STRUCTURAL AND KINETIC CHARACTERIZATION OF PROTEIN

AMPYLATION BY VOPS FIC DOMAIN

APPROVED BY SUPERVISORY COMMITTEE

Kim Orth, Ph.D.

Xuewu Zhang, Ph.D.

Diana Tomchick, Ph.D.

Kevin Gardner, Ph.D.

Margaret Phillips, Ph.D.

STRUCTURAL AND KINETIC CHARACTERIZATION OF PROTEIN AMPYLATION BY VOPS FIC DOMAIN

by

PHI HOANG LUONG

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

November, 2011

DEDICATION

To Xiaomo Jiang

ACKNOWLEDGEMENTS

I thank my cool mentor, Kim Orth. You are cool times triple infinity. Thank you for your cool endless encouragement, your cool mentorship, and your cool support. You have been a positive influence in my graduate studies and in my life. You are always optimistic, which I should learn from you when I get a good result. Thank you so much for your guidance and please allow me to bug you once in a while when I'm gone.

I want to thank Yi-Heng Hao. Yi-Heng was my teacher throughout my graduate studies and has taught me everything he knows. Yi-Heng made my day to day experiments a lot easier. You always helped answer my questions. I was very lucky to have you as my teacher.

I thank my committee members Diana Tomchick, Margaret Phillips, Xuewu Zhang, and Kevin Gardner. Thank you, Diana, for all of your guidance. There was absolutely no way the structure of VopS would have been solved without you. You are a beautiful person and I am very fortunate you were my mentor. Thanks, Meg, for your invaluable guidance on the enzymology project. I needed a lot of input for the kinetic studies and I was glad I could pick at your brain. Thank you, Xuewu, for being my committee chair. I felt comfortable just hopping into your lab to ask for things and to talk to you. Thanks, Kevin, for being attentive about my well-being during the years and during my post doc searching. I picked you as a committee member, because I thought you are absolutely awesome and you make me laugh.

Yan Li, thank you for being such a great help for mass spectrometry. I wish only the best for you. Chad Brautigam, thank you for your help with the structure project and comments on the VopS manuscript. Thank you, Radha Akella and John Humphreys, for your continuous support in reagents and your expertise in structure and enzymology. You guys are always fun to talk to. I thank Lisa Kinch who provided powerful insight with her

god-like skills in bioinformatics. Thank you, Carlos Huerta, for being very helpful during my kinetic studies. Yan, Chad, Radha, John, Lisa, and Carlos, thank you so much for letting me pick at your brains throughout the years.

I want to thank Ron Taussig. At various random times I would show up in your lab and ask for help. You were always willing to provide the tools and reagents I needed for the experiments. You are pretty cool yourself.

Thanks Howard Hang and Markus Grammel for such a great collaboration. Your expertise made the AMPylation “click” project go very smoothly.

I want to thank past and current lab members for making the lab a lively and productive place to work in. Thanks Melanie Yarbrough and Dara Burdette for laying the ground work for my studies. I thank my friends and colleagues at UT Southwestern that helped made graduate school and research easier. Xiaomo, Tina, Jessica, Greg, Andy, Thi, Francis, Laura, Jenny, Ben, and Nishanth you guys are amazing. Graduate school was fun and I wouldn’t mind staying a little longer, but I guess it’s time to move on.

BIOCHEMICAL AND STRUCTURAL CHARACTERIZATION OF PROTEIN AMPYLATION BY VOPS FIC DOMAIN

PHI HOANG LUONG

The University of Texas Southwestern Medical Center at Dallas, 2011

Supervising Professor: KIM ORTH, Ph.D.

The bacterial pathogen *Vibrio parahaemolyticus* manipulates host signaling pathways by injecting type III effectors into the cytoplasm of the target cell. One of these effectors, VopS, blocks actin assembly by AMPylating a conserved threonine residue in the switch 1 region of Rho GTPases. The modified GTPases are no longer able to interact with downstream effectors due to steric hindrance by the covalently linked AMP moiety.

Herein we analyze the structure of VopS and its evolutionarily conserved catalytic residues. We describe features of the VopS crystal structure, including a hairpin element that is responsible for protein–protein interaction and residues involved in ATP binding. Steady-state analyses of VopS point mutants provide kinetic understanding on the functions of conserved residues for the AMPylation activity. Further mechanistic analysis of VopS with its two substrates, ATP and Cdc42, demonstrates that VopS utilizes a sequential mechanism to AMPylate Rho GTPases. The structure of VopS and its ternary reaction mechanism provide critical groundwork for future studies on AMPylators, a novel family of enzymes that modify hydroxyl-containing residues with AMP.

We also developed molecular tools to facilitate the study of protein AMPylation in collaboration with Howard Hang at The Rockefeller University. An ATP analogue, N^6pATP, was developed that utilizes click chemistry to allow for the detection of AMPylated proteins by fluorescent or biotin tags. N^6pATP can be utilized in *in vitro* AMPylation reactions catalyzed by known AMPylators including Fic domain and adenylyltransferase domain proteins. Further, we showed that N^6pATP can be used for the detection and purification of endogenous AMPylated proteins.

Preliminary studies were performed on another effector protein of unknown function, VopQ from *Vibrio parahaemolyticus*. The protein sequence of VopQ does not resemble any known protein domains. Various constructs were made for VopQ, and here I describe the purification and crystallization of VopQ.

TABLE OF CONTENTS

LIST OF PUBLICATIONS

Grammel M*, **Luong P***, Orth K, Hang HC. (2011) A chemical reporter for protein AMPylation. *Journal of the American Chemical Society*. Published online 23 September 2011; 10.1021/ja205137d.

Hao YH, Chuang T, Ball HL, **Luong P**, Li Y, Flores-Saaib RD, Orth K. (2011) Characterization of a rabbit polyclonal antibody against threonine-AMPylation. *Journal of Biotechnology*. 151(3): 251-254.

Woolery AR, **Luong P**, Broberg CA, and Orth K. (2010) AMPylation: Something old is new again. *Fronteirs in Cellular and Infection Microbiology,* 1 (113). Review

Luong P, Kinch KN, Brautigam CA, Grishin NV, Tomchick DR, Orth K. (2010) Kinetic and structural insights into the mechanism of AMPylation by VopS Fic domain. *The Journal of Biological Chemistry*, 285(26): 20155-63.

Shen H, **Luong P**, Huq E. (2007). The F-box protein MAX2 functions as a positive regulator of photomorphogenesis in Arabidopsis. *Plant Physiology*, 145(4): 1471-83.

* These authors contributed equally to this work

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2XYT	2 times yeast tryptone
ADP	Adenosine diphosphosphate
AMP	Adenosine monophosphate
AMPCPP	α,β-Methyleneadenosine 5′-triphosphate
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
CBD	Chaperone binding domain
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EHEC	*Enterohaemorrahagic Escherichia coli*
EPEC	*Enteropathogenic Escherichia coli*
Fic	Filamentation induced by cyclic AMP
FPLC	Fast protein liquid chromatography
GDP	Guanosine diphosphate
GS-ATase	Glutamine synthetase adenylyltransferase
GST	Glutathione S-transferase
GTP	Guanosine triphosphate
GTPase	Guanosine triphosphatase
His	Histidine
IPTG	Isopropyl-beta-D-thiogalactopyranoside
kDa	Kilodalton
MWCO	Molecular weight cutoff
PBS	Phosphate-buffered saline
PDB	Protein data bank

PPi	Pyrophosphate
RSMD	Root mean square deviation
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
S.E.	Standard error
SeMet	Selenium methionine
T3SS	Type III secretion system
T3SS1	Type III secretion system one
T3SS2	Type III secretion system two
TPR	Tetratricopeptide repeat
V. para.	*Vibrio parahaemolyticus*
Vop	Vibrio outer protein

CHAPTER 1

Review of Literature

The Type III Secretion System and *Vibrio parahaemolyticus*

Pathogens exert multiple virulence mechanisms that promote pathogenesis and disrupt host immune defenses during infection. Many human bacterial pathogens encode a common virulence mechanism called Type III Secretion System (T3SS) (1). The T3SS are essential for bacterial virulence and human food-borne illnesses caused by pathogenic bacteria, such as species from *Salmonella*, *Shigella*, *Yersinia*, and *Vibrio* genera and pathogenic *Escherichia coli* strains.

The T3SS apparatus forms a direct contact between the bacterium and the host cell (Figure 1). It contains a base that spans across bacterial membranes, and a needle-like structure that extends through the bacterial cell wall and across the host cell membrane. A bacterial translocator complex forms the pore in the host membrane. The formation and extension of the needle complex into the host cell allow a set of bacterial proteins, called effectors, to be translocated through the needle into the host cytosol (2). These Type III secreted effectors are quiescent in the pathogen but mimic or capture an endogenous eukaryotic host activity inside the host cell to disrupt host function and promote pathogen survival (1).

