERC) at the perinuclear region (Figs. 15D and G). In contrast, little or no co-localization of Sac2CS was observed with late endosomal markers Rab7 and Rab9, and the lysosomal marker LAMP1 (Figs. 15E and G, Figs. 16A, B, and G). To exclude the possibility that the perinuclear signal could be Golgi apparatus, we also stained cells expressing GFP-Sac2CS with antibodies against the Golgi resident protein.
GPP130. Very little overlap between Sac2CS and GPP130 was seen (Fig. 16C). Furthermore, when these cells were treated with Brefeldin A (BFA), the GPP130 signal disappeared at the perinuclear region whereas GFP-Sac2CS remained unaffected (Fig. 16D), suggesting that GFP-Sac2CS does not localize to the Golgi apparatus. Collectively, these data are consistent with Sac2 being associated with the peripheral early endosomes and juxanuclear ERC.

The cellular localization of Sac2 at early endosomal structures implies a role of Sac2 in early endocytic events. We thus examined the co-localization of Sac2 with endocytic cargoes. GFP-Sac2CS expressing cells were either stained with monoclonal anti-TfnR antibodies or labeled with Alexa594-Tfn and chased for 15 min. Sac2CS displayed significant co-localization with Tfn

---

**Fig. 15. Sac2 localizes to early and recycling endosomes.** (A-E) Colocalization of either GFP- or mCherry-Sac2CS with various endocytic markers. A single plane of spinning disk confocal microscopy is shown. All fluorescent fusion proteins were transiently transfected. (F) Colocalization analysis of GFP-Sac2CS with Tfn. N2A cells were transfected with GFP-Sac2CS and were treated with Alexa594-Tfn for 15 minutes before fixation for imaging (scale bar = 10 μm). (G) Colocalization analysis of Sac2CS with various endocytic markers. The mean values of the Pearson's correlation coefficient (PCC) from three cells and the SEM are shown.
and TfnR in the punctate structures (Figs. 15F and G and Figs. 16E and G). To assess whether Sac2 specifically functions at the early stages of the endosomal recycling pathway rather than the late endosomal-lysosomal degradation pathway, we examined the co-localization of Sac2CS with EGF, an endocytic cargo targeted to lysosomes for degradation. HeLa cells, which express high levels of EGFR, were transfected with GFP-Sac2CS and labeled with Alexa555-EGF followed by chase for 15 min. In contrast to Tfn, GFP-Sac2CS showed little co-localization with EGF (Figs. 16F and G). These observations suggest that Sac2 is selectively recruited to endocytic cargoes that are programmed for recycling.

Fig. 16. Colocalization of Sac2CS with different endosomal and organelle markers in N2A and HeLa cells. (A) Colocalization of GFP-Rab9 and mCherry-Sac2CS. (B-E) Cells were transfected with either GFP-Sac2CS or mCherry-Sac2CS, and immunostained with indicated antibodies. (D) Cells were treated with 5 μg/ml of Brefeldin A (BFA) for 15 min at 37°C. Images shown are a single plane section taken from a spinning disk confocal microscope. (F) HeLa cells were transfected with GFP-Sac2CS and labeled with Alexa555-EGF. Images were taken after 15 min chase. (G) Colocalization analysis of Sac2CS with various endocytic markers. The mean values of the Pearson’s correlation coefficient (PCC) of three cells and the SEM are shown. (scale bar = 10 μm).
3.2.5 Sac2CS mutant alters the cellular distribution of TfnR and perturbs its recycling

To examine the role of Sac2 in the endocytic recycling of cell surface receptors, we first measured the steady state ratio of TfnR levels on the cell surface versus the interior. N2A cells were transfected with GFP, GFP-Sac2CS or GFP-Sac2 for 48 hours, and then shifted to 4°C to halt endocytosis. Cell surface proteins were biotinylated with Sulfo-NHS-SS-biotin. Biotinylated surface proteins were isolated from cell lysates using streptavidin agarose beads, and the amount of surface TfnR was quantified using Western blot. A significant reduction (~30%) in the protein levels of surface TfnR was observed in cells expressing GFP-Sac2CS compared to either GFP control or WT Sac2, whereas the total levels of TfnR were similar in all three conditions (Figs. 17A and B). These data suggest that exogenous expression of Sac2CS reduces the surface presentation of TfnR at steady state. We next examined the effect of Sac2 on the dynamics of Tfn and TfnR recycling. The kinetics of Tfn recycling was monitored by pulse-chase flow cytometry assays. Within the first 10 min, the Tfn fluorescent signal in control cells was reduced to 45.8%. However, the Tfn signal in cells expressing GFP-Sac2CS decreased only to 65.3% (Fig. 17C). Retardation of Tfn recycling persisted to later time points (Fig. 17C). Together, the pulse-chase data support a role of Sac2 in the regulation of Tfn recycling.

3.2.6 CRISPR-mediated knock-out of Sac2 delays TfnR recycling

To further investigate the role of Sac2 in Tfn recycling, we generated N2A mutant cell lines with a complete disruption of Sac2 expression by the CRISPR technique [211, 212]. Similar to
cells transiently expressing Sac2CS, Sac2 null cells showed a reduced surface distribution of TfnR (Figs. 18A and B). In agreement with the notion that Sac2 plays a role in Tfn recycling, pulse-chase flow cytometry analyses showed a delayed recycling of Tfn in Sac2 null cells (Fig. 18C).

![Fig. 18. Defects of Tfn recycling in Sac2 null cells.](image)

(A) Western blot analysis of biotin surface labeled TfnR. (B) Quantification was performed as in Fig. 17B. Values are normalized to WT cells (surface/total). Data are from three replicate experiments (mean ± SEM). (C) Flow cytometry analysis of Tfn recycling in WT and Sac2 null N2A cells. Data are from three replicate experiments (n = 5000 in each time point, mean ± SEM). (D) Representative images of Tfn recycling at the indicated times points. Last row: cells transfected with Flag-Sac2 (green). (scale bar = 10 μm). **p < 0.05; ***p < 0.01.
More importantly, this delay of Tfn recycling was restored by the re-expression of WT GFP-Sac2 in Sac2 null cells (Fig. 18C). The effect on Tfn recycling in Sac2 null cells was visualized by confocal microscopy. Cells were labeled with Alexa647-Tfn and chased at indicated time points (Fig. 18D). A prominent retention of intracellular Tfn signals was observed in Sac2 null cells at later time points, which again suggests a delay in Tfn recycling (Fig. 18D). A prominent retention of intracellular Tfn signals was observed in Sac2 null cells at later time points, which again suggests a delay in Tfn recycling (Fig. 18D). Notably, the reduced surface signal and delayed Tfn recycling in Sac2 null cells were rescued by expressing WT Sac2 (Fig. 18D). Together, these data demonstrate that Sac2 is an important regulator in the Tfn/TfnR recycling pathway.

3.2.7 CRISPR-mediated knock-out of Sac2 perturbs integrin distribution and recycling

To assess whether Sac2 has a general impact on the recycling of other endocytic cargos, we analyzed integrin recycling in Sac2 null cells. We first compared the surface distribution of integrin β1 in WT and ΔSac2 cells by flow cytometry. Surface labeling revealed that the cell surface β1 integrin levels were reduced by approximately 20% in Sac2 null cells, whereas the total β1 integrin levels were similar (Figs. 19A-C). Immunostaining with anti-integrin β1 under non-permeabilized and permeabilized conditions further showed that the β1 integrin signal was significantly weaker at the cell surface in ΔSac2 cells compared to WT cells (Fig. 19D, top panels). In contrast to the cell surface levels of β1 integrin, a stronger intracellular β1 integrin signal was detected in ΔSac2 cells after permeabilization (Fig. 19D, lower panels). These results suggest that similar to TfnR, the intracellular distribution of β1 integrin was also perturbed when Sac2 was deleted. We then examined the effect of Sac2 on integrin recycling by an antibody-labeling pulse-chase experiment. In WT cells, internalized antibody-receptor complexes efficiently recycled back to the cell surface within 20 min, whereas in Sac2 null cells, significantly lower amount were returned (Fig. 19E, top panels). On the contrary, more labeled integrins were retained intracellularly in ΔSac2 cells compared to WT cells (Fig. 19E, lower panels), suggesting that integrin recycling is delayed in Sac2 null cells.
To assess whether Sac2 specifically affects endocytic recycling processes rather than other membrane trafficking events such as retrograde trafficking, we examined the localization of a chimeric form of the cation-independent mannose 6-phosphate receptor (CI-M6PR), which at steady state resides mostly in the perinuclear TGN region, but in conditions where retrograde trafficking is impaired, the receptor switches from a perinuclear region to peripheral structures or gets degraded [213, 214]. WT and Sac2 null cells were co-transfected with GFP-CI-M6PR and DsRed-GaTL, a TGN marker. We did not observe any changes in co-localization with DsRed-GaTL or fluorescence signal of CI-M6PR between WT and Sac2 null cells (Fig. 20). Taken together, these data suggest that Sac2 is specifically involved in the endocytic recycling of a number of cell
surface receptors but not in other membrane trafficking pathway such as retrograde trafficking.

**Fig. 20. The cellular distribution of CI-M6PR is unaffected in Sac2 null cells.** (A) WT and Sac2 null cells were co-transfected with GFP-M6PR and DsRed-GalT. (scale bar = 10 μm). (B) Colocalization analysis of GFP-M6PR with DsRed-GalT. The mean values of the Pearson's correlation coefficient (PCC) from three cells and the SEM are shown. (C) The mean total cell fluorescence intensity and the SEM are shown (n = 10).

### 3.2.8 Tfn-containing compartment is positive for PI(4)P in Sac2 null cells

Our data showed that Sac2 specifically hydrolyzes PI(4)P and regulates Tfn recycling. These data led us to hypothesize that in Sac2 null cells, PI(4)P may accumulate at the compartments where Tfn is enriched. To detect this pool of PI(4)P, we used two commonly used PH domains from FAPP1 and OSBP [213-215] and the PI(4)P-binding domain of SidC (P4C) [216] as
specific PI(4)P probes. WT and Sac2 null cells were transfected with GFP-2xPHFAPP1 for overnight and then labeled with Alexa647-Tfn. Live imaging was carried out to monitor the in vivo PI(4)P dynamics. In WT cells, the PI(4)P probe predominantly labeled the Golgi apparatus with very little co-localization with Tfn (Figs. 21A and B). However, in Sac2 null cells, the PI(4)P probe not only strongly labeled the Golgi area, but also labeled punctate structures scattered in the cytoplasm. Interestingly, some of these PI(4)P puncta were positive for Tfn, suggesting that PI(4)P indeed transiently accumulates on Tfn-containing endosomal structures in Sac2 null cells. The increased levels of PI(4)P on Tfn-positive endosomal structures were completely reversed when WT mCherry-Sac2 was re-introduced into Sac2 null cells (Figs. 21A and B). Similar results were observed when OSBP (Figs. 21C and D) or P4C probes were used (data not shown). These observations suggest that Sac2 deletion causes a transient accumulation of PI(4)P on Tfn-positive endosomal structures and the hydrolysis of PI(4)P by Sac2 is critical for the endocytic recycling of cell surface receptors.