Vibrio parahaemolyticus (*V. para*). is the leading cause of seafood-associated food poisoning in humans through the consumption of undercooked shellfish. Infection with *V. para* causes gastroenteritis associated with clinical symptoms including diarrhea, nausea, vomiting, fever, and headache (3,4). Pathogenic strains

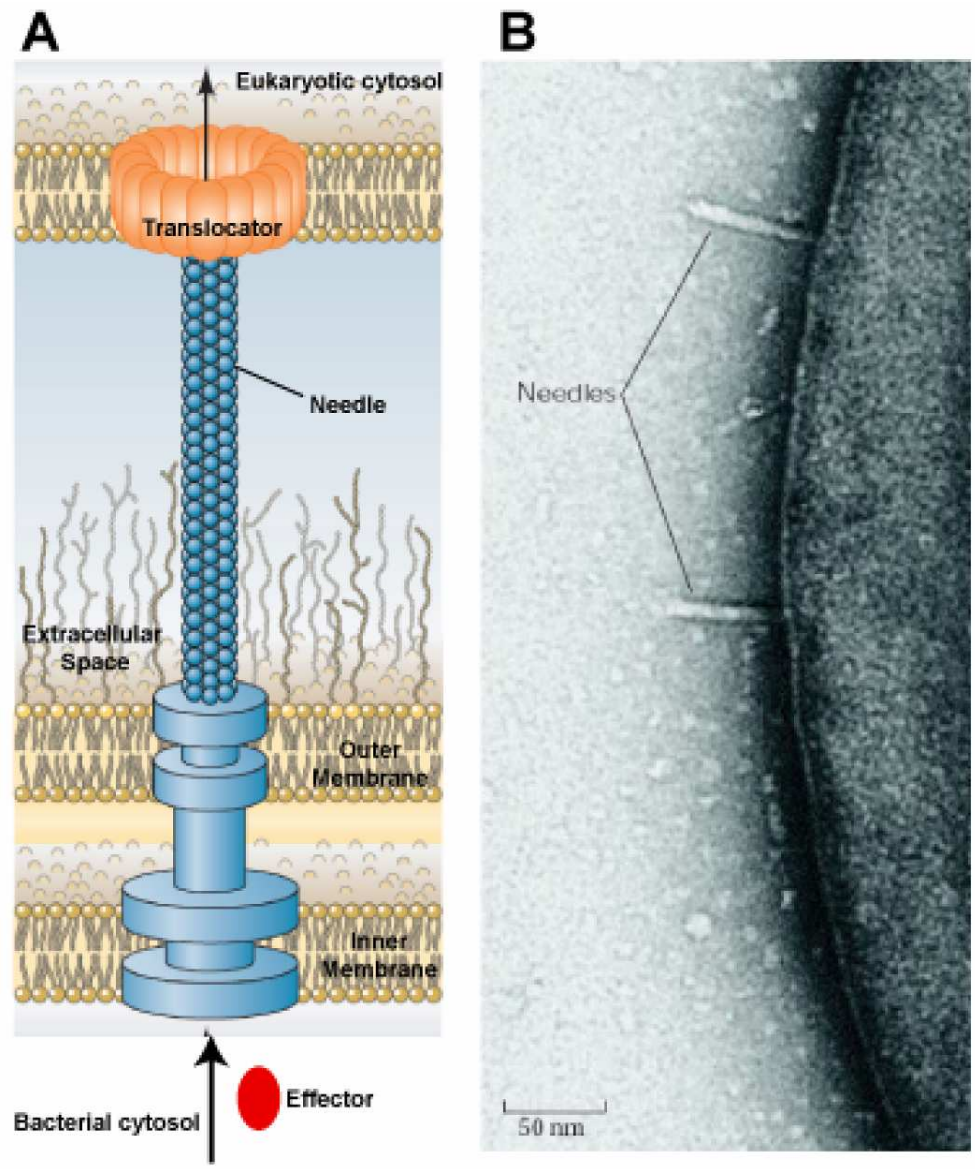

Figure 1. The Type III Secretion Machinery. (A) The Type III secretion needle system spans both the inner and outer membranes of the bacteria and forms a pore in the eukaryotic plasma membrane. **(B)** Electron micrograph of Type III secretion needles from *Yersinia enterocolitica*. Figures adapted from (5).

of the Gram-negative bacteria *V. para* encode two different T3SSs, and each system translocates a unique repertoire of effectors. The first secretion system, T3SS1, in *V. para* is necessary for cytotoxicity observed in infected tissue culture cells, while the second secretion system, T3SS2, has been associated with enterotoxicity, in a model of rabbit ileal loop infection (6). Known *V. para* effectors secreted by T3SS1 include VopQ (VP1680), VopR (VP1683), VopS (1686), and VPA0450. Together these four effectors contribute to a specific cytotoxicity mechanism that involves the induction of autophagy, followed by cell rounding, and finally cell lysis (7). The mechanism of VopR-mediated cytotoxicity is unknown; VPA0450 has been identified as an inositol polyphosphate 5-phosphatase that cleaves the 5 position of inositides inside host cells during infection (8). My projects focus on VopQ and VopS, discussed below.

Bacterial Effector Chaperones

Effector chaperones keep effectors in an inactive state in the bacterium, and deliver effectors to the needle for secretion. Different effector chaperone proteins have high structural similarities that are usually not reflected by their primary sequences (9). These chaperones are typically around 15-17 kDa, and are acidic in nature with an isoelectric point of around 4. Chaperones form homodimers, or heterodimers with other chaperones, resulting in a heart shaped structure. Several bacterial effector chaperones have been shown to interact with the N-terminal regions of effector proteins (Figure 2A and 2B). Crystal structures of chaperones in complex with the chaperone binding domains (CBD) of effector proteins reveal that a chaperone homo dimer interacts with a monomer of the effector (10,11).

In the *V. para* genome, positioned in tandem with VopQ, VopR, and VopS genes are their putative chaperones, VP1682, VP1684, and VP1687, respectively. VP1682, VP1684, and VP1687 were predicted to be chaperones due to their structural

characteristics. (12). Only VP1682 has been biochemically characterized to be the chaperone for VopQ (13). VP1684 and VP1687 are predicted to be the chaperones for VopR and VopS because they are positioned in tandem to each other. It is unclear if VPA0450 has a designated chaperone.

The Vibrio parahaemolyticus effector VopS is an AMPylator of Rho GTPases

Initial infection studies with wild-type lab strain and a VopS mutant strain of *V. para* demonstrated that VopS inhibits the function of Rho GTPases, such as Rho, Rac and Cdc42 (14). The Rho family GTPases are small G proteins that have essential roles in actin cytoskeletal dynamics (14). Bioinformatic analysis revealed that the C-terminus of VopS contains a protein domain called Fic (filamentation induced by cyclic adenosine monophosphate) (15). Yarbrough and colleagues (2009) discovered that the Fic domain of VopS uses ATP (adenosine triphosphate) to directly transfer an AMP (adenosine monophosphate) moiety to conserved threonine residues (threonine 35, Thr35) on Cdc42 and Rac1, and threonine 37 (Thr37) on RhoA. This newly discovered posttranslational modification by VopS Fic domain was named AMPylation (15). The conserved threonine of Rho family GTPases is located on the switch 1 region, and is responsible for coordinating the magnesium involved in GTP binding. AMPylation of this residue prevented Rho GTPases from interacting with downstream effectors, thereby inhibiting actin assembly in infected cells (Figure 3). The AMPylation activity of VopS was abolished when a histidine in a conserved motif in the Fic domain was mutated to alanine (15). Within months of the publication of this work on VopS, a second Fic domain containing protein, IbpA from *Histophilus somni* (*H. somni.)*, was described to AMPylate Rho GTPases on a conserved tyrosine residue (Tyr32) in the switch 1 region. AMPylation by the Fic domain of IbpA also resulted in a cell rounding phenotype (16).

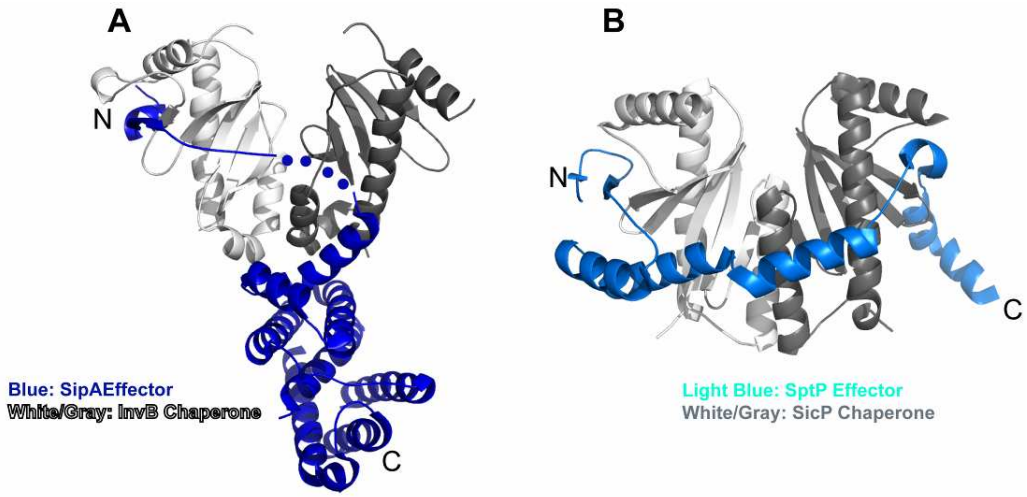

Figure 2. Structures of T3SS bacterial chaperone complexed with the CBD domain of effectors. (A) Structure of chaperone InvB dimer (white/grey) complexed with SipA (blue) CBD [PDB ID 2FM8] (10). **(B)** Structure of chaperone SicP dimer (white/grey) complexed with CBD region of SptP (light blue). N and C-terminus of bacterial effector are designated [PDB ID 1JYO] (11).