3.2.9 Sac2 null cells are defective in cell migration

Endocytic recycling of cell surface cargo receptors is essential for cell motility and migration [217, 218]. Copious studies have shown that perturbations on endocytic functions that result in reduced plasma membrane proteins, such as integrin and TfnR, lead to defects in cell migration [219-221]. To test whether Sac2 play a role in cell migration, we performed wound-healing assays. In response to monolayer wounding, the migration velocity of Sac2 null cells was significantly slower (~1.1 μm/hr) compared to wild type cells (2.6 μm/hr) (Figs. 22A and B). The defect in cell migration in Sac2 null cells was rescued to a rate of ~1.8 μm/hr by exogenous expression of GFP-Sac2 (Figs. 22A and B). Together, our data support a role of Sac2 in cell migration.

3.2.10 Structure of the homology Sac2 domain (hSac2)

Sac2 has a unique conserved hSac2 domain and our data suggest that the hSac2 domain contributes to its intracellular localization (Fig. 23). To understand the molecular basis, we
performed structural studies of this unique domain (aa. 593-760). The hSac2 domain contains approximately 120 amino acids of unknown function and structure. Sequence homology searches reveal homology sequence in proteins encoded by the transformation-related protein 63 regulated 1 (TPRG1) [222] and the tumor protein p63-regulated gene 1-like genes (TPRGL) [223]. The hSac2 domain was expressed, purified and crystallized. The hSac2 crystal diffracted to 2.7 Å and the structure was solved by single wavelength anomalous dispersion method (SAD) (Table 2, page 53). The hSac2 domain consists of a core of two perpendicularly apposed β-sheets with a C-terminal α-helix that seals the gap between two β-sheets (Fig. 23A). This core has a similar architecture to the pleckstrin homology (PH) domain. In addition to the core, hSac2 has an extra N-terminal α-helix and a short β-strand at the C-terminal end. PH domains are known to bind to phosphoinositides (PIs). However, hSac2 did not show any binding to any phospholipids in our liposome
Fig. 22. Sac2 deletion delays cell migration. (A) Wound healing assay of ΔSac2 cells. Dashed lines measure the distance of the wound. FITC images are shown to indicate transfected cells. (B) Quantification of the distance traveled (μm) after 48 hours as measured by the dashed lines. Data are from three replicate experiments (mean ± SEM., **p < 0.01).

Sedimentation assays (data not shown). Electrostatic surface potential calculation did not reveal any surface patch with significant positive electrostatic potentials, which are frequently observed in many lipid-binding modules. Our structure further revealed that the hSac2 domain may form a dimer (Fig. 23B). The dimer interface mainly consists of the β1 strand and α2 helix in each monomer and embeds several hydrophobic residues. The dimer is further stabilized by four pairs of main chain hydrogen bonds formed between the two β1 strands within each monomer. Furthermore, dimerization of hSac2 creates a continuous anti-parallel β-sheet with a large hydrophobic surface, which is covered by the N-terminal α1 helix in a swapped manner (Fig. 23B). To test whether Sac2 indeed forms a dimer and whether the hSac2 domain contributes to the Sac2 dimerization, co-immunoprecipitation experiment was performed using constructs containing the catalytic and hSac2 domains (aa. 1-798). Flag-tagged, and GFP-tagged Sac2 1-798 were cotransfected in HEK293T cells and cell lysates were immunoprecipitated with anti-Flag beads. Co-immunoprecipitated proteins were analyzed by Western blot probing for both Flag and GFP tags.
Flag-Sac2 1-798 co-precipitated a significant amount of GFP-Sac2 1-798 (Fig. 23C). However, the interaction was significantly reduced with GFP-Sac2 1-567 (lacking the hSac2 domain). Taken together, our structural data suggest that the hSac2 domain is at least partially responsible for the Sac2 dimerization and may contribute to the endosomal localization of Sac2 through potential protein-protein interactions but not by direct lipid binding.

Fig. 23. Crystal structure of the hSac2 domain of Sac2. (A) Ribbon diagram of the overall structure of the hSac2 domain. The hSac2 domain consists of a core of two perpendicularly apposed β-sheets (pink) with a C-terminal α-helix (blue). (B) Ribbon diagram of the hSac2 dimer. Inset: a zoom-in view showing the four pairs of main chain hydrogen bond formed between the two β1 strands within each monomer. (C) Western blot showing Flag-Sac2 1-798 coimmunoprecipitating with GFP-Sac2 1-798 but much reduced with GFP-Sac2 1-567.
Table 2. hSac2: Data collection and structural refinement statistics.

<table>
<thead>
<tr>
<th>A. Data collection statistics</th>
<th>P2₁2₁2₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>a = 61.77 Å, b = 82.85 Å, c = 157.70 Å, α = 90°, β = 90°, γ = 90°</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td>Selenium</td>
</tr>
<tr>
<td>Synchrotron beam lines</td>
<td>MCCHESS A1</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9789</td>
</tr>
<tr>
<td>Maximum resolution (Å)</td>
<td>2.62</td>
</tr>
<tr>
<td>Observed reflections</td>
<td>181,173</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>25,539</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.0(100.0)</td>
</tr>
<tr>
<td>&lt;I&gt;/&lt;σ&gt;²</td>
<td>29.0(4.13)</td>
</tr>
<tr>
<td>R_{sym}ᵃᵇ (%)</td>
<td>6.7(61.3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. Refinement statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
</tr>
<tr>
<td>R_{crys}/ R_{free} (%)</td>
</tr>
<tr>
<td>Rms bond length (Å)</td>
</tr>
<tr>
<td>Rms bond angles (°)</td>
</tr>
<tr>
<td>Ramachandran plot</td>
</tr>
<tr>
<td>Most favored/Additional (%)</td>
</tr>
<tr>
<td>Generous/Disallowed (%)</td>
</tr>
</tbody>
</table>

ᵃValues in parenthesis are for the highest resolution shell.
ᵇR_{sym} = ΣₜΣₗ[Iₜ(h)] - <I(h)>/ΣₜΣₗ[Iₜ(h)].
ᶜR_{crys} = Σ[F_{obs}−k|F_{cal}|]Σ|F_{obs}|. R_{free} was calculated for 5% of reflections randomly excluded from the refinement.

3.3 Discussion

It is well established that PI(4,5)P₂ is enriched in and serves as a marker for the plasma membrane [2]. PI(4,5)P₂ functions as a co-receptor at the plasma membrane to recruit a number of endocytic proteins including subunits of AP-2, AP180/CALM, epsin, and dynamin to trigger the formation of clathrin coated pits (CCPs) [224]. These PI(4,5)P₂ molecules are quickly degraded during the subsequent endocytic stages. Several PI phosphatases contribute to the removal of PI(4,5)P₂. For example, the PI phosphatase synaptojanin is recruited to CCPs and dephosphorylates PI(4,5)P₂ to facilitate the uncoating of CCPs [23] and the release of PI(4,5)P₂ binding adaptor proteins. The PI 5-phosphatase SHIP2 (SH2 domain-containing inositol 5-phosphatase) is localized to endocytic CCPs and controls the maturation of CCPs [225]. OCRL (oculocerebrorenal syndrome of Lowe), another inositol 5-phosphatase that acts on both PI(4,5)P₂ and
PI(3,4,5)P3, is associated with newly formed clathrin coated vesicles (CCVs) [17, 18]. By converting PI(4,5)P2 into PI(4)P during the endocytic pathway, PI 5-phosphatases may contribute to the establishment of a new PI identity of early endosomes to regulate downstream receptor trafficking and sorting. However, the enzymatic cascade following the actions of 5-phosphatases that eventually leads to the conversion of PI(4,5)P2 to PI(3)P at early endosomes is still not well understood. One plausible pathway is through the interplay of PI kinases and phosphatases. Recent data show that the class II PI-3 kinase C2α controls clathrin-mediated endocytosis by the phosphorylation of PI(4)P to PI(3,4)P2 [226]. It also has been shown that the PI 4-phosphatase, INPP4A, which specifically hydrolyzes PI(3,4)P2 to PI(3)P [134], contributes to the production of PI(3)P on early endosomes [227].

In this work, we report that Sac2 has an enzymatic activity towards PI(4)P. We provide evidence that Sac2 associates with early endocytic compartments, particularly with the endocytic recycling cargo Tfn and its receptor TfnR. Our data identify Sac2, a previously poorly

![Fig. 24. Model for Sac2 function during endosomal recycling.](image-url) The recycling process is dependent on the proper maintenance and conversion of PIs. Initial stage of endocytosis requires the hydrolysis of PI(4,5)P2 to PI(4)P via PI 5-phosphatases, such as OCRL. Sac2 is then recruited to endocytic intermediates to hydrolyse PI(4)P to phosphatidylinositol, which is a substrate for Vps34 to generate PI(3)P. The spatiotemporal conversion from PI(4,5)P2 to PI(3)P-enriched endosome is essential for endosomal maturation and subsequent recycling events.
characterized PI phosphatase, as a novel player in the endocytic pathway. These results uncover an alternative mechanism on the conversion of PI(4,5)P$_2$ during endocytosis and allow us to propose a working model (Fig. 24). In this model, 5-phosphatases, such as OCRL, may function upstream to hydrolyze PI(4,5)P$_2$ to PI(4)P. Sac2 is recruited to CCVs and early endosomes to relay the dephosphorylation reaction by hydrolyzing PI(4)P to phosphatidylinositol, which can be directly phosphorylated by hVps34 to generate PI(3)P [228]. Thus, through an enzymatic cascade of 5-phosphatases, Sac2, and PI-3-kinases, the initial PI(4,5)P$_2$ that is required for clathrin dependent endocytosis is converted to PI(3)P, which is a key lipid for subsequent endosomal functions. As predicted by this model, ablation of Sac2 function by CRISPR mediated genome editing causes transient accumulation of PI(4)P on the early endosomes and delays the recycling of Tf$n$ and integrin.

We also show that the catalytically inactive mutant of Sac2CS is localized to punctate structures in the cell, which are positive for endocytic markers and recycling cargos. The formation of Sac2CS positive puncta is at least partially due to the effect of substrate trapping and as a result, Sac2CS likely behaves as a dominant-negative enzyme. Our data show that Sac2CS has a predominant co-localization with Tf$n$R but with little to no overlaps with another endocytic cargo EGF. These data suggest that Sac2 is a PI(4)P phosphatase specifically targeted to recycling endosomes. It is still unclear how Sac2 is selectively recruited to endocytic compartments. However, since the C-terminal portion is required for the formation of Sac2CS puncta, it is likely that the hSac2 domain and/or proline-rich domain play a role in the recruitment of Sac2 to the endosomal intermediate structures via the association with unknown factors specifically localized on endosomes, but not the Golgi complex.

In conclusion, our data demonstrate that, in contrast to previous findings, Sac2 is a PI 4-phosphatase that specifically hydrolyzes PI(4)P. We further provide evidence that Sac2 participates in endocytic recycling processes by facilitating the lipid conversion of PI(4,5)P$_2$ along the vesicle trafficking route from CCPs to early endosomes (Fig. 24). Our data underscore the importance of spatial and temporal control of PIs along the endocytic pathway by a cascade of PI-
metabolizing enzymes.

3.4 Materials and Methods

3.4.1 Cloning, mutagenesis, and plasmids

PCR products for full-length Sac2 (aa. 1-1133) and different Sac2 truncations amplified from cDNA encoding human Sac2 were digested with BamHI and NotI restriction enzymes and inserted into pEGFP-C1 or pCmCherry in frame with the N-terminal GFP or mCherry tag, respectively. Flag-tagged Sac2 was generated using BglII and SalI restriction sites in the plasmid p3xFlag-CMV7.1. Single amino acid substitution (C458S) of Sac2 was introduced by site-directed mutagenesis. For bacteria expression, hSac2 (aa. 593-760) was cloned into a pET28a-based vector in frame with an N-terminal His-SUMO tag. The expression plasmids encoding clathrin light-chain, Rab4, Rab5, Rab7, Rab9, Rab11, M6PR, and GalT were obtained from Dr. Bill Brown (Cornell University) and Dr. Anthony Bretcher (Cornell University). The PI(4)P probes P4C was cloned from Legionella pneumophila SidC (614-744aa).

3.4.2 Cell culture and transfection

N2A and HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS (Cellgro), and 1% Pen/Strep (Cellgro) at 37°C in 5% CO₂ atmosphere. All plasmids were transfected using polyethyleneimine (PEI) reagent.

3.4.3 CRISPR

Guide sequence in the first exon of Sac2 (GCTGCAGGCCGCCGTCGCGC) was designed and chosen using Feng Zhang lab's target finder (http://crispr.mit.edu). The insert was cloned into pX330 (Addgene) via BbsI site. The pX330 plasmid containing the Sac2 guide insert was transfected with PEI reagent. After transfection for 24 h, N2A cells were counted and diluted in complete medium to a density of 100 cells. The cells were grown 5-7 days until the appearance of
colony formation. Individual colony was trypsinized and re-seeded to 24-well plate. A total of 30 clones were screened for endogenous Sac2 protein level by Western blot with anti-Sac2 antibody. Two Sac2 null clones were isolated.

3.4.4 Immunofluorescence

Cells were seeded on glass coverslips and fixed in 4% paraformaldehyde/PBS for 15 minutes. Cells were washed twice with PBS and permeabilized in PBS containing 0.05% saponin, and 3% BSA for 30 minutes. Cells were immunostained with proper primary and secondary antibodies. Confocal images were taken using a CSU-X spinning disk microscope (Intelligent Imaging Innovations) with a spherical aberration correction device, 63 × 1.4 NA objective on an inverted microscope (Leica), and acquired with a QuantEM EMCCD camera using Slidebook software (Intelligent Imaging Innovations). Anti-Lamp1 (BD Pharmingen), anti-TfnR (Invitrogen), and anti-GPP130 (Covance) antibodies were used at 1:1000, 1:500, and 1:500, respectively. Disruption of Golgi apparatus was carried out with 5 μg/ml of Brefeldin A (BFA) (Sigma) for 15 min at 37°C.

3.4.5 Quantitative analysis of colocalization and total cell fluorescence

Fiji (http://fiji.sc/About) program was used to process images. To quantify the degree of colocalization between Sac2CS and different organelle markers, 3 cells were selected, and for each cell, the entire cell region was examined. Pearson's correlation coefficient (PCC) was calculated using Fiji plug-in Coloc2 program on a single plane between the two indicated fluorescent signals [229]. ImageJ was used to measure cell fluorescence. P value was calculated using student's t-test (http://www.graphpad.com/quickcalcs/ttest1.cfm)

3.4.6 Time-lapse microscopy

For live cell imaging, cells were plated in a 35 mm petri dish, 14 mm glass bottom microwell. Cells were transfected as previous. Before imaging, medium was replaced with HEPES-buffered DMEM without phenol red (Lonza). All time-lapse movies were captured with a Zeiss
LSM 700 confocal microscope (Carl Zeiss) using a ×63 oil immersion objective (Carl Zeiss) at 37°C. Single section images were recorded at 5-10 sec interval and processed using the Zen software (Carl Zeiss). For imaging Tfn recycling, cells were first starved for 2 hours in DMEM-free medium, and prior to image capture, were replaced with complete medium containing 5 μg/ml Alexa647-Tfn at 37°C.

### 3.4.7 Immunoprecipitation

HEK293T cells grown to ~80% confluence were co-transfected with appropriate plasmids using PEI reagent. After ~18 h transfection, cells were lysed in RIPA buffer. Soluble fractions were collected by centrifugation at 18,000 rpm for 15 min at 4°C. Immunoprecipitates were prepared by 2 h binding at 4°C, followed by washing with 50 mM Tris-HCl pH 8.0 and 150 mM NaCl. To prepare Sac2 proteins for enzymatic assays, endogenous mouse Sac2 proteins from N2A cells were immunopurified using 1 μg of anti-Sac2 antibodies and 15 μl of protein A beads. GFP-Sac2 and Flag-Sac2 proteins were purified using GFP-Trap beads (Chromotek) and anti-Flag M2 affinity gel (Sigma), respectively. Final beads were resuspended in reaction buffer (50 mM Tris-HCl pH 8.0 and 150 mM NaCl). Endogenous integrin β1 was detected using rabbit anti-integrin β1 (#4706; Cell Signaling).

### 3.4.8 Enzymatic assays

All diC8-phosphoinositides were purchased from Cell Signals, Inc. All reactions were performed in a polystyrene 96-well plate for 20 min at 37°C with a total volume of 50 μl reaction mixture, which contains reaction buffer (50 mM Tris-HCl pH 8.0 and 150 mM NaCl), 1 nmol of lipids, and 2 μl of bead slurry of immunoprecipitated purified proteins. Phosphate release was measured at OD620 absorbance with the addition of malachite green reagent as described by Maehama et al. [230].

58
3.4.9 Transferrin and integrin recycling assay

For immunostaining, cells were grown on glass coverslips and starved for 30 min in DMEM-free medium at 37°C and then labeled with 10 μg/ml of Alexa594-Tfn for 30 min in 1X HBSS, 20 mM HEPES, and 2 mg/ml BSA on ice for 30 min. After wash with cold PBS, cells were chased for different time points in complete medium containing 50-fold excessive of unlabeled-Tfn and 100 μM deferoxamine to prevent reinternalization. At the end of each time point, cells were washed with 0.1 M glycine, pH 3.5 for 1 min, once with PBS, and then fixed in 4% paraformaldehyde (PFA). The coverslips were washed twice in PBS and then placed on a microscope slide with Fluoromount-G for imaging.

For the antibody-induced recycling assay, cells were incubated with 5 μg/ml of rat anti-integrin β1 (FCMAB269F; Millipore) in complete medium for 1 h at 37°C. The cells were first washed with PBS, and treated with 0.1 M glycine, pH 2.5, twice for 1 min to remove surface-bound antibody. After additional wash with PBS, cells were returned to 37°C in complete medium to chased internalized antibody back to the cell surface. For the analysis of surface receptor level, cells were washed in PBS, fixed in 4% PFA for 20 min, blocked in 1% BSA in PBS for 15 min, and incubated with anti-rat Alexa Fluor 488. For the analysis of internal receptor level, cells were treated with 0.1 M glycine, pH 2.5, twice for 1 min to remove recycled surface antibody-receptor complex. Cells were fixed in 4% PFA, permeabilized in 0.1% saponin for 30 min, and incubated with the secondary antibody anti-rat Alexa Fluor 488.

3.4.10 Flow cytometry

N2A cells were detached using trypsin and serum-starved in suspension for 30 min at 37°C in DMEM. For labeling, cells were incubated with 10 μg/ml of Alexa647-Tfn in buffer solution containing 1X HBSS, 20 mM HEPES, and 2 mg/ml BSA at 4°C for 30 minutes. Unbound transferrins were removed by washing the cells twice with 1X PBS. Cells were then supplemented in complete medium with 50 μg/ml of unlabeled Tfn and 100 μM deferoxamine and transferred to 37°C. After 5 min of incubation, cells were sampled at time 0, 10, 20 and 40 min. At each time
point, Cells were washed in 0.1 M glycine pH 3.5 followed by fixation in 2% PFA for 15 minutes. Fixed cells were resuspended in PBS before subjecting to flow cytometry analysis. 5000 cells were analyzed by measuring Alexa647 fluorescence. Time points are normalized time 0 and plotted for cell-associated Alexa647 signal intensity.

For detection of surface integrin β1, cells were detached with 10 mM EDTA in PBS for 20 min at 37°C, followed by washing with PBS. The resuspended cells were incubated in 50 μl of diluted anti-integrin β1 (1:15) (FCMAB269F; Millipore) in PBS containing 2% FBS for 30 min on ice. Cells were washed 1X in PBS and then incubated with anti-Rat Alexa Fluor 488 (1:100) in PBS containing 10% FBS for 30 min on ice. Cells were wash again with PBS and fixed in 2% paraformaldehyde for 15 min. 10,000 cells were analyzed on an Accuri flow cytometer.

3.4.11 Biotin surface labeling

N2A cells were starved in DMEM-free medium for 2 hours at 37°C in 5% CO2 incubator. Cells were then washed twice with 1X HBSS buffer prior to labeling with 1 mg/ml of Sulfo-NHS-SS-Biotin (ProteoChem) in 1X HBSS for 30 minutes on ice. Cells were washed twice with 1X HBSS to remove excess biotin and subsequently harvested in 1X HBSS containing 1% triton and protease inhibitor. Following centrifugation, an aliquot was collected as input for total protein levels, and the lysate was incubated with streptavidin resin (GenScript) at 4°C for 2 hours. The resin was then washed 3 times in lysis buffer followed by addition of 1X SDS loading buffer. Western blot was carried out and probed with mouse anti-TfnR and rabbit anti-Sac2 antibodies.

3.4.12 Scratch wound healing assay

Wild type N2A cells and ΔSac2 cells were transfected with the corresponding plasmid. Next day, transfected cells were seeded in triplicate in a 96-well plate in a density that after 24 hours, 90% confluence is reached as a monolayer. Wound was applied by scratching the monolayer with a 10 μl pipette tip across the center of the well. Each well was washed 1X with PBS to remove the detached cells and replaced with fresh medium. Images were captured at 0 h, 24 h and
48 h, with an ImageXpressMICRO microscope (Molecular Devices) using a 4X objective transmitted light, and FITC filter set to identify transfected cells. Experiment was performed twice in triplicate (n = 6).