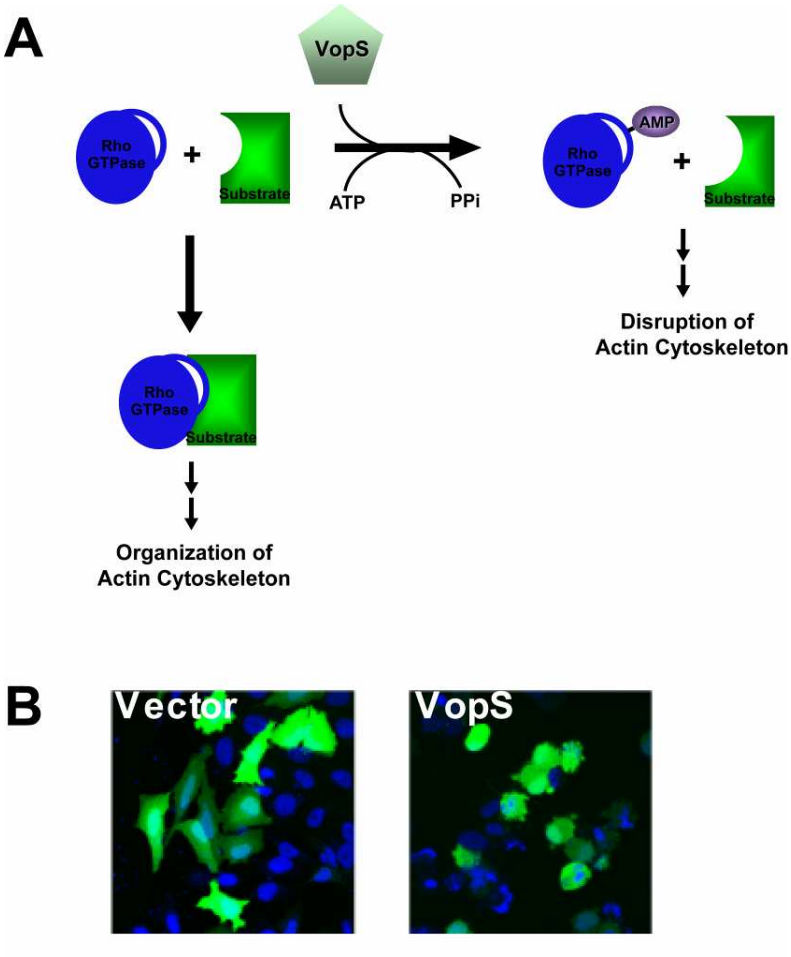

Figure 3. AMPylation of Rho GTPases disrupts actin cytoskeleton. (A) AMPylation of Rho GTPases by VopS sterically hinders substrate binding leading to disruption of actin cytoskeleton dynamics. **(B)** Transfection of HeLa cells with VopS induces cell rounding. Nuclei stained blue and green fluorescent protein used as transfection marker. Figure 3B adapted from [8].

The Rho family GTPases

Rho family GTPases control actin cytoskeleton dynamics by defining cell polarity, shape and movement (17). Rho family GTPases are approximately 21kDa proteins and belong to the Ras superfamily of small G proteins. GTPases are molecular switches that adopt two distinct conformations depending on the nucleotide bound state. GTPases are inactive when bound to GDP and active when bound to GTP. In the active conformation, GTPases use their conserved structural elements, called switch 1 and switch 2 regions, to interact with protein substrates and stimulate downstream pathways controlling actin dynamics (18). GAPs (GTPase-activating proteins), GEFs (guanine nucleotide exchange protein), and GDIs (guanine nucleotide dissociation inhibitors) are regulators of GTPases (19). GAPs stimulate the GTPase to hydrolyze GTP to GDP which will inactivate the GTPase. GEFs promote GTPase activation by inducing a release of bound GDP in the GTPase in exchange for GTP. GDIs bind to GTPases to inhibit nucleotide exchange and also prevent GTPases from localizing to the membrane (20). The GTPase itself has an enzymatic activity that can hydrolyze GTP to GDP and this hydrolysis activity is promoted by interaction with GAP (GTPase-activating proteins) proteins.

The first Rho family member, RhoA, was discovered in 1985 and currently there are at least 20 known members of this family (21). The best studied members are RhoA, Rac1, and Cdc42 and their activation induces cytoskeletal effects in the form of stress fibers, lamellipodia and filopodia, respectively (Figure 4).

The Fic Domain

Fic domains are evolutionarily conserved from prokaryotes to eukaryotes and fall into a domain superfamily called Fido, which consists of three protein domain families:

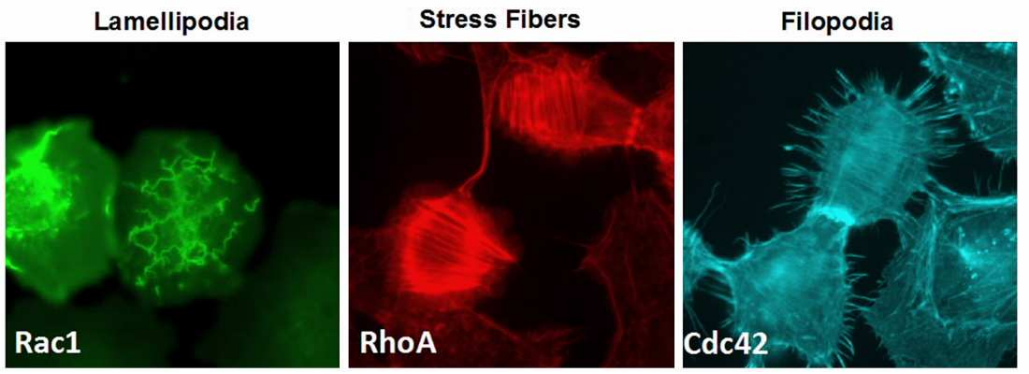

Figure 4. Actin structures. Activation of Rac1, RhoA, and Cdc42 induces formation of lamellipodia, stress fibers, filopodia actin structures, respectively. Images are pseudocolored highlighting cellular actin. Image courtesy of Neal Alto.

Fic domains, Doc (death on curing) domains, and AvrB (avirulence protein B) (15,22). All members of the Fido superfamily have similar tertiary topology (Figure 5A). Fic and Doc domains have the conserved central HPFx[D/E]GN[G/K]R motif while AvrB members lack the conserved histidine in this motif (22). Higher eukaryotes such as human, fly and worm possess only one Fic domain containing protein in each species (22). The eukaryotic Fic proteins appear to have a single-spanning transmembrane region at their N-termini that is not observed in bacterial Fic proteins.

Structures of Fic-containing proteins from a range of bacterial species including *Helicobacter pylori* (*H. pylori*) [PDB ID 2F6S], *Shewanella oneidensis* (*S. oneidensis*) [PDB ID 3EQX previous PDB ID was 2QC0], *Bacteroides thetaiotaomicron* (*B. thetaiotaomicron*) [PDB ID 3CUC], and *Neisseria meningitidis* (*N. meningitides*) [PDB ID 2GO3] have been determined by the Protein Structure Initiative (Figure 5B and 5C) (22,23). These proteins have not been biochemically characterized and their functions are unknown. A multiple sequence alignment of the AvrB, Doc, and Fic family are shown in Figure 5D. The alignment shows Fic protein from *H. pylori* [PDB ID 2F6S], *S. oneidensis* (labeled as 2QC0 in the alignment), and *N. meningitides* [PDB ID 2GO3] along with the predicted VopS Fic domain region (Figure 5D). The Fic domain family displays a conserved topology, with a central 2-helix bundle encircled peripherally by up to 6 helices (Figure 5A). The Fic motif HPFx[D/E]GN[G/K]R lies within the loop connecting the 2 central helices, a4 and a5. A conserved beta hairpin/loop structural element depicted as b1 and b2 in Figure 5A is positioned near the Fic motif and is also present in all solved structures (22). Individual Fic proteins may also have unique non-conserved structural features. For example, as shown in Figure 5C, the non-conserved regions of a Fic domain protein from Bacteroides thetaiotaomicron are colored in white. VopS also has a distinctive N-terminal region encompassing two-thirds of the protein not