### 3.4.13 Protein expression and purification

The hSac2 domain was expressed as His-SUMO fusion in *E. coli* and was purified by Co\(^{2+}\) resins. The His-SUMO tagged was cleaved by sumo protease Ulp1 and the protein was further purified by gel filtration and protein was concentrated to 10 mg/ml in 20 mM Tris (pH =7.8) and 20 mM NaCl for crystallization. For Sac2 full-length protein expression, *Tni* insect cells were seeded in a 500 ml culture flask containing 200 ml of culture media recombinant proteins were purified with GST sepharose resins.

### 3.4.14 Crystallization

Crystals were grown at room temperature by the hanging-drop vapor diffusion method by mixing 1 µl of protein (11 mg/ml) with an equal volume of reservoir solution containing 0.2 M ammonium sulfate, 0.1 M Tris-HCl pH 8.0, and 20% PEG3350. Rod-shaped crystals were formed within 2-3 days.

### 3.4.15 Data collection, processing, structure determination and refinement

Diffraction data sets for selenomethionine protein crystals were collected at the Cornell high energy synchrotron source. All data sets were indexed, integrated and scaled with HKL-2000. Selenomethionines were identified in the crystal using the program HKL2MAP [231]. Iterative cycles of model building and refinement were carried out with the program COOT [232] and the refmac5 program [233] in the CCP4 suite [234] to complete the final model.
PART II: PI-phosphatase and PI-binding effectors encoded by Legionella pneumophila

Chapter 4: Structural Basis for substrate recognition by a unique Legionella phosphoinositide phosphatase


Abstract. Legionella pneumophila is an opportunistic intracellular pathogen that causes sporadic and epidemic cases of Legionnaires’ disease. Emerging data suggest that Legionella infection involves the subversion of host phosphoinositide (PI) metabolism. However, how this bacterium actively manipulates PI lipids to benefit its infection is still an enigma. Here we report that the L. pneumophila virulence factor SidF is a phosphatidylinositol phosphate 3-phosphatase that specifically hydrolyzes the D3 phosphate of PI(3,4)P2 and PI(3,4,5)P3. This activity is necessary for the anchoring of PI(4)P-binding effectors to bacterial phagosomes. Crystal structures of SidF and its complex with its substrate PI(3,4)P2 reveal striking conformational rearrangement of residues at the catalytic site to form a cationic pocket that specifically accommodates the D4 phosphate group of the substrate. Thus, our findings unveil a novel Legionella PI phosphatase essential for the establishment of lipid identity of bacterial phagosomes.

4.1 Introduction

The Legionella genus is mainly constituted by environmental bacteria. Several species, in particular Legionella pneumophila and Legionella longbeachae, are pathogenic to humans [235-237]. Because of the development of artificial water systems, such as air conditioning, Legionnaires’ disease has emerged as a significant health threat in modern societies [238]. Inhalation of L. pneumophila in contaminated aerosols allows the pathogen to reach the alveoli of the lung, where they can be phagocytosed by host macrophages. Once engulfed by macrophages,
*L. pneumophila* delivers a large array of effector proteins into the host cell through a specialized secretion system called defective organelle trafficking (Dot)/intracellular multiplication (Icm) type IV secretion system (T4SS) [239, 240]. The translocated Dot/Icm substrates hijack host cellular processes, particularly the membrane trafficking pathways to bypass the default phagosome maturation pathway. In fact, several Dot/Icm substrates mediate the recruitment of secretory vesicles derived from endoplasmic reticulum to establish a replication-permissive compartment called the *Legionella*-containing vacuole (LCV) [241, 242]. Because membrane trafficking is extensively regulated by phosphoinositides (PIs), studies on how host cell PI signaling and metabolism pathways are exploited by intracellular bacterial pathogen have recently been placed on the center of focus.

Interference of the temporal and spatial distribution of intracellular PIs often leads to abnormal cellular functions, which has been capitalized by virulent invaders [144, 145]. Bacterial pathogens have evolved a variety of mechanisms to subvert PI metabolism in host cells. For examples, *Shigella flexneri*, the causative agent of human dysentery, modifies PI metabolism in host cells to favor its internalization through the PI-4-phosphatase activity of the virulent factor IpgD [243]. *Salmonella typhimurium*, which is responsible for most food-borne gastroenteritis [244], delivers the PI phosphatase SigD/SopB into the host. By hydrolyzing PI(3,4,5)P₃, SopB contributes to the localized membrane ruffling that leads to bacterial internalization in nonphagocytic cells [245, 246].

Modulation of host PI metabolism by *Legionella* is important for the establishment of the LCV within which the bacterium replicates [144, 145, 241]. It has been suggested that PI(4) P is enriched on the LCV membrane, which among other functions, anchors effectors such as SidM/DrrA, SidC, and SdcA to promote the recruitment and fusion of the endoplasmic reticulum derived vesicles with the LCV [247, 248]. In the amoebae host *Dictyostelium discoideum*, the *Legionella* effector protein LpnE appears to recruit the host PI-5-phosphatase OCRL to the LCV, leading to restriction of intracellular bacterial growth [249]. Although significant progress has been made toward our understanding of the roles of PI metabolism in bacterial pathogenesis, our
knowledge on how bacterial pathogens actively exploit host cell PI metabolism and signaling is still in its infancy. Currently, no virulence factors that directly modify host PIs have been identified in *Legionella*. Hence, we performed bioinformatic and biochemical studies on *Legionella* effector proteins. We identified the *Legionella* effector SidF as a unique PI phosphatase that specifically hydrolyzes PI(3,4)P₂ and PI(3,4,5)P₃. In agreement with this enzymatic activity, we found that deletion of SidF results in the reduced recruitment of effector proteins that anchor on the LCV via binding to PI(4)P. We further report the crystal structures of SidF and its complex with a bound short-chain (dibutanoyl) derivative of PI(3,4)P₂. The structures show that the conserved CX₅R catalytic motif is located in a large cationic groove. Remarkably, structural analysis reveals key features responsible for substrate specificity. Residue His233, located in a loop region between α6 and β5, translocates approximately 20 Å to the catalytic site and together, with a serine and three lysine residues, forms a pocket that specifically accommodates the D₄ phosphate group of the substrate. Our findings uncover a family of bacterial PI phosphatases and establish a role of SidF in the maintenance of the lipid composition of LCV.

4.2 Results

4.2.1 *Legionella* effector SidF is a phosphoinositide 3-phosphatase

To search for *Legionella* effector proteins that may directly modify host PIs, we used a sequence pattern based method to retrieve proteins containing the CX₅R motif, a signature sequence of the catalytic residues in PI phosphatases [250], in the genome of *L. pneumophila* strain Philadelphia 1. More than 400 hypothetical proteins were found to possess this motif. Among these candidates, 29 proteins have been identified as substrates of the Dot/Icm transporter [251]. Some of these proteins were then expressed, purified, and examined for in vitro PI phosphatase activities by a malachite green-based assay [230]. SidF, a Dot/Icm substrate shown to be involved in modulating host cell death [252], was found to possess PI phosphatase activity (Fig. 1A). Mutation of the catalytic cysteine to serine (C645S) abolishes SidF PI phosphatase activity (Fig. 1A).
SidF is comprised of 912 residues with a large N-terminal domain (1–760) of unknown function and two predicted transmembrane motifs at the C terminus. Ectopically expressed GFP-SidF localizes to the ER membrane in mammalian cells (Fig. 2). Interestingly, deletion of the C-terminal portion of SidF including the two transmembrane motifs changes its localization to the cell periphery (Fig. 2), suggesting that the N-terminal cytosolic portion of SidF has the propensity to associate with membranes (discussed below). In agreement with the prediction of SidF as a membrane protein, endogenous SidF delivered into the host cell associates with the LCV membrane during Legionella infection (data not shown). These observations imply a role of SidF in controlling the lipid composition of the LCV membrane.

Although SidF is a membrane protein, deletion of the putative transmembrane domains did not affect its enzymatic activity and, thus, the N-terminal portion (1–760) of SidF was used in our in vitro activity assays. SidF exhibited phosphatase activities against PI(3,4)P₂ and PI(3,4,5)P₃ with a preference for PI(3,4)P₂ (Figs. 1A and B). To further investigate the enzymatic function of SidF, a fluorescent phosphoinositide-based TLC method [253] was used to determine the specific phosphate group hydrolyzed by SidF (Figs. 1C–F). SidF hydrolyzed PI(3,4)P₂ to a single phosphorylated PI product (Fig. 1C, lane 5), and this product could not be further digested by the specific PI-3-phosphatase MTM [120]. However, it could be hydrolyzed to phosphatidylinositol (PtdIns) by the Sac domain of yeast Sac1, a phosphatase that hydrolyzes both PI(3)P and PI(4)P [26] (Fig. 1C, lanes 6 and 7). This result suggests that SidF can specifically dephosphorylate PI(3,4)P₂ at the D3 position of the inositol ring. Similarly, when PI(3,4,5)P₃ was used as the substrate, double phosphorylated PI species were generated and this species could be further hydrolyzed by OCRL, a PI-5-phosphatase that hydrolyzes the D5 phosphate of PI(4,5)P₂ and PI(3,4,5)P₃ [18] (Fig. 1E, lanes 5 and 6). Therefore, these results demonstrate that SidF is a PI-3-phosphatase that specifically hydrolyzes PI(3,4)P₂ and PI(3,4,5)P₃ to PI(4)P and PI(4,5)P₂, respectively.

4.2.2 SidF facilitates the anchoring of effector proteins to bacterial phagosomes

The identification of a PI phosphatase from L. pneumophila has addressed the long-stand
Fig. 1. *Legionella* effector SidF is a phosphoinositide phosphatase. (A) Phosphoinositide substrate specificity of purified wild-type and C645S mutant SidF as determined by the malachite green assay (green color indicates the release of free phosphate). PI(3,4)P₂ and PI(3,4,5)P₃ are the preferred substrates. (B) Quantification of the amount of released phosphates. Data are from three replicate experiments (mean ± SEM). (C) Determination of SidF substrate specificity by fluorescent lipids. Phosphatase reactions were carried out with di-C8-Bodipy-FL-PI(3,4)P₂ and PI phosphatases as labeled. In lane 6 and 7, the reactions were first carried out with SidF, and the products were further hydrolyzed by the addition of a specific 3-phosphatase MTM (lane 6) or Sac1 (lane 7), that hydrolyzes both PI(3)P and PI(4)P. (D) Schematic diagram to illustrate the enzymatic reactions shown in C. (E) TLC results of the hydrolysis of PI(3,4,5)P₃ by SidF. In lane 6, the reaction was first carried out with SidF, and the products were further hydrolyzed by the addition of OCRL, a 5-phosphatase that hydrolyzes PI(4,5)P₂. (F) Schematic illustration of the reactions in E.