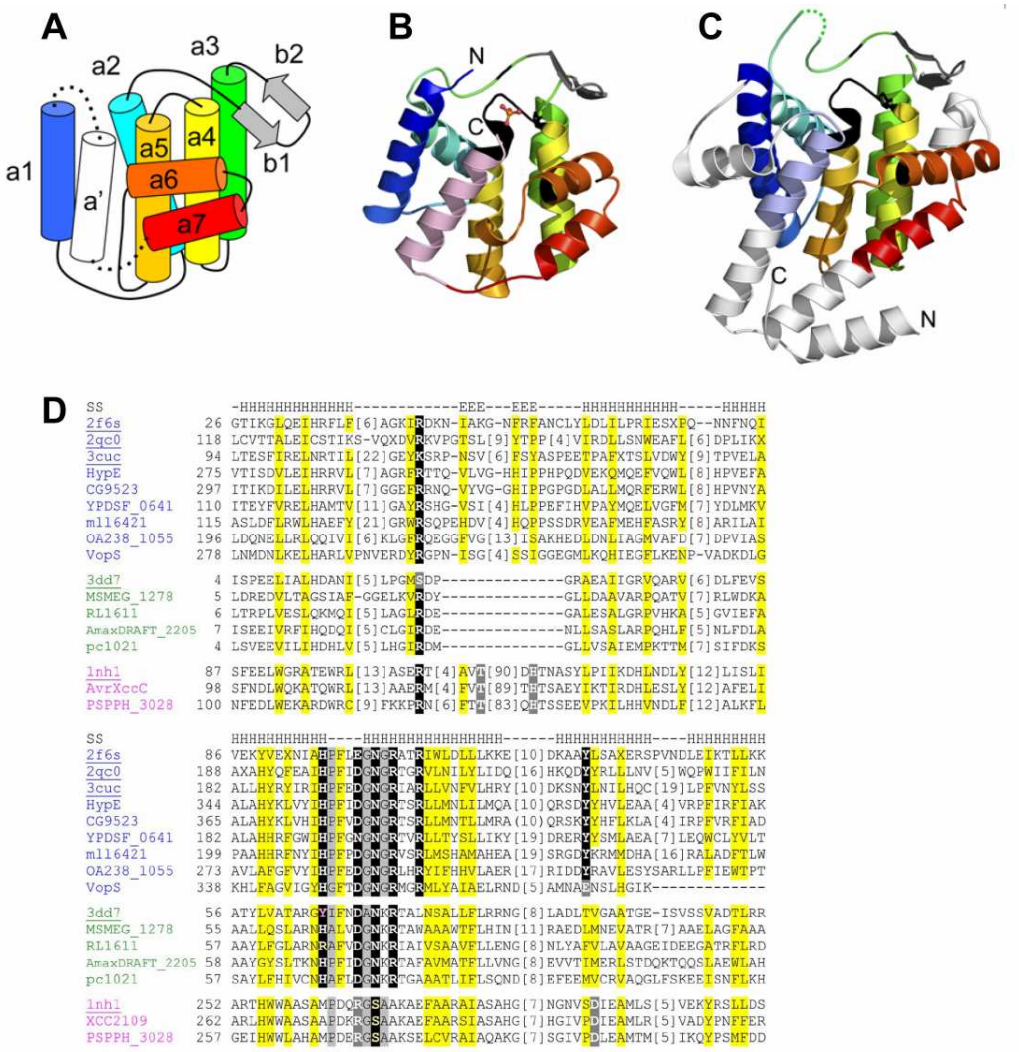

Figure 5. Examples of Fic protein structures. (A) Cartoon of the core conserved Fic domain and letter assignment of conserved structural elements. **(B)** Structure of Fic protein from *H. pylori* [PDB ID: 2f6s]. **(C)** Structure of Fic protein from *B. thetaiotaomicron* [PDB ID: 3cuc] White colored secondary elements designates nonconserved structural features in the protein. N and C-terminus are designated N and C respectively. **(D)** Multiple sequence alignment of the Fido domain family members. Fic (blue), doc (green) and AvrB (magenta) family members were aligned where hydrophobic residues are highlighted yellow, conserved small residues are grey, and conserved polar residues are colored black. VopS belong to the Fic family. Notice that doc members lack conserved beta-strands and AvrB members do not have the conserved histidine that is present in Fic family members. Secondary structure (SS), helix (H), and beta-strands (E) are designated above each alignment. Figures were adapted from (22).

conserved in other Fic proteins according to primary sequence analysis. In addition, the tyrosine AMPylation activity of IbpA requires a nonconserved region to the N-terminus of its Fic domain (16). The identification of AMPylation activity was the first biological function ascribed to a Fic domain-containing protein. The function of Fic domains as enzymes capable of phosphotransfer reactions was not known until the biochemical characterization of VopS, therefore, we set out to determine the structure of VopS, a Fic-containing AMPylator, and compare it to known Fic domain structures with unknown functions (24).

The Discovery of Protein AMPylation A Half Century Ago.

Protein AMPylation (adenylylation) was discovered by Earl Stadtman in the 1960s when he found that glutamine synthetase adenylyltransferase (GS-ATase) in *E. coli* can transfer AMP to glutamine synthetase. GS-ATase domains possess a G-X_{11}-D-X-D motif that is required for their activity (25). GS-ATase has two independent catalytic domains that can catalyze the removal or addition of AMP to glutamine synthetase, respectively (Figure 6A) (26). The structures of these two adenylyltransfearase (ATase) domains have been solved individually. The N and C terminal ATase domains share 24% sequence identity and high structural similarity with an RMSD of 2.4 Å (Figure 6B) (27) Both de-AMPylation and AMPylation activities target tyrosine 397 of glutamine synthetase. (27,28)

The physiological function of GS-ATase is to control nitrogen metabolism (29). When intracellular nitrogen levels are high, such as high glutamine levels, GS-ATase mediates AMPylation to inhibit glutamine synthetase to inhibit glutamine synthesis. Under low nitrogen conditions, such as low glutamine levels, the GS-ATase removes the

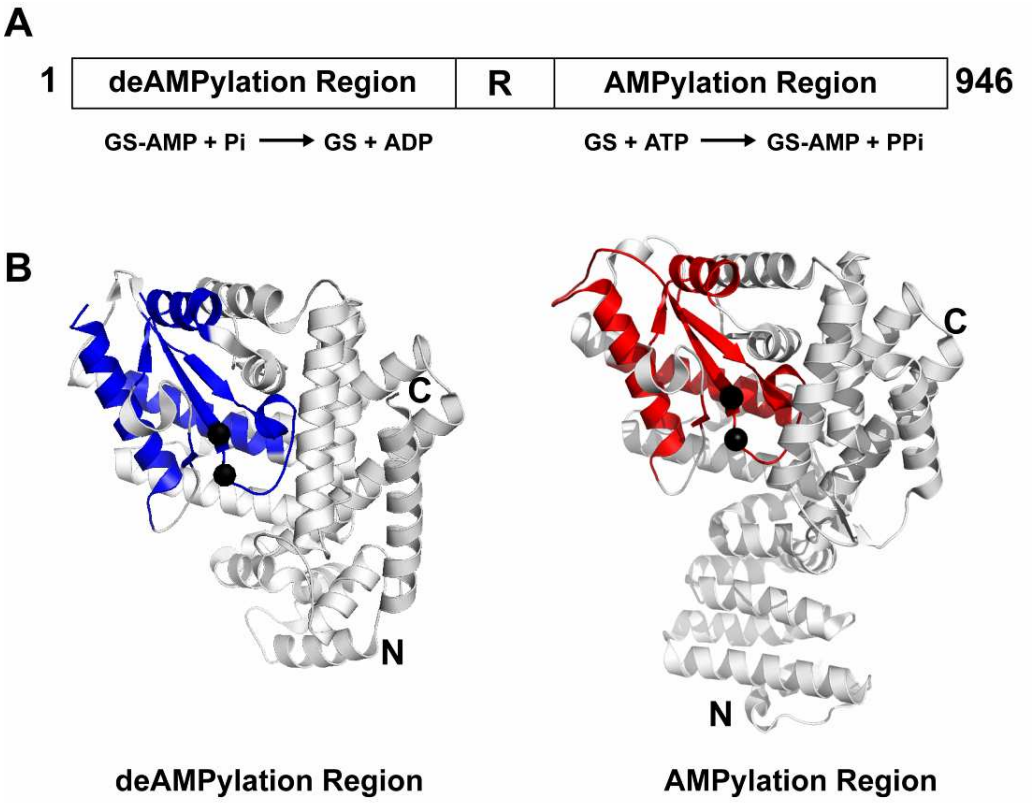

Figure 6. **Domain organization and structures of GS-ATase. (A)** Domain architecture of GS-ATase from *E. coli*. R: Regulatory region of GS-ATase. N-terminal deAMPylation region produces ADP as product. C-terminal AMPylation region uses ATP to AMPylate GS. **(B)** Structures of N- and C-terminal deAMPylation [PDB ID: 1V4A] and AMPylation [PDB ID: 3K7D] region of GS-ATase. The conserved alpha/beta fold belonging to the adenylylation domain family is colored for the deAMPylation region (blue) and for the AMPylation region (red). Black spheres designate the two catalytic aspartate residues of adenylylation domains.

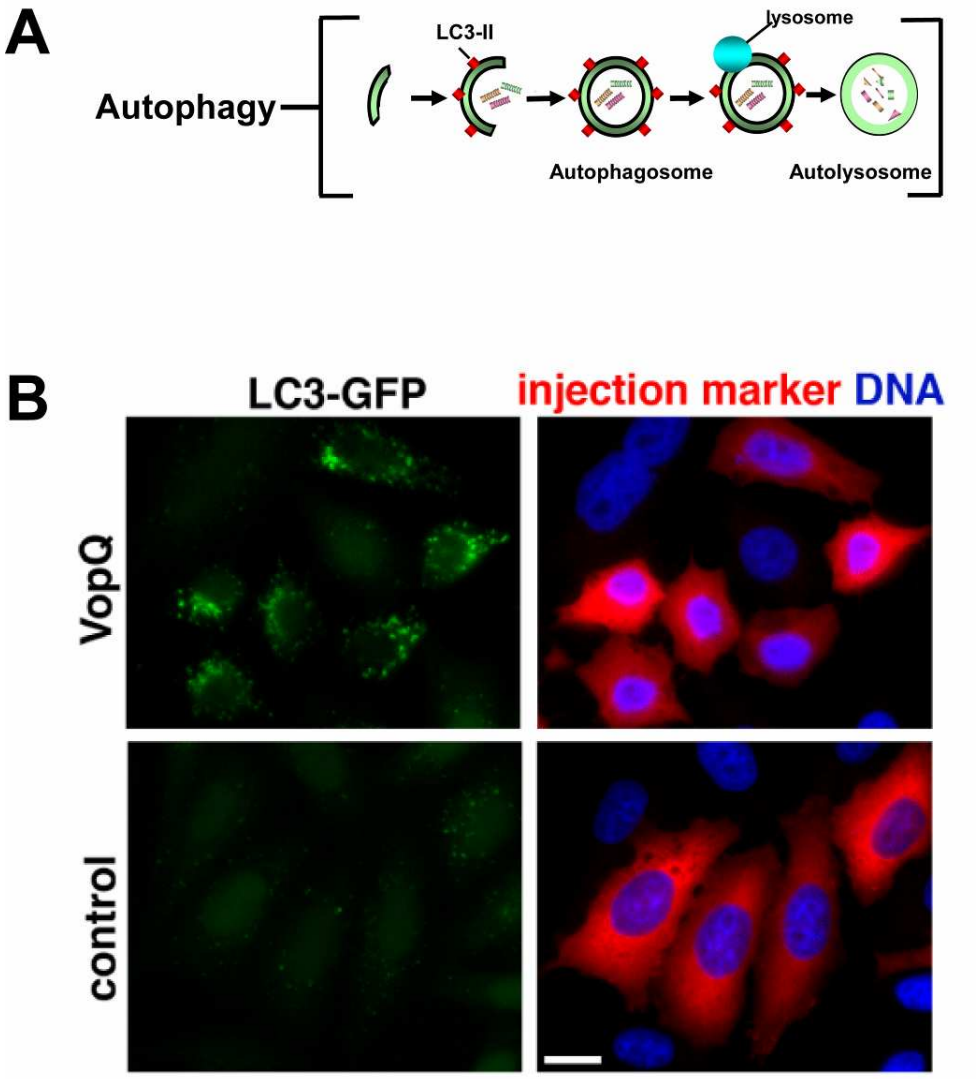

Figure 7. VopQ is necessary and sufficient to induce autophagy in mammalian cells. **(A)** Cartoon of autophagy highlighting engulfment of host proteins, the formation of an autophagosome and degradation of proteins in the autolysosome. **(B)** HeLa cells stably expressing GFP-LC3 were microinjected with recombinant VopQ. GFP-LC3 punctae (green) forms upon VopQ microinjection. DNA (blue) was stained with Hoechst (blue) and Texas-Red dextran (red) was used as an injection marker. Figure 6B adapted from (30).

AMP with its de-AMPylation activity to relieve the inhibition of glutamine synthetase. De-AMPylation mediated by GS-ATase utilizes inorganic phosphate as substrate and generate ADP as product, and requires additional protein components to promote the full activation of GS-ATase de-AMPylation activity (29).

The Vibrio parahaemolyticus effector, VopQ

VopQ was shown to be necessary and sufficient to induce autophagy in mammalian cells but the biochemical and molecular basis of VopQ activity is not known. VopQ contains 492 amino acids and has a molecular weight of 54kDa. BLAST searches showed VopQ is conserved only in other *Vibrio* species. Structural predictions suggest no similarity with any other proteins in the Protein Data Bank (PDB).

Autophagy is a major protein degradation pathway that is independent of the 26S proteasome system and is ubiquitous in eukaryotes (31). Under nutrient stressed conditions autophagy is induced to promote the degradation and recycling of cellular contents in order to generate nutrients for cellular processes (Figure 7A). During autophagy, cytosolic proteins are enveloped and sequestered into double membrane vesicles called autophagosomes. The autophagosome then fuses with the lysosome, forming the autolysosome where the hydrolases provided by the lysosome degrade the proteins (32).

Using HeLa cells that stably express GFP-LC3, Burdette et al 2010, demonstrated that recombinant VopQ microinjected into HeLa cells was sufficient to induce autophagy within minutes (Figure 7B). GFP-LC3 HeLa cells infected with a *Vibrio* strain coding for VopQ induced autophagy, indicated by GFP punctae formation in GFP-LC3 HeLa cells (30). In contrast, infection with VopQ deficient *Vibrio* strain could not induce autophagy. Electron microscopy of *Vibrio* infected RAW 264.7

macrophages further supported the requirement of the VopQ gene for the induction of autophagy.

Goal of Experimental Chapters

Chapter 3 describes the determination of the VopS atomic structure using X-ray crystallography, providing a structural basis for Fic domain-mediated AMPylation. I first describe the overall structure of VopS, and then characterize the functions of conserved Fic motif residues using the solved structure and *in vitro* AMPylation assays. I also describe the regions on VopS that mediate its interaction with ATP and GTPase. In addition, I discuss crystallization attempts to co-crystallize VopS with small molecule and protein substrates. In the discussion section I talk about structures later determined by other groups that provide insights into Fic domain mediated AMPylation.

Chapter 4 describes the kinetic contribution of conserved residues in protein AMPylation mediated by VopS. I discuss the Michaelis-Menten values of VopS AMPylation activity and characterize at a kinetic level residues important for activity. I determine the relative contribution of residues involved in ATP and Cdc42 binding using steady-state kinetic analyses and discuss the reaction mechanism of VopS mediated AMPylation of Rho family GTPases.

Chapter 5 describes the development and characterization of a chemical probe as a molecular tool to study AMPylation. I describe the characterization of the chemical probe using enzymes from both Fic domain (VopS and IbpAFic2) and AT-ase domain (GT-ATase) families. Further, I demonstrate and evaluate the applicability of the probe for the identification and purification of endogenous AMPylation substrates.

In Chapter 6 I explain attempts to solve the crystal structure of VopQ in order to gain insights into the function of VopQ. I discuss purification and crystallization trials of

VopQ. The purification and crystallization of VopQ attempts in this chapter provide a general scheme on how to purify stable and soluble bacterial effector proteins for biochemical characterization.

In the Discussion and Future Directions section, I provide a summary of the experimental findings. I also provide an update on recent findings in the field after the completion of my thesis project. I then discuss questions that are most interesting to me and propose experimental approaches on how to address them.

CHAPTER 2

Materials and Methods

Protein Expression and Purification

VopS-(75-387) was cloned into pGEX-TEV vector to generate a (GST)-tagged VopS-(75-387) construct for purification using GST affinity chromatography (Sigma). The construct was transformed and expressed in *Escherichia coli* BL21 (DE3) cells with 0.4mM final isopropyl β-D-thiogalactopyranoside (IPTG) (Roche Applied Science) for 20 hours at 20°C. Cell pellets were stored at -80°C. Cells were then lysed using a cell disrupter (EmulsiFlex-C5, Avestin Inc.) and affinity purified via GST chromatography (Sigma). The GST tag was removed with overnight cleavage at 4°C using recombinant His-TEV protease. The protein was then loaded onto a 1-ml HiTrapQ HP column (Buffer A 25mM Tris pH 8.0, Buffer B 25mM Tris pH 8.0 + 1M NaCl). VopS fractions were pooled and subsequently loaded onto a HiLoad 16/60 Superdex 75 (GE Healthcare) column in a final buffer containing 10 mM Tris pH 7.5, 50 mM NaCl, 1 mM DTT. Selenomethionine-labeled VopS-(75-387) protein was expressed in the methionine auxotroph *E. coli* strain B834 (DE3) (Novagen). Cells were grown using SelenoMet™ Medium supplemented with selenomethionine (Molecular Dimensions Limited) and VopS was expressed and purified similarly to the native protein. Wild-type GST-VopS and mutant GST-VopS (31-387) constructs were cloned, expressed, and purified with GST affinity chromatography as described above. VopS mutants were made using site-directed mutagenesis that was performed according to manufacturer's instructions (Stratagene). For kinetic studies the wild-type and mutant GST-VopS-(31-387) constructs were expressed and purified using GST affinity chromatography as previously above. The GST-tag was cleaved as described above and the protein was loaded onto a 1ml

HiTrapQ HP column to collect cleaved wild-type or mutant VopS-(31-387). Protein was buffer exchanged in 10 mM Tris (pH 8.0), 100 mM NaCl, 1 mM DTT, 10% glycerol with Amicon 10K concentrators and stored at -80°C. The His-Cdc42-(1-179)-(Q61L) construct was generated using a pET28a vector (Novagen). The clone was transformed in BL21 (DE3) cells and protein expression was similar as described above. His-Cdc42-(1-179)-(Q61L) pellets were purified using immobilized metal ion affinity chromatography (Sigma). Protein substrate was buffer exchanged and stored at -80°C as described previously. Protein concentration was assayed by the Bradford method (Bio Rad) and purity by SDS-PAGE analysis.

Protein constructs for click chemistry experiments include VopS-(31-387) pGEXTEV and VopS-(H348A)-(31-387) pGEXTEV, GST-Fic2 pGEXTEV, GST-Fic2-(H3717A) pGEXTEV, Rab1a (Canis) pET28a, DrrA (1-339) pET28a, RacV12 pET28a, RhoA pGEX-KG, His-Cdc42-(1-179)-(Q61L, T35S) pPROEXHTa, His-Cdc42-(1-179)-(Q61L, T35A) pPROEXHTa and His-Cdc42-(1-179)-(Q61L, Y32A) pPROEXHTa were expressed and transformed in BL21 (DE3) cells. Protein expression was performed overnight at room temperature with 0.4 mM IPTG. Protein pellets were purified using glutathione-agarose (Sigma) or immobilized metal ion affinity chromatography (Sigma). VopS active and VopS mutant constructs were cleaved overnight and purified with HiTrap QHP column as mentioned above. Proteins were buffer exchanged in 20 mM Tris pH 8.0, 100 mM NaCl, 10 % glycerol, 1 mM DTT with an Amicon 10K concentrator and stored at -80°C. Protein concentrations were assayed by the Bradford method (Bio-Rad) and purity by SDS-PAGE analysis.

Protein expression of GST-VopQ (pGEX-TEV) constructs were expressed in BL21 overnight in 0.4mM IPTG at room temperature unless noted. GST-VopQ constructs were purified by GST affinity chromatography as described above and cleaved overnight at 4°C with His-TEV protease. Where noted a QHP HiTrap Column was used

for further purification as described above. VopQ QHP fractions that were loaded onto Superdex 75 used buffer with 10mM Tris pH7.5 50mM NaCl.

For coexpression studies, VopQ (pPROEXHta) constructs and VP1682 (encodes His tag while VopQ (pET28a) has no tag unless noted. Two types of VP1682 chaperone were made. The VP1682 (pPROEXHta) contains a His tag while VP1682 (pET28a) is tagless. Also VopQ (pPROEXHTa) encodes for a His tag while the VopQ (pET28a) is tagless.