The hypothesis that this bacterium employs its own PI phosphatase(s) to actively modify the PI composition on bacterial phagosomes. Indeed, SidF converts PI(3,4)P₂ and PI(3,4,5)P₃, which are two PI species generated on phagosomes at early stages of phagocytosis [254, 255], to PI(4)P and PI(4,5)P₂, respectively. Intriguingly, PI(4,5)P₂ may be further converted to PI(4)P by the host PI-5-phosphatase OCRL [249]. Thus, one plausible hypothesis is that SidF plays a role in the establishment of LCVs with a lipid composition enriched in PI(4)P. It has been suggested that PI(4)P provides specific anchors on the LCV for *Legionella* effectors, such as SidC/SdcA (18) and SidM(DrrA) [256]. Through binding to PI(4)P, these effectors presumably facilitate the fusion of ER derived vesicles with the LCV. Hence, we examined the role of SidF in the association of SidC with LCVs. Strikingly, compared with the wild-type strain, the association of SidC with phagosomes formed by the SidF deletion mutant was significantly reduced, and such defect can be almost fully restored by expressing wild-type SidF from a plasmid but not the catalytically inactive
SidF C645S mutation (Figs. 3A and B). Instead, our data suggest that SidF facilitates the anchoring of PI(4)P binding effectors, such as SidC to the LCV membrane through the generation of PI(4)P.

Fig. 2. Intracellular localization of transfected SidF in mammalian cells. (A–C), Exogenously overexpressed EGFP-SidF colocalizes with ER marker, RFP-KDEL, in Cos7 cells. (D–F). The cytosolic portion (1–760) of SidF localizes to the plasma membrane when overexpressed in Cos7 cells. (Scale bars: 10 μm).

4.2.3 Crystal structure of SidF

To understand the molecular mechanism of the catalytic function of SidF, the cytosolic portion (residues 1–760) of SidF (Fig. 4) was purified and crystallized. The crystal structure of SidF was determined by Single Isomorphous Replacement with Anomalous Scattering (SIRAS) method (Table 1). The crystal structure revealed that the entire N-terminal portion of SidF is folded into one large single domain (Fig. 5). This domain is comprised of 19 α-helices with lengths ranging from 6 to 60 amino acids surrounding a 10-pleated β-sheet core. The overall shape of the phosphatase domain resembles a cowboy hat. The bottom of the hat has a flat surface with two protrusion loops (Figs. 5A and C colored in gold and highlighted with a square). Interestingly,
Fig. 3. SidF is required for anchoring SidC to the bacterial phagosomes. (A) Representative images of SidC immuno-staining of mouse bone marrow-derived macrophages infected with indicated *L. pneumophila* strains at an MOI of 1 for 1 h. (B) Quantitation of SidC positive bacterial phagosomes. Phagosomes positive for SidC staining are normalized against total phagosomes. Data shown are from two independent experiments performed in triplicate in which at least 100 phagosomes were scored per coverslip. **P < 0.01, paired Student t test. WT: *L. pneumophila* Philadelphia-1 wild type strain Lp02; dotA-: the type IV secretion system defective strain Lp03(dotA-); ΔSidF: the sidF deletion mutant Lp02 strain; ΔSidF(pSidF): the sidF deletion Lp02 strain complemented with a plasmid expressing SidF and ΔSidF(pSidF C/S): the sidF deletion Lp02 strain complemented with a plasmid expressing SidF C/S catalytically dead mutant.

these two loops are mainly comprised of hydrophobic residues (Fig. 5). The catalytic CX\(_2\)R motif (shown in spheres) resides in the middle of a groove that nearly bisects the bottom surface (Figs. 5B and D). Like other PI phosphatases (32), this groove is enriched with positively charged residues, which contribute a highly basic character to the groove. These architectural arrangements suggest a molecular mechanism for PI hydrolysis by SidF at the membrane interface. The flat bottom surface of SidF may associate with the membrane with the two hydrophobic loops penetrating into the bilayer. When scooting on the membrane surface, the overall positive charge in the groove may facilitate the loading of negatively charged PI lipids into the catalytic site. Structural homology search by the DALI program [257] indicated that SidF has no overall structural homologs in the PDB database; however, the catalytic core of SidF bears similar topological fold with other PI phosphatases. Among these PI phosphatases, Sac1 has a highest Z-score of 8.6 and a rmsd of 4.1 Å for 240 aligned residues. The other phosphatase PTEN has a Z-score of 7.3 and an rmsd of 2.7 Å for 111 aligned residues (further discussed below).
Fig. 4. Multiple sequence alignment of SidF: The sequences corresponding to the SidF 1–760 region from different *Legionella* species were aligned by MutiAlin and colored by ESPript 2.2. Secondary elements are drawn above the alignment. The catalytic residues are marked with red stars. Residues important for D4 phosphate recognition are highlighted with blue triangles. Two hydrophobic loops are marked with gold circles. The flexible loop containing His233 is marked with a green box. Entrez database accession numbers are as follows: SidF_Phila, gi: 52842790; SidF_130b, gi: 307611464; SidF_Lens, gi: 54295421; SidF_Corby, gi: 148255306; SidF_Alcoy, gi: 296108226; and SidF_Paris, gi: 54298573.
4.2.4 Structure of SidF-PI(3,4)P$_2$ complex

To address how SidF can specifically hydrolyze PI(3,4)P$_2$ and PI(3,4,5)P$_3$, catalytically inactive (C645S) recombinant SidF (1–760) proteins were prepared. The mutant proteins were mixed with substrate diC$_4$-PI(3,4)P$_2$ [a dibutanoyl derivative of PI(3,4)P$_2$] and screened for crystals of the protein–lipid complex. Diffraction data were processed and used in the refinement against the apo protein structure. After the first round of refinement, the difference Fourier electron density clearly revealed a well-defined substrate molecule diC$_4$-PI(3,4)P$_2$ bound at the catalytic site (Figs. 6A and B). Remarkably, the binding of PI(3,4)P$_2$ to SidF induces a large conformational change of the loop connecting α6 and β5. His233 on this loop shifts approximately 20 Å to the substrate binding site (Fig. 6C), where His233, together with Ser647 and three lysines (Lys646, Lys717, and Lys740), forms a highly cationic pocket that selectively accommodates the D4 phosphate group of the substrate (Fig. 6D). The phosphate group at the D3 position of PI(3,4)P$_2$ is also heavily involved in the interaction with SidF. The D3 phosphate group...
subject to hydrolysis forms intensive hydrogen bonds and electrostatic interactions with five main chain amide groups of the catalytic CX₅R loop and the guanidinium group of the CX₅R arginine Arg651 (Fig. 7). The D1 phosphate group of the substrate is less involved in the interaction with the enzyme. It forms electrostatic interactions with Arg651 and one hydrogen bond with the amido group of Asn419 (Fig. 7). It is also notable that the diacylglycerol moiety of the substrate molecule makes significant contact with several hydrophobic residues, including Trp420 and Phe421. These hydrophobic residues are located within the two hydrophobic protrusion loops
Fig. 6. Substrate recognition by SidF. (A) Difference electron density map (Fo−Fc at 3σ, green mesh) calculated near the catalytic site. The substrate diC4-PI(3,4)P₂ molecule shown in sticks fits nicely in the electron density. (B) A view of the SidF–substrate complex represented in surface. The substrate colored in yellow binds deeply in the positively charged groove at the catalytic site. (C) The binding of substrate induces a large conformation change of a loop containing residue His233. The apo structure is shown in pink with the His233 loop shown in red. The complex structure is colored in cyan with the corresponding loop in blue. The diC4-PI(3,4)P₂ molecule is shown in sticks and enveloped in a yellow surface. His233 forms a hydrogen bond with the D4 phosphate group of the substrate. (D) Specific recognition of the D4 phosphate group of the substrate. Five residues (Lys717, Lys740, Lys466, Ser647, and His233) form a basic pocket that holds the D4 phosphate through an intensive hydrogen bond network and electrostatic interactions. Hydrogen bonds are indicated by dashed lines, and the distance in Å is labeled.

and may penetrate into the lipid bilayer during hydrolysis.

SidF also hydrolyzes PI(3,4,5)P₃ but with less efficiency compared with PI(3,4)P₂. Based on our complex structure, it can be predicted that the D5 phosphate will be exposed to solvent and no significant interactions can be formed with the protein. Instead, the presence of a glutamate residue (Glu370) near the D5 position may repel the binding of D5 phosphate of
PI(3,4,5)P₃, which may explain the lower activity of SidF against PI(3,4,5)P₃ (Fig. 1A).

4.2.5 Structure comparison of SidF to other phosphatases

Despite the lack of detectable overall sequence and structural fold similarity, comparison of the structure of SidF with other CX₅R motif-based protein and lipid phosphatases reveals that the topology of the catalytic core of SidF is conserved with other phosphatases. All CX₅R motif-based phosphatases share a common architecture of a central β-sheet consisting of four parallel β strands and one α helix. The peptide containing the catalytic CX₅R motif connects the carboxyl end of one of the β strands with the amino terminus of the α helix in the structural core (Figs. 8A–C) [89, 91, 258, 259]. It is interesting to note that the electric dipole of this α helix contributes net-positive electrostatic potentials to its amino terminus, where the catalytic site resides. Hence, this structural organization may facilitate the docking of the negatively charged phosphate group of the substrate.
Structural comparison of SidF with other PI phosphatases, such as the myotubularin phosphatases [92, 93] and the tumor suppressor PTEN [91], further reveals the difference in the overall shape of the active site pocket, which is one of the key determinants for substrate specificity. In SidF, the active site pocket is deep and wide at the bottom to fit the two adjacent phosphate groups at D3 and D4 position (Fig. 8D). However, in MTMR2, which hydrolyzes both PI(3)P and PI(3,5)P₂, the width of the active site pocket is much narrower at the bottom region that limits the hydrolysis of substrates with two consecutive phosphate groups attached. Steric clashes between the protein and the lipid prohibit the binding of PI molecules phosphorylated at the D4 position.

Fig. 8. Structure comparison of SidF with other phosphoinositide phosphatases. (A) Ribbon diagram showing the topology of the SidF active site. This conserved structural core contains four parallel β strands and one α helix (colored in cyan). The catalytic CX₅R motif (colored in red and the conserved cysteine and arginine residues shown in sticks) is located between the C terminus of one of β strands and the first turn of the α helix. (B and C) The active site topology of Sac1 and protein tyrosine phosphatase 1b (PTB1b). (D) Slice of the active site surface showing the docking of substrate diC₄-PI(3,4)P₂ at the active site. diC₄-PI(3,4)P₂ is shown in sticks. In this particular slice, Trp420 can be seen to form hydrophobic interactions with the lipid tails of the substrate. (E) Slice of the active site surface of MTMR2 with a diC₄-PI(3,5)P₂ molecule shown in sticks bound at the active site. (F) Slice of the active site surface of PTEN. The active site pocket is occupied by a tartrate molecule. Comparison of these three PI phosphatases suggests the overall shape of the catalytic site pocket may also play a role in substrate selectivity.
to MTMR2 (Fig. 8E). The active site in PTEN is much wider, consistent with the larger size of its preferred substrate PI(3,4,5)P3 (Fig. 8F).