Protein Methylation

Protein in 1X PBS at a concentration of 2 – 4 mg/ml was labeled with 20 µl of dimethylamine borane complex to 1 ml of protein. 40 µl of 1.0 M formaldehyde was added per ml of protein sample. Aluminum foil was wrapped around the protein sample to exclude light during the reaction and the protein was incubated for two hours at 4 °C. An additional 20 µl of dimethylamine borane complex and 40 µl of 1 M formaldehyde per ml of protein sample was added and incubated for two hours at 4 °C in the dark. 10 µl of dimethylamine borane complex was added and incubated overnight at 4 °C. Protein was loaded on a gel filtration column in 10mM Tris pH 7.5, 150mM NaCl to remove methylation reagents and to buffer exchange the sample.

Limited Proteolysis

Stock trypsin dissolved at 1 mg/ml was serially diluted (1:10, 1:100, and 1:1000) in 10 mM HCl. 1 µl of trypsin dilution was added to recombinant VopS-(31-387) in 50µl volume. Each reaction was incubated at room temperature for 30 minutes and 5µl of 0.1 mg/ml pefabloc (Roche) to stop the proteolytic reaction. Reactions were analyzed by SDS-PAGE, stained with Colloidal Blue (Invitrogen) and samples were N-terminally sequenced. Samples subjected to MALDI-TOF to determine the total molecular weight.

Limited proteolysis of papain was performed at 1:10, 1:100, and 1:1000 dilution from a 1mg/ml stock. 1 µl of papain was added in 30 µl volume. Protein was incubated at room temperature for 30 minutes and stopped with 5X sample buffer and boiled 10 minutes at 95 °C.

Protein Crystallization

Initial screening of purified VopS-(75-387) at 12 mg/ml was performed in sitting-drop vapor diffusion mode in 96-3 well Intelli-Plates (Art Robbins Instruments) at 20°C. Initial hits were seen in the PEGs Suite (Qiagen). Crystal condition #37 with 0.1M HEPES (pH 7.5) and 25% (w/v) PEG 3000 was selected for optimization. Optimized crystallization conditions were 0.1 M HEPES (pH 7.25 - pH7.5) and 21% PEG 3500 at 4°C with 6 mg/ml protein. Initial screening for VopQ constructs were performed in sitting-drop vapor diffusion mode in 96-3 Intelli-Plates as described above.

Data Collection, Structure Determination, and Refinement

Diffraction data were collected at the Structural Biology Center at the Advanced Photon Source (Argonne National Laboratory). Data sets were indexed, integrated, and scaled using the HKL-3000 program package (33). The native VopS-(75-387) crystal had the symmetry of space group P21 with unit cell parameters a= 66.67 Å, b= 62.32 Å, c= 75.76 Å, β= 91.3º, diffracted X-rays to a minimum Bragg-spacing, dmin, of 1.80 Å, and contained two molecules in the asymmetric unit. Phases were obtained from a selenomethionyl-substituted protein crystal by the single-wavelength anomalous dispersion (SAD) method using X-rays with energy near the selenium K absorption edge. Selenium sites were located using the program SHELXD (34). Phases were refined with MLPHARE (35) resulting in an overall figure of merit of 0.17 for data between 30.0 and 2.28 Å. Phases were improved via density modification and two-fold symmetry

averaging in the program dm (36), resulting in a figure of merit of 0.65. An initial model containing ~93% of all residues was constructed automatically using the program ARP/wARP (37). Manual model building was performed with the program Coot (38). Refinement was performed with native data to a resolution of 1.8 Å using the program PHENIX (39) with a random 5.07% of all data set aside for Rfree calculation. Molprobity was used during refinement to check on the model quality (40). All protein structure figures were constructed using the MacPyMOL program (41).

Filter Binding Assays

VopS-(31-387) activity was assayed by AMPylation of His-Cdc42-(1-179)-(Q61L) with ^{32}P-α-ATP purchased from Perkin Elmer. P81 Whatman filters were used for filter binding. Buffer conditions were 20 mM Hepes pH 7.4, 5 mM $MgCl_2$, 100 mM NaCl, 1 mg/ml BSA, and 1 mM DTT. Adenosine triphosphate (ATP) was dissolved in 20 mM Hepes pH 7.4. Reactions were performed in triplicate at 25ºC and initiated with ^{32}P-α-ATP (30-120 cpm/pmol) and time points taken at 15, 30, 45, 60, 75 for mutants and additional time point at 90 seconds for the wild-type protein. A 30µl reaction volume per time point collected was used. Reactions were stopped by pipetting 20µl of the reaction onto P81 Whatman filters. Filters were then immediately immersed in 75 mM phosphoric acid and stirred in a beaker on a platform shaker. Filters were subsequently washed three times with 75 mM phosphoric acid, rinsed with acetone and allowed to air dry. Each filter was placed in a vial with scintillation fluid and counts per minute were measured using a Beckman LS6500 scintillation counter. To determine apparent steady-state kinetic values for ATP with VopS-(31-387), His-Cdc42-(1-179)-(Q61L), protein substrate was kept at a constant concentration of 600 µM while varying ATP concentrations (40, 80, 120, 160, 250, 400, 700, 1000, 1500, 2000 µM) were studied. Apparent kinetic values for ATP with mutant VopS constructs were performed at 600 µM

protein substrate, varying ATP with concentrations of 100, 200, 300, 500, 750, 1000, 1500, and 2000 μM. Assays of the R299A mutant had the same ATP concentrations as other VopS mutants, but an additional measurement at 150 μM ATP was taken. To determine apparent steady-state kinetic values for His-Cdc42-(1-179)-(Q61L), the ATP concentration was fixed at 2000 μM while varying His-Cdc42-(1-179)-(Q61L) concentrations (50, 75, 100, 150, 200, 300, 500, 800, 1200 μM). The R299A mutant was assayed at 2000 μM ATP, while the following concentrations were used for the protein substrate: 50, 75, 100, 200, 300, 500, 800, and 1200 μM. To determine the reaction mechanism, ATP concentrations were fixed (50, 75, 100, 200, 1000 μM) while varying the concentration of His-Cdc42-(1-179)-(Q61L) (50, 75, 100, 200, 500 μM).

Inhibition studies with AMP, AMPCPP, and PPi were performed under steady-state conditions.. For AMP and AMPCPP inhibition, 5 nM VopS, 160 μM ATP and 300 μM Cdc42 were incubated with different concentrations of AMP or AMPCPP. Reactions were stopped at 75 seconds and performed in duplicates. To test PPi inhibition, 5nM VopS, 1 mM ATP, at various Cdc42 (60 μM, 100 μM, 150 μM, 200 μM, 500 μM), and PPi (0 μM, 150 μM, 400 μM, 900 μM) concentrations were used.

Kinetic Data Analysis

Single-substrate kinetic measurements were fitted into the Michaelis-Menten equation (Equation 1) using GraphPad Prism 5,

$v = V_{max}[S]/(K_m + [S])$ (Eq. 1)

Bisubstrate kinetic studies fitted best to a (sequential) random, rapid-equilibrium model using Sigma Plot 11.0 Enzyme Kinetics 1.3. Equations for random rapid-equilibrium (Eq. 2), ordered rapid-equilibrium (Eq. 3), and ping pong (Eq. 4) are listed below, respectively.

$v = V_{max}[A][B]/(\alpha K_aK_b+K_b[A]+K_a[B]+[A][B])$ (Eq. 2)

$v = V_{max}[A][B]/(K_aK_b + K_b[A]+[A][B])$ (Eq. 3)

$v = V_{max}[A][B]/(K_a[B]+ K_b[A]+[A][B])$ (Eq.4)

V_{max} is maximum velocity and A and B represents substrates. K_a and K_b are Michaelis-Menten constants of the substrates and α is the interaction factor between A and B.

***In vitro* gel based AMPylation Assays**

AMPylation assays used purified recombinant protein with 200 μM cold ATP and 0.1-0.4 μCi ^{32}P-α-ATP. 100 μM of His-Cdc42-(1-179)-(Q61L) was incubated with 5 nM of GST-VopS-(31-387) wild-type or mutant protein. Assays were performed at 25ºC in buffer containing 20 mM Hepes pH 7.4, 5 mM $MgCl_2$, 100 mM NaCl, 0.1mg/ml BSA, 1 mM DTT. Reactions were stopped with SDS sample buffer. Samples were boiled and separated by SDS-PAGE and visualized by autoradiography.

Cell Lysate preparation for Click Chemistry Reactions.

HeLa lysates were prepared from 10 cm 80-90% confluent plate and washed 2 X 2.5ml cold 1X PBS. Cells were washed 1X 2.5ml 1X PBS + Roche protease inhibitor tablet or cells were scraped and then Roche protease inhibitor from 100X Roche tablet stock was added to eppendorf tube. A final 0.5% Triton was added from 10% Triton stock and allowed to sit on ice from 30 to 60 minutes. Cells were spun 20,000g 4°C and supernatant were dialyzed 2 times in 1 Liter 1X PBS + 0.5% Triton in a 2000 kDa MWCO Slide-A-Lyzer (Promega).

***In vitro* AMPylation reactions with N6pATP**

N6pATP and azido-rhodamine was generated and provided by Markus Grammel and Howard Hang at Rockefeller (42). Reactions with N^6pATP at 100μM were

performed in 1X PBS pH 7.4 + 5mM $MgCl_2$ with a total volume of 15 µl and incubated at 1 hour 30°C and reactions were then stopped by methanol/chloroform precipitation. Tubes at room temperature were filled with 85 µl water to 100 µl. 400 µl of cold methanol and 100 µl of chloroform was added and the mix was vigorously vortexed. Then, 300 µl water was added, the mix was vortexed again and centrifuged at 20,000 g for 5 min at room temperature (RT). The upper phase was carefully removed without disturbing the interphase. Tubes were placed on ice and add 1 ml of ice-cold acetone was added and the samples were vortexed. The samples were centrifuged at 20,000g for 10 min at 4 °C. Supernatant was decanted and the samples were dried for about 5 min at RT or until dry. The resulting protein pellets were dissolved in 15 µl 1X PBS + 4% SDS

For in-gel fluorescence detection of N^6pATP modified substrate proteins, resolubilized protein pellets from AMPylation reactions were reacted with azido-rhodamine (az-rho).[6] For this purpose a click-chemistry master mix was prepared with 7.5 µl 4ST, 0.25 µl 10 mM azido-rhodamine, 0.5 µl 50 mM TCEP (tris(2-carboxyethyl)phosphine), 1.25 µl 2 mM TBTA (tris((1-benzyl-1H-1,2,3-triazol-4-yl)methyl) amine) in 1:4 DMSO:n-butanol and 0.5 µl $CuSO_4$ per individual sample. 10 µl click-chemistry master mix was added to each individual protein sample to yield a total volume of 25 µl with 0.1 mM az-rho, 1 mM TCEP, 0.1 mM TBTA and 1 mM $CuSO_4$. Click-chemistry was carried out for 1 hour at room temperature. Click-chemistry was stopped by addition of 10 µl sample buffer. Samples were incubated at 95 °C for 5 min and run for analysis on SDS-PAGE).

Protein gels were rinsed with deionized water and incubated for multiple hours in 50% water, 40% methanol, 10% acetic acid at 4°C. Before in-gel fluorescence analysis, protein gels were transferred to deionized water and incubated for at least 30 min at room

temperature. Protein gels were scanned on an Amersham Bioscience Typhoon 9400 variable mode imager (excitation 532 nm, 580 nm filter, 30 nm band-pass).

List of Primers

NAME	SEQUENCE
143 N107VP1680 HINDIII BAMHI FWD	ATCGAAGCTTGGATCCGAGCGAAAAATCTTGGAAGTG
165 CDC42 NCO1 FWD	ATCGCCATGGATGCAGACAATTAAGTGTGTTG
168 C12 CDC42 XHO1 REV	ATCGCTCGAGTCAAGGCTCCAGGGCAGCCAAT
171 N74 VP1686 BAMH1	ATCGGGATCCGCGATCACAAAAGCAGTGTTTG
172 CDC42 Q61L FWD	GGACTTTTTGATACTGCAGGGCTTGAGGATTATGACAA TTACGA
176 CDC42 Q61L REV	TCGTAATCTGTCATAATCCTCAAGCCCTGCAGTATCAA AAAGTCC
188 VOPS F350A	GGT GTG ATT GGT TAT CAC GGC GCT ACC GAT GGC AAC GGA CGC ATG
189 VOPS F350A	CAT GCG TCC GTT GCC ATC GGT AGC GCC GTG ATA ACC AAT CAC ACC
194 VOPS N354A FWD	GGTTATCACGGCTTTACCGATGGCGCCGGACGCATGG GGCGCATG
195 VOPS N354A REV	CATGCGCCCCATGCGTCCGGCGCCATCGGTAAAGCCG TGATAGCC
196 N76VOPQ BAMH1 FWD	ATCGGGATCCCAGGCACTGCTTAAGGAAGAA
198 VOPS R296A	CGT TTG GTA CCA AAT GTA GAA GCA GAT TAC CGT GGG CCA AAT ATC
199 VOPS R296A	GATATTTGGCCCACGGTAATCTGCTTCTACATTTGGTA CCAAACG
200 VOPS R299A	CCA AAT GTA GAA CGC GAT TAC GCA GGG CCA AAT ATC TCT GGT GGC
201 VOPS R299A	GCCACCAGAGATATTTGGCCCTGCGTAATCGCGTTCTA CATTTGG
202 VOPS Y298A	GTA CCA AAT GTA GAA CGC GAT TTC CGT GGG CCA AAT ATC TCT GGT
203 VOPS Y298A	ACCAGAGATATTTGGCCCACGGAAATCGCGTTCTACAT TTGGTAC
212 N78 VOPS BAMH1 FWD	ATCGGGATCCGCA GTG TTT GAC AAC GAA CAG
215 VOPS D352A FWD	ATT GGT TAT CAC GGC TTT ACC GCA GGC AAC GGA CGC ATG GGG CGC
216 VOPS D352A REV	GCG CCC CAT GCG TCC GTT GCC TGC GGT AAA GCC GTG ATA ACC AAT
227 VOPS L308A FWD	CCA AAT ATC TCT GGT GGC ACA GCA CCA TCT AGT ATT GGT GGG GAA
228 VOPS L308A REV	TTCCCCACCAATACTAGATGGTGCTGTGCCACCAGAGA TATTTGG
230 VOPS N100	ATCGGGATCCCAGATGTTGTTTAAAGACGCG

BAMH1 FWD	
233 FIC2 BAMH1 FWD	ATCGGGATCCAAATCATCTCCGCAAGAGGGA
235 FIC2 XHO1 REV	ATCGCTCGAGTTATTTTTTTGCCAACTCTTTTAA
239 VOPS N150 BAMH1 FWD	ATCGGGGATCCTCTGGCAAGAACCTCAAAGCG
240 VOPS N200 BAMH1 FWD	ATCGGGATCCGTGATGCACCTAAAAGCAGCG
241 VOPS N175 BAMH1 FWD	ATCGGGATCCGGTGGTGGCGTGGCGGCTGCT
242 VOPS N225 BAMH1 FWD	ATCGGGATCCGAAGGTCTTCCTGAAGACGCT
245 CDC42 T35A FWD	TTTCCATCGGAATATGTACCGGCGGTTTTTGACAACTA TGCAGTC
246 CDC42 T35A REV	CCA AAT ATC TCT GGT GGC ACA CGC CCA TCT AGT ATT GGT GGG GAA
280 RAC1 T35S FWD	TTTCCTGGAGAATATATCCCTGCCGTCTTTGACAATTA TTCTGCC
281 RAC1 T35S REV	GGCAGAATAATTGTCAAAGACGGCAGGGATATATTCT CCAGGAAA
340 N80 VOPQ BAMH1 FWD	ATCGGGATCCAAGGAAGAAAAACCAGAAACC
354 VOPQ XHO1 REV	ATCGCTCGAGTTAAATCCAGCCTTCGGCTAAG
389 N15 VOPQ NCO1 FWD	ATCGCCATGGCCATGAACACGATTCAACCACTG
392 N247 VOPQ NCO1 FWD	ATCGCCATGGCCGCCACACTTGGTATGGCG
403 VOPQ N28 NCO1	ATCGCCATGGCCCGAGTAATCAGCAAAAGAGGC
424 N101 VOPQ NCO1 FWD	ATCGCCATGGGCGGAGAACCATTAACCGAGCGA
425 VOPQ CBD FWD	ATTGCCGCGCAAAAAGATGACCGAGGAGAACCATTAA CCGAGCGAAAA
426 VOPQ CBD REV	TTTTCGCTCGGTTAATGGTTCTCCTCGGTCATCTTTTTG CGCGGCAAT
427 VOPQ FL NCO1 FWD	ATCGCCATGGCCATGGTGAATACAACGCAAAAA
428 1682 NCO1 REV NO STOP	ATCGCCATGGGCACACGCAGTGGTTGAACATG
431 1682 NCO1 NO START	ATCGCCATGGGCAACACGATTCAACCACTGCTC
432 1682 XHO1 VERSION 2	ATCGCTCGAGCTACACACGCAGTGGTTGAACATGAC
436 N101 VOPQ ALGINO NCO1 FWD	ATCGCCATGGGCGGTGAGCCACTTACTGAGCGT
438 VOPQ ALGINO XHO1 REV	ATCGCTCGAGTTACACCCAGCCTTCTGCCAA

442 C15 VP1682 XHO1 REV	ATCGCTCGAGTTATGAAAGATCTTGATCCAACAT
452 VOPQ HARVEYI BB120 XHO1 REV	ATCGCTCGAGTTAAATCCAGCCTTCTGCTAA
453 VOPQ HARVEYI BB120 N163 NCO1 FWD	ATCGCCATGGGCGGTGAACCGCTAACGGAACGA
467 N29 VOPQ NCO1 FWD	ATCGCCATGGGCGTAATCAGCAAAAGAGGCGAA
476 NC17 VA1682 XHO1 REV	ATCGCTCGAGCTACGCTTGATCCAACATTGCGAA
486 C10VP1682 XHO1 REV	ATCGCTCGAGCTAATGATTTGAAGTGGCTGAAAG
487 C5VP1682 XHO1 REV	ATCGCTCGAGCTAAACATGACTTTGTTCATGATT
489 C15VP1682 SPE1 REV NO STOP	ATCGACTAGTTGAAAGATCTTGATCCAACAT
490 C10VP1682 SPE1 REV NO STOP	ATCGACTAGTATGATTTGAAGTGGCTGAAAG
491 N101 VOPQ SPE1 FWD	ATCGACTAGTGGAGAACCATTAACCGAGCGA
493 C15VP1682 LINKER 1 SPE1 REV	ATCGACTAGTGCTGGAGAGGGACGAATCAGATGAGAG GTTTGAAAGATCTTGATCCAA
495 C15VA1682 XHO1 REV	ATCGCTCGAGCTATGGAATCGCTTGATCCAACAT
496 C15VA1682 SPE1 REV NO STOP	ATCGACTAGTTGGAATCGCTTGATCCAACAT
499 VOPQ GS TAIL XHO1REV	ATCGCTCGAGTTAGCCCGAGCCCGAGCCAATCCAGCC TTCGGCTAAG
506 CDC42 Y32A FWD	AACAAATTTCCATCGGAAGCCGTACCGACTGTTTTTGA CAAC
507 CDC42 Y32A REV	GTTGTCAAAAACAGTCGGTACGGCTTCCGATGGAAAT TTGTT
510 VOPQ K363Y FWD	ATGAGTTCACCTCAAGCGCCATACGAAGCGCAAGAGA TGGAAGCC
511 VOPQ K363Y REV	GGCTTCCATCTCTTGCGCTTCGTATGGCGCTTGAGGTG AACTCAT
514 DRRA NDE1 FWD	ATCGCATATGAGCATAATGGGGAGAATT
516 C340 DRRA XHO1 REV	ATCGCTCGAGACGTTGAACACCCAGCTCTCT
520 RAB1A DOG NDE1 FWD	ATCGCATATGTCCAGCATGAATCCCGAA
521 RAB1A DOG XHO1 REV	ATCGCTCGAGTTAGCAGCAACCTCCACCTGA