4.3 Discussion

Although PI phosphatases have been reported from other bacterial pathogens, such as the PI-4-phosphatases IpgD from *Shigella flexneri* [243], SigD/SopB from *S. typhimurium* [245, 246], no such enzymes have been reported from *Legionella* species. Our discovery of a PI phosphatase in *L. pneumophila* not only establishes an archetypal family of PI phosphatase, but also opens a new avenue toward the understanding of the roles of PI signaling and metabolism in *L. pneumophila* infection.

Our results further demonstrated a role of SidF in maintaining the PI composition of LCV. By hydrolyzing PI(3,4)P2 and PI(3,4,5)P3 that are generated on phagosomes at their early stage [254, 255], SidF (possibly in coordination with other possible PI metabolizing enzymes either from host or bacterium) converts the LCV into a PI(4)P enriched organelle. As a result, the lipid composition of LCV would resemble that of the cis-Golgi compartment and may render the LCV a better recipient site for secretory vesicles originating from ER. PI(4)P enrichment at the LCV also provides specific membrane anchors for PI(4)P binding effector proteins. Deletion of SidF from *L. pneumophila* significantly reduced the anchoring of PI(4)P binding effectors such as SidC on LCV compartments (Fig. 3). The residual association of SidC with the LCVs may result from additional bacterial effectors with activity similar to that of SidF, or host proteins involved in the production of PI(4)P, or a combination of both.

The hydrolysis of PI(3,4)P2 and PI(3,4,5)P3 by SidF at the D3 position may also play a role in preventing the conversion of these two lipids into PI(3)P by other 4- or 5-phosphatases, such as SHIP1 and INPP4A during endocytic processes. It has been shown that the accumulation of PI(3)P at the phagosome membrane facilitates the fusion of phagosome with endosomes and lysosomes [260, 261] and promotes the degradation of phagosomal contents. However, the precise lipid composition of LCV is not known. A thorough understanding of the role of SidF in
controlling the PI composition and the fate of LCV would require lipid composition analysis on purified LCVs under a variety of genetic backgrounds and infection stages.

SidF has been shown to play a role in conferring cell death resistance in infected macrophages by interacting through its C-terminal portion with BNIP3 and Bcl-rambo, two prodeath members of the Bcl2 protein family [252]. Here, we further show that the N-terminal part of SidF is a specific PI-3-phosphatase. Given the pleiotropic effects of PI(3,4)P2 and PI(3,4,5)P3 on cellular physiology, including the cell surviving process [2-4], it is clear that SidF is a multifunctional protein that may be involved in diverse biological processes of the host by distinct mechanisms. Future investigations are required to elucidate the role of SidF in host PI3K signaling pathways and their potential interplays with the host cell-death process under infection conditions.

Our findings led us to propose a functional model of SidF on the bacterial phagosome (Fig. 9). In this model, SidF is delivered into the host through the Dot/Icm complex and anchors on the LCV membrane via two C-terminal transmembrane domains. The flat surface of the catalytic domain is interfaced with the LCV membrane, and the two hydrophobic loops are inserted into the hydrophobic lipid bilayer. SidF then hydrolyzes PI(3,4)P2 and PI(3,4,5)P3, which may cause the accumulation of PI(4)P on LCV and the subsequent recruitment of other effectors that anchor on the LCV membrane through the binding to PI(4)P. By controlling the lipid composition of LCV, SidF may facilitate the programming of LCV to an amenable niche for bacterial growth to escape from the default degradative phagolysosomal pathway.

4.4 Materials and Methods

4.4.1 Cloning and mutagenesis

PCR products for SidF amplified from L. pneumophila genomic DNA was digested and inserted into a pET28a-based vector in frame with an N-terminal His-SUMO tag. All constructs were confirmed by DNA sequencing. Point mutations were generated by site directed mutagenesis.
Fig 9. Functional model of SidF. The *Legionella* effector protein SidF is a PI-3-phosphatase that specifically hydrolyzes PI(3,4)P₂ and PI(3,4,5)P₃. By the action of SidF and/or other unknown mechanisms, a PI(4)P enriched LCV membrane is established. PI(4)P enrichment may allow specific anchoring of Dot/Icm effectors, such as SidC, to the LCV, thus facilitates the recruitment and fusion of ER-derived vesicles with the LCV. (Inset) Molecular mechanisms of SidF. SidF anchors on the LCV membrane through its C-terminal double transmembrane motifs. The flat surface of the cytosolic domain of SidF interfaces with the LCV membrane and the two hydrophobic loops protruding out from the flat surface penetrate into the bilayer. The basic charges in the catalytic groove facilitate the loading of substrate into the catalytic site.
4.4.2 Protein expression and purification

For protein expression, *Escherichia coli* Rosetta strains harboring the expression plasmids were grown in Luria–Bertani medium supplemented with 50 μg/mL kanamycin to midlog 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 at pH 8.0, 200 mM NaCl, and protease inhibitor mixture (Roche) and were lysed by sonication. Soluble fractions were collected by centrifugation at 40,000 × g for 20 min at 4°C and incubated with cobalt resins (Clontech) 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 SidF from the His-SUMO tag. Eluted protein samples were further purified by FPLC size exclusion chromatography. The peak corresponding to SidF was pooled and concentrated to 10 mg/mL in a buffer containing 20 mM Tris at pH 7.4 and 150 mM NaCl. To express recombinant OCRL proteins, OCRL gene was cloned into a pFAST-based vector in frame with an N-terminal His-GST tag. Baculovirus was generated by using standard protocols (Invitrogen). Tni cells at 2 × 10⁶ cells per mL were infected with baculovirus for 2 d. Cells were harvested and lysed as described above. Recombinant OCRL proteins were affinity purified with glutathione Sepharose resins (GE).

4.4.3 Bioinformatic Search

The protein sequences of the *Legionella pneumophila* strain Philadelphia 1 were downloaded from NCBI. A perl script was constructed to search the complied sequence file for proteins that contains the CX5R signature motif.

4.4.4 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 Tris-HCl pH8.0, 10% PEG3350, and 10% ethylene glycol. Rod-shaped crystals were formed within 2-3 days. For phase determination, protein crystals were soaked in cryoprotectant (0.1 M
Tris-HCl pH 8.0, 10% PEG3350, and 25% (v/v) ethylene glycol) with the addition of 10 mM ethylmercury chloride (kind gift from Dr. Steve Ealick at Cornell University) for 2 h at room temperature. For SidF-substrate complex crystal formation, purified SidF C645S mutant proteins (10 mg/ml) was mixed with a final concentration of 0.3 mM diC4-PI(3,4)P2 (Echelon Biosciences Inc.). The protein-substrate complex was crystallized in a similar condition as described above.

4.4.5 Data collection and processing

Diffraction data sets for native and mercury derivative crystals were collected at the Cornell synchrotron light source, MacCHESS beam line A1. Data set for SidF-substrate complex crystal was collected at Brookhaven National Laboratory, X4C beamline. All data sets were indexed, integrated and scaled with HKL-2000 [262]. The crystals belong to space group P2₁2₁2₁ with a = 71.23 Å; b = 114.83 Å; c = 124.37 Å; α = β = γ = 90° (Table 1). The calculated Matthews coefficient Vₘ=3.03 and with 59.3% of solvent in the crystal and one protein molecule in an asymmetric unit [263].

4.4.6 Structure determination and refinement

Three mercury sites corresponding to residues C206, C635 and C602 were identified in the crystal using the program HKL2MAP. 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 COOT [232]. Iterative cycles of model building and refinement were carried out with the with refmac5 [233] in the CCP4 suite [234] to complete the final model.

4.4.7 Enzymatic assays

All diC₈-phosphoinositides were purchased from Cell Signals, Inc. All reactions were performed in a polystyrene 96-well plate for 20 min at 37°C with a total volume of 50 μl reaction mixture, which contains reaction buffer (50 mM Tris-HCl pH 8.0 and 150 mM NaCl), 1 nmol of
lipids, and 0.1 μg of purified enzymes. Phosphate release was measured at OD$_{620}$ absorbance with the addition of malachite green reagent as described by Maehama et al [230].

### 4.4.8 Thin-layered chromatography

The green fluorophore, Bodipy*-FL labeled diC6 phosphoinositides were purchased from Echelon Research Laboratories (Salt Lake City, UT). Most reactions were conducted in 20 μl of buffer containing 50 mM ammonium carbonate (pH 8.0) and 2 mM dithiothreitol, and with 1 μg of lipid substrates and 0.1 μg of purified proteins for 20 min at 37°C. For reactions containing two enzymes, the second enzyme was added after the completion of the first SidF catalyzed reaction and was incubated for additional 20 min. Reaction products were dried in a Speed-Vac for 30 min at 45°C. The dried pellets were resuspended in 10 μl of methanol/isopropanol/acetic acid (5/5/2) and spotted onto a TLC Silica gel 60 F$_{254}$ (EMD) that was pretreated by soaking in methanol/water (3:2) containing 1% potassium oxalate and then 1 h drying in 10 μl of methanol/isopropanol/acetic acid (5/5/2) and spotted onto a TLC Silica gel 60 F$_{254}$ (EMD) that was pretreated by soaking in methanol/water (3:2) containing 1% potassium oxalate and then 1 h drying in a 65°C oven. The TLC plates were developed in a solvent system consisting of 1-propanol/2 M acetic acid (65%:35%). Fluorescent lipids were visualized under UV lights using a Bio-Rad Gel dock system.

### 4.4.9 Cell culture and transfection and fluorescence microscopy

Cos7 cells were maintained and grown in Dulbecco’s modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (Cellgro) and penicillin-streptomycin solution (Cellgro). Cells were transfected with 0.5 μg of each plasmid using polyethylenimine (PEI) reagent at 1:5 ratio. Overnight transfected Cells were fixed with 4% formaldehyde. Fluorescent microscopy images were acquired with a spinning disc confocal microscopy system (3I Corp) using a DMI 6000B microscope (Leica) and a digital camera (QuantEM; Photometrics).
4.4.10 Bacterial infection

The *L. pneumophila* Philadelphia-1 wild type strain Lp02, the type 4 secretion system defective strain Lp03(dotA-), the sidF deletion mutant and the complementation strain were described previously [252]. *L. pneumophila* was grown and maintained on CYE medium as previously described [264]. For infection, bacterial strains were grown to the post-exponential phase as determined by optical density of the cultures (OD$_{600}$=3.3-3.8) as well as an increase of bacterial motility.

4.4.11 Immunostaining

Mouse bone marrow-derived macrophages were infected with relevant strains for indicated time points. Samples were fixed following standard protocols. SidF and SidC were stained using specific antibodies described in earlier studies [252, 265].
Chapter 5: The *Legionella* effector SidC defines a unique family of ubiquitin ligases important for bacterial phagosomal remodeling


Abstract. 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*.

5.1 Introduction

The ubiquitous bacterium, *Legionella pneumophila*, is found in aquatic environments where it infects freshwater protozoa. Development of the potentially fatal Legionnaires’ disease occurs when susceptible individuals inhale contaminated aerosols [238]. 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 the creation of a specialized *Legionella*-containing vacuole (LCV) that provides the environment for robust intracellular
bacterial growth [242]. The mature LCV is characterized by an enrichment of a particular phosphoinositide lipid, PI(4)P [248, 266, 267], and by the accumulation of endoplasmic reticulum (ER) proteins, presumably captured by intercepting vesicles derived from the ER [268, 269]. Another feature of the LCV is enrichment of poly-ubiquitin conjugates around the vacuolar membrane [270]. This sophisticated membrane remodeling process is achieved by the coordinated activity of effector proteins delivered into the host through the Dot/Icm apparatus [251, 271]. However, the mechanism underlying these processes is still not well established. Recently, the Legionella effector proteins SidC (Substrate of Icm/Dot transporter) [265] 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 [247, 248, 265]. Bacterial vacuoles harboring the ΔsidC-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 [247]. 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 [272], cell signaling [197], and membrane trafficking [273, 274]. 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 [275]. 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 [276, 277]. E3s in the RING family facilitate the transfer of Ub from the E2 catalytic cysteine directly to a substrate [276]. However, HECT-domain E3s utilize their own active cysteine to form a thioester bond-linked E3−Ub intermediate before transferring 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 [278, 279]. For example, the Pseudomonas syringae protein AvrPtoB contains a RING-like domain and functions
as a ubiquitin ligase to inhibit host programmed cell death [280]. 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 [281]. 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 [282].

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 [283, 284]. Several of such predicted F-box containing effectors, including LegAU13/AnkB [285-287], LegU1, and LicA [285] 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 [285], and LubX [288]. 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 [288] and the Legionella effector SidH [289]. The F-box protein LegU1 specifically directs the ubiquitination of the host chaperone protein BAT3 [285]. 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 [270]. 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 re-modeling its protein composition.

5.2 Results

5.2.1 Crystal structure of the N-terminal domain of SidC reveals 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. 10A). The N-terminal portion of SidC has

![Image](image_url)

Fig. 10. Crystal structure of the N-terminal SNL domain of SidC (amino acids 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 subdomain within the SNL domain. (B) Ribbon diagram of the overall structure of the SNL domain. The SNL domain has two subdomains. The main subdomain contains a two-layered β-sheet (pink) flanked by two clusters of α-helices (blue). The all α-helical insertion subdomain is shown in green. (C) Multiple sequence alignment reveals two clusters of conserved residues. The first cluster contains an invariant cysteine C46, and the second cluster contains invariant H444 and D446. (D) Mapping of the invariant 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.
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. We named this conserved domain as SNL domain (SidC N-terminal ubiquitin Ligase 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 (Table 2). The structure reveals that the SNL domain has a crescent-like overall shape consisting of two sub-domains (Figs. 10A and B). A small all-alpha helical domain (in green) is inserted into the rest of the protein between residues 224-327. The main domain contains a core of two layered β-sheets, which is sandwiched between two clusters of α helices (Fig. 10B). Structural homology search with the Dali server [290] 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. 10C). 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. 10D). 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 [291] and deubiquitinases (DUB) [292], and other cysteine-based enzymes. 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.

5.2.2 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 (data not shown). 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. 11A). Some ubiquitinated species were more prominent (indicated by *) while others were absent (indicated by arrow heads) in transfected cells (Fig. 11A). 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 [293]. HEK293T cells were first co-transfected with HA-Ub and GFP-tagged SNL domain (either the wild type or its C46A mutant). After labeling 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. 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 3).
Fig. 11. Ectopic expression of SidC altered the intracellular ubiquitination pattern. (A) HEK293T cells were cotransfected 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), anti-GFP, and anti-GAPDH as a loading control (Right). Arrowheads highlight several bands that are positive in GFP and GFP-SNL C46A controls but are diminished in the presence of wild-type SNL domain. Asterisk denotes bands that are more prominent in the sample expressing wild-type SNL domain. (B) HEK293T cells were cotransfected 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.

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. 11B). In the presence of DTT, the E2~Ub bands were shifted down relative to their corresponding E2 band (Fig. 11B, 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. These data further suggest that SidC is not a DUB, which cleaves the isopeptide linkage between Ub and target proteins or within
Table 3. SILAC hits of proteins with increased ubiquitination.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Peptide count</th>
<th>( \text{gmean of log_2} ) ratio (Wt/mut)</th>
<th>Peptide count</th>
<th>( \text{gmean of log_2} ) ratio (mut/Wt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UBE2NL</td>
<td>9</td>
<td>-3.60566</td>
<td>4</td>
<td>3.08539</td>
</tr>
<tr>
<td>UBE2L3/UbcH7</td>
<td>11</td>
<td>-3.46593</td>
<td>12</td>
<td>6.35531</td>
</tr>
<tr>
<td>UBE2K/Ubc1</td>
<td>10</td>
<td>-2.64236</td>
<td>7</td>
<td>5.3087</td>
</tr>
<tr>
<td>UBE2S</td>
<td>9</td>
<td>-2.56502</td>
<td>7</td>
<td>3.80895</td>
</tr>
<tr>
<td>UBA1/Ube1</td>
<td>97</td>
<td>-2.53918</td>
<td>97</td>
<td>2.70132</td>
</tr>
<tr>
<td>BRAP</td>
<td>4</td>
<td>-2.37871</td>
<td>2</td>
<td>4.29822</td>
</tr>
<tr>
<td>UBE2T/HSPC150</td>
<td>9</td>
<td>-2.02673</td>
<td>3</td>
<td>2.64006</td>
</tr>
<tr>
<td>USP5/ISOT</td>
<td>35</td>
<td>-2.02175</td>
<td>34</td>
<td>3.8251</td>
</tr>
<tr>
<td>AUP1</td>
<td>8</td>
<td>-1.34807</td>
<td>2</td>
<td>2.34722</td>
</tr>
</tbody>
</table>

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. This hypothesis also explains the observation shown in Fig. 11A 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).

5.2.3 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. 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). 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. 12A). Since SidC appeared most active in the presence of UbcH7 (Fig. 12A), which was also among the top SILAC hits (Table 3), 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
Fig. 12. The SNL domain of SidC has E3 ubiquitin ligase activity. (A) In vitro ubiquitination assays with the SNL domain of SidC (amino acids 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, whereas 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 polyubiquitinated 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.

H444A and D446A showed reduced activity, whereas the activity was completely abolished in the C46A mutant (Fig. 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 were retained in the stacking gel (Fig. 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. 12D). Western blot with antiserum against UbcH7 also revealed poly- or multiple mono- ubiquitination of this E2 (Fig. 12E). These ubiquitinated E2s contributed to the ubiquitin signals derived from ubiquitin species in the middle range of molecular weights observed in Fig. 12C. E3 ligases often display autoubiquitination in in vitro assays, and in most instances this activity is considered as a mechanistically relevant readout [294].

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. 13). 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. 14). 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.

---

**Fig. 13.** 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).
5.2.4 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 chains 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. 15A). 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. 15B). These results indicate that SidC is able to use multiple lysine residues for ubiquitin polymerization with a preference for K11 and K33-linked chains.

Fig. 15. 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.

5.2.5 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 [247]. We thus investigated the relevance of the E3 ligase activity of SidC in Legionella infection. A Dictyostelium discoideum strain stably expressing GFP-HDEL [295] (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 (Figs. 16A 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 (Figs. 16A and B). 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 [270]. This recruitment is nearly abolished in the infection with the ΔsidC-sdcA mutant strain [296]. 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 ΔdotA or the ΔsidC-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 [287]. However, here we showed that in strain Lp02 [297], the recruitment of poly-ubiquitin species was not impaired (Fig. 17). Thus, the E3 ligase activity of SidC/SdcA, but not AnkB, is required for the recruitment
of ubiquitinated species to the LCV in strain Lp02.

**Fig. 17.** 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; ΔsidC-sdcA: the SidC and SdcA double deletion mutant of the Lp02 strain; ΔsidC-sdcA(pSidC) and ΔsidC-sdcA(pSidC C46A): ΔsidC-sdcA strain complemented with a plasmid expressing wild type or C46A mutant SidC; ΔsAnkB: the AnkB deletion mutant of the Lp02 strain. (B) Percent of cells containing ubiquitin positive LCVs counted from three independent assays under the conditions infected with the indicated *Legionella* strains.

**5.3 Discussion**

Through the Dot/Icm transporter, *L. pneumophila* delivers approximately 300 experimentally verified substrates into the host [251, 271]. 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 [247]. Recently, crystal structures of the N-terminal portion of SidC were reported [296, 298]. 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 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 [299, 300]. 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 [301-303]. However, the catalytic cysteine of the SNL domain is localized at the center of a surface of its main sub-domain (Figs. 10B 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* [241]. Given the fact that ubiquitination regulates a myriad of cellular functions, including cell cycle, cell death, trafficking, and immune responses [272], 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 [278]. 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 [285, 288]. 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 rewire 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 [296]. However, SidC apparently does not ubiquitinate Rab1 directly, since co-transfection of SidC and Rab1 fails to cause mono-ubiquitination of Rab1 [296]. 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 Rab1(S22N), or the constitutive active Rab1(Q70L), no ubiquitinated Rab1 was detected (data not shown). These data suggest that the mono-ubiquitination of Rab1 is mediated through an indirect unknown mechanism as proposed by the previous report [296].

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

5.4 Materials and Methods

5.4.1 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 [304]. 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 [305] inserted with wild type or C46A mutant SidC was used to complement the ΔsidC-sdcA mutant [265]. 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.

5.4.2 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 [306], 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.

5.4.3 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.

5.4.4 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. The crystals belong to space group P2₁2₁2₁ with a = 68.64 Å; b = 134.45 Å; c = 172.68 Å; α = β = γ = 90° (Table 2). The calculated Matthews coefficient Vₘ = 3.32 and with
62.9% of solvent in the crystal and two protein molecules in an asymmetric unit [263].

5.4.5 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 [231]. 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 [307]. Iterative cycles of model building and refinement were carried out with the program COOT [232] and the refmac5 program [233] in the CCP4 suite [234] to complete the final model.

5.4.6 In vitro deubiquitination assay

Lys-48- or Lys-63-linked polyubiquitin chains (Ub1-7) 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).

5.4.7 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.
5.4.8 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% CO2 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.

5.4.9 SILAC labeling and immunoprecipitation

HEK293T cells were grown for 2 weeks in complete media containing normal lysine and arginine (“light”) or [13C6,15N2] lysine and [13C6,15N4] 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. [308].

5.4.10 Legionella strains and infection

Strains of *L. pneumophila* used include the wild type Lp02 [297], the Dot/Icm deficient Lp03 [297] and the *sidc-sdcA* mutant [265]. *Dictyostelium discoideum* strain AX4 stably expressing
HDEL-GFP was cultured at 21.5°C 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 2×10⁵ cells per well. After incubation for 2 hrs at 25°C, cells were infected with *L. pneumophila* grown to post-exponential phase for 2 hrs at an MOI of 2. To detect translocated SidC, 2×10⁷ 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°C. The supernatants (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. [308]. 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, 2×10⁵ 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.

### 5.4.11 Antibodies, immunostaining, and Western blot

Anti-*L. pneumophila* [309] and anti-SidC [265] were described previously. Anti-tubulin antibody was purchased from (DSHB, University of Iowa). Infected *D. discoideum* samples were fixed and stained as described earlier [295]. For ubiquitin immunofluorescent analyses, U937 cells were fixed and stained by standard procedures [310]. 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.
Western blots were performed following standard protocols [309]; anti-SidC and anti-tubulin antibodies were used at 1:10,000; 1:20,000 and 1:5000, respectively.
Chapter 6: Bacterial effectors utilize host phosphoinositides to facilitate intracellular replication

6.1 Introduction

It has been increasingly recognized that modulation of host PI signaling and metabolism is critical for the pathogenicity of many human pathogens [145, 311]. One strategy to exploit host PI pathways by several bacterial pathogens is to inject PI metabolizing enzymes into the host. Many of these enzymes, particularly PI phosphatases have been documented. The enterobacteria *Salmonella enterica*, which causes food-borne gastroenteritis and typhoid fever in human, delivers a PI phosphatase SigD/SopB into the host and is essential for enteropathogenicity [312]. SopB is shown to have sequence homology with mammalian inositol polyphosphate 4-phosphatases and have a broad spectrum of substrate specificity [250, 313]. *Shigella flexneri*, the causative agent of human dysentery injects a SopB ortholog, IpgD into the host cells. IpgD prefers to hydrolyze PI(4,5)P\(_2\) to generate PI(5)P [243] leading to membrane blebbing and actin filament remodeling and promote the entry of the bacterium. *Mycobacterium tuberculosis*, the causative agent for tuberculosis, secretes two PI phosphatases SapM and MptpB. SapM hydrolyzes PI(3)P and inhibits phagosome-late endosome fusion in vitro, and inhibits phagosomal maturation [314]. MptpB was first reported as a protein tyrosine phosphatase [315] and was later demonstrated to preferentially hydrolyze PI(3)P and PI(3,5)P\(_2\) [316]. The marine bacterium, *Vibrio parahaemolyticus*, which causes gastroenteritis in humans, encodes a PI-5-phosphatase-like effector VPA0450 [317]. VPA0450 hydrolyzes the D5 phosphate from the plasma membrane PI(4,5)P\(_2\) causing plasma membrane blebbing and facilitating cell lysis [317].

Recently, we discovered in *Legionella pneumophila*, the causative agent of Legionnaires’ disease [266, 267], encodes two novel CX\(_5\)R-based PI phosphatases named SidF (Chapter 4) and SidP [269], and an E3 ubiquitin ligase, SidC (Chapter 5). Both SidF and SidP are PI-3-phosphatases with different substrate specificity. SidF specifically hydrolyzes the D3 phosphate of PI(3,4)P\(_2\) and PI(3,4,5)P\(_3\) [266], while SidP hydrolyzes PI(3)P and PI(3,5)P\(_2\) [267]. SidC is a
PI(4)P-binding effector and is required for efficient recruitment of ER-derived vesicles through its ubiquitin ligase activity. Through the concerted action of these three PI-utilizing enzymes, *Legionella* may be able to change the lipid identity and protein composition of the *Legionella*-containing vacuoles, thus establishing an amenable niche for bacterial intracellular propagation. In this final chapter, we will discuss the structural similarity between SidF and SidP, and apply the structural information to propose a general catalytic mechanism for bacterial CX₅R-based PI phosphatases.

### 6.2 Crystal structures of SidF and SidP

SidF and SidP have unique primary sequences. Except for the sequence near the catalytic core, SidF and SidP lack detectable sequence homology with other known PI phosphatases. However, the 3D structures of the phosphatase catalytic domains of SidF and SidP are highly similar with each other. Both SidF and SidP consist of a central 10-11 pleated β-sheet sandwiched within 18-19 α-helices (Chapter 1, Figs. 4J-L). The catalytic core of these two enzymes, which contains four parallel β strands and one α-helix with the catalytic P-loop connecting between one of β strands and the α-helix, are conserved among all CX₅R motif-based phosphatases (Chapter 4, Fig. 4K). In SidF, the catalytic site falls in a groove, which nearly bisects the catalytic domain. This groove is highly positively charged and may facilitate the loading of lipid substrate to the catalytic site (Chapter 4, Fig. 4L). The determination of SidF with a lipid substrate molecule diC₄ PI(3,4)P₂ highlighted the structural basis for membrane interaction, substrate recognition, and catalytic mechanism [266].

### 6.3 Membrane binding

The complex structure of SidF with diC₄ PI(3,4)P₂ immediately reveals that the acyl chains of the substrate interface with two hydrophobic loops (colored in gold in Fig. 18B). These two hydrophobic loops are likely to penetrate into the hydrophobic lipid bilayer while the substrate is loaded to the catalytic site. In addition to hydrophobic residues, two cationic residues (R366, R422) may contribute to the electrostatic interactions with negatively charged lipid head.
groups. Thus, similar to 5-phosphatases, both hydrophobic and electrostatic interactions facilitate the stable association of the enzyme with membrane bilayers to enhance the processivity of catalysis.

### 6.4 Substrate specificity

The complex structure of SidF with diC₄ PI(3,4)P₂ delineates a cationic pocket mainly formed by residues H233, K646, S647, K717, and K740 for the recognition of the D4 phosphate of PI(3,4)P₂ [266] (Fig. 18C). These residues distribute rather dispersedly in primary sequence, but cluster together to form a pocket critical for substrate selectivity. Similar to the P4IM described in 5-phosphatase, extensive hydrogen bonds and electrostatic interactions are involved to accommodate the D4 phosphate. One the other hand, the binding of D1 phosphate group is mediated by R651 through salt bridge and N419 through hydrogen bond interactions. Besides the two phosphate groups, a hydrogen bond between the 2’ hydroxyl group of the inositol ring with the side chain of D650 also plays a key role in the orientation of the substrate at the catalytic site. These specific interactions allow the positioning of the scissile D3 phosphate close to the catalytic residue C645.

### 6.5 Catalytic mechanism

The catalytic mechanism is conserved among all CX₅R-based phosphatases, including the well-studied protein tyrosine phosphatases [258, 318]. In the case of SidF, the D3 scissile phosphate is held in a position close to the Sγ-atom of C645 for nucleophilic attack through intensive polar interactions with five main-chain amide groups of the catalytic P-loop and with the guanidinium group of the CX₅R arginine. The reaction can be divided into two steps shown in Fig 18D. The first step is the formation of a phospho-cysteine intermediate. The cleavage starts with a nucleophilic attack by the deprotonated Sγ-atom of C645. The breakage of the scissile P-O bond between the D3 phosphate and the inositol ring is likely be facilitated by D650, which protonates the 3’ hydroxyl as a general acid. As a result, the transient intermediate, phospho-cysteine
is formed. The second step is the hydrolysis of the cysteinyl-phosphate intermediate. The hydrolysis requires a water molecule, which is like to be activated by the same D650 accepting a proton from the attacking water molecule as a general base (Fig. 18D).

Fig. 18. Membrane binding and substrate specificity of SidF. (A) Ribbon diagram of the structure of SidF with bound diC4 PI(3,4)P2 (in sticks) (PDB ID: 4FYG). (B) Zoom-in view of the SidF MIM (in gold). (C) Zoom-in view of the interactions between SidF with the head-group of PI(3,4)P2. Residues H233, K740, K646, S647, and K717 form a strong cationic pocket for the specific docking to the D4 phosphate of the substrate. Residue N419 forms a hydrogen bond with D1 phosphate. Like other CX5R phosphatases, the scissile phosphate is positioned close to the catalytic cysteine by interactions with the P-loop main-chain amide groups and R651. (D) The reaction has two steps. In the first step, the catalytic cysteine attacks the scissile phosphate to break the bond between the phosphate and hydroxyl group of the lipid. D650 functions as the general acid to donate a proton to the leaving hydroxyl, resulting in a phospho-cysteine intermediate. In the second step, D650 now functions as the general base to activate a water molecule. The activated water molecule hydrolyzes the phospho-cysteine intermediate to restore the enzyme.

6.6 PI-binding effector proteins

In Legionella, at least three effectors have been identified to encode PI-binding domain.

DrrA/SidM is a guanine nucleotide exchange factor (GEF) and also has adenylyl transferase activity specific for Rab1 [319]. Interestingly, DrrA/SidM contains a PI(4)P binding module that
is important for binding to PI(4)P-enriched LCV and exhibits strong PI(4)P-dependent membrane insertion [320]. RidL is a PI(3)P-binding effector that binds simultaneously to the retromer subunit, Vps29, and PI(3)P. As a result of this coincidence detection, RidL inhibits the host ability to restrict *Legionella* replication [321]. SidC is another effector that contains not only a PI(4)P binding domain but also an enzymatic domain to carry out the ubiquitination reaction [322]. Together, these effectors, once secreted, decorate the PI(4)P-enriched LCV via its PI(4)P binding module and execute its action at the precise location. These findings illustrated the evolutionary advantage for intracellular pathogens to encoded not only PI-metabolizing enzymes but also PI-binding effector proteins to facilitate their survival.

**Summary**

Our studies on the role of both mammalian and *Legionella* PI phosphatases have provided valuable insights into its substrate specificity as well as the catalytic mechanism, in particular, the “CX₅R”-based enzymes. Interestingly, we found that Sac2 is a PI(4)P phosphatase that functions specifically at endosomal structures. This is the first reported case of a PI 4-phosphatase to hydrolyze a distinct pool of PI(4)P other than the Golgi. The role of Sac2 appears to be tissue-specific as the protein level is highest in the brain. However, the physiological role of Sac2 in the brain remains unclear and requires further investigation in the mouse model.

The discovery of two novel PI phosphatases and the role of a PI(4)P-binding effector in *Legionella* raises the question as to how the host carries out its defense mechanism during infection. Future studies that examine the interacting partners of SidF, SidP, and SidC would provide useful clues. In addition, proteomic approach to search for the substrate of SidC would shed light on the components involved in host immune signaling and membrane trafficking. In a constant battle for survival, deciphering these signaling events is paramount to developing better therapeutics.
REFERENCES