List of Constructs

Construct	Description	Source
pGEXTEV VopS-(31-387)	Soluble active construct of VopS; BamH1 and Xho1	(15)
pGEXTEV VopS-(31-387)-(H348A)	Histidine 348 to alanine mutant; BamH1 and Xho1	(15)
pGEXTEV VopS-(31-387)-(F350A)	Phenylalanine 350 to alanine mutant	This study
pGEXTEV VopS-(31-387)-(D352A)	Aspartate 352 to alanine mutant	This study
pGEXTEV VopS-(31-387)-(N354A)	Asparagine 354 to alanine mutant	This study
pGEXTEV VopS-(31-387)-(R356A)	Arginine 356 to alanine mutant	This study
pGEXTEV VopS-(31-387)-(R299A)	Arginine 299 to alanine mutant	Cloned by Soonjoung Kim (rotation student)
pGEXTEV VopS-(31-387)-(L308A)	Leucine 308 to alanine mutant	This study
pGEXTEV VopS-(75-387)	Construct used to solve VopS structure; BamH1 and Xho1	This study
pGEXTEV VopS-(75-387)-(H348A)	Histidine 348 to alanine mutant; BamH1 and Xho1	This study
pGEXTEV VopS-(79-387)	BamH1 and Xho1	This study
pGEXTEV VopS-(79-387)-(H348A)	BamH1 and Xho1	This study
pGEXTEV Fic2	Fic2 domain from Histophilus *somni;* BamH1 and Xho1	This study
pGEXTEV Fic2-(H3717A)	Histidine 3717 to alanine mutant; BamH1 and Xho1	This study
pGEXTEV VopQ-(108-492)	BamH1 and Xho1	This study
pGEXTEV VopQ-(77-492)	BamH1 and Xho1	This study
pGEX-KG RhoA	Human RhoA construct	Gift of N. Alto (UTSW)
pGEXTEV VopS-(101-387)	BamH1 and Xho1	This study
pGEXTEV VopS-(151-387)	BamH1 and Xho1	This study
pGEXTEV VopS-(176-387)	BamH1 and Xho1	This study
pGEXTEV VopS-(201-387)	BamH1 and Xho1	This study
pGEXTEV VopS-(226-387)	BamH1 and Xho1	This study
pET28a VopQ-(29-492)	Nco1 and Xho1; tagless construct	This study
pET28a VP1682-(1-137)	Nco1 and Xho1; tagless construct	This study

pET28a Cdc42-(1-179)-(Q61L)	Dominant active Cdc42; Kan resistance; Nde1 and Xho1	This study
pET28a Rab1a	Rab1a from *Canis;* Nde1 and Xho1	This study
pET28a DrrA-(1-339)	N-terminal construct containing AMPylation region; Nde1 and Xho1	This study
pET28a RacV12	Dominant active construct of Rac1	(15)
pPROEXHTa Cdc42-(1-179)-(Q61L, T35S)	Dominant active with threonine 35 to serine mutation; Nco1 and Xho1	This study
pPROEXHTa Cdc42-(1-179)-(Q61L, T35A)	Dominant active with threonine 35 to alanine mutation; Nco1 and Xho1	This study
pPROEXHTa Cdc42-(1-179)-(Q61L, Y32A)	Dominant active with tyrosine 32 to alanine mutation; Nco1 Xho1	This study
pPROEXHTa Cdc42-(1-179)-(Q61L)	Dominant active Cdc42; Amp resistance; Nco1 and Xho1	This study
pPROEXHTa VopQ-(248-492)	Nco1 and Xho1	This study
pPROEXHTa VopQ-(29-492)	Nco1 and Xho1	This study
pPROEXHTa VopQ-(101-492)	Nco1 and Xho1	This study
pPROEXHTa VP1682-(1-137)	Nco1 and Xho1	This study
pPROEXHTa VP1682-(1-152)	Nco1 and Xho1	This study
pPROEXHTa VopQ *alginolyticus*-(102-492)	Nco1 and Xho1; *V. alignolyticus*	This study
pPROEXHTa VA1682-(1-135)	Nco1 and Xho1; *V. alignolyticus*	This study
pPROEXHTa Fusion VP1682-(1-137)-VopQ-(102-492)	Protein fusion; Chaperone cloned into Nco1 and Spe1; Effector cloned into Spe1 and Xho1	This study
pPROEXHTa Fusion VP1682-(1-142)-VopQ-(102-492)	Protein fusion; Chaperone cloned into Nco1 and Spe1; Effector cloned into Spe1 and Xho1	This study
pPROEXHTa Fusion GS Tail VP1682-(1-137)-VopQ-(102-492)	Protein fusion; Chaperone cloned into Nco1 and Spe1; Effector cloned into Spe1 and Xho1	This study

CHAPTER 3

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF THE FIC PROTEIN, VOPS

Introduction

VopS AMPylates the conserved threonine of Rho GTPases to disrupt the actin cytoskeleton. Several Fic protein structures were previously solved and deposited in the PDB database by the Protein Structure Initiative, but when this project was initiated there had been no descriptions for the AMPylation mechanism by a Fic domain protein (22). Here we provide a structural understanding on how VopS mediates the AMPylation of Rho GTPases. The role of the Fic motif HPFX(D/E)GN(G/K)R residues in AMPylation were not clear except for the conserved histidine. Mutation of the histidine to alanine abolished VopS biochemical activity. In this chapter, structure and functional characterization of VopS protein and Fic motif residues are described. In addition, comparisons of VopS with deposited Fic protein structures are made to determine the ATP and protein substrate binding site for VopS.

Results

VopS-(75–387) Is Proteolytically Stable and Catalytically Active

To identify the optimal fragment of VopS protein for crystallization trials, we used limited proteolysis with the catalytically active VopS-(31–387) that has the first 30 amino acid residues of the protein removed to increase protein solubility. Limited proteolysis with trypsin uncovered a relatively stable cleavage product (Figure 8A). N-

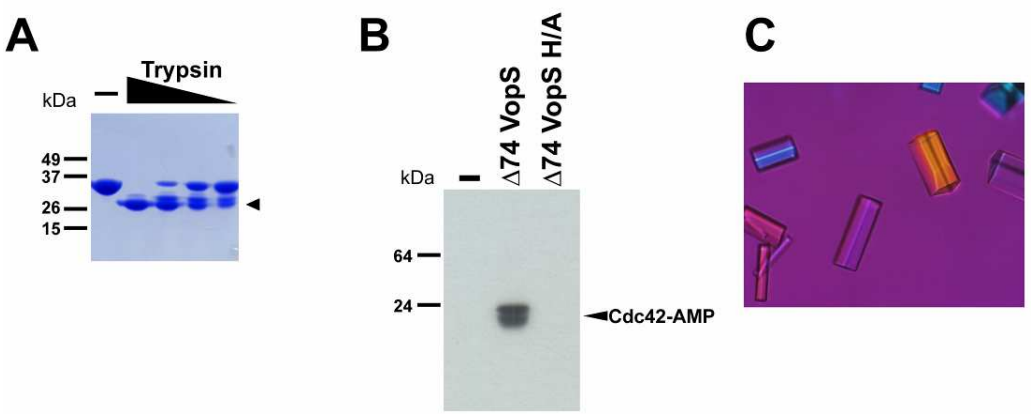

Figure 8. Expression and crystallization of VopS. (A) Limited proteolysis of VopS-(31-387) with decreasing concentrations of trypsin reveal proteolytically stable product indicated with arrow. Stable product was sent for N-terminal sequencing. **(B)** *In vitro* AMPylation assay with Δ74 VopS and Δ74 VopS H348A mutant. Reactions were incubated with 100 μM His-Cdc42-(1-179)-(Q61L) and ^{32}P-α-ATP at 25ºC alone or with Δ74 VopS or Δ74 VopS H348A mutant. The assay was stopped at 90 seconds with loading buffer. **(C)** Optimized crystals of VopS-(75-387) grown in 0.1 M HEPES (pH 7.25 - pH7.5) and 21% PEG 3500.

terminal sequencing and mass spectrometry revealed that the cleavage after lysine 74 yielded the stable product. VopS-(75–387) and VopS-(75–387)-H348A, a catalytically inactive mutant where the conserved histidine 348 of the Fic motif was mutated to an alanine, were cloned, purified from *E. coli*, and assayed for *in vitro* enzymatic activity. Consistent with previous observations, VopS-(75–387), but not the mutant VopS-(75–387)-H348A, was catalytically active in an *in vitro* AMPylation assay using ^{32}P- α-ATP and dominant active His-Cdc42-(1–179)-Q61L as substrates (Figure 8B).

Overall Structure of VopS

The enzymatically active VopS-(75–387) was purified and used for crystallization trials. After optimization of crystallographic conditions, VopS-(75–387) crystals grew overnight at 4°C in 0.1 M HEPES, pH (7.25–7.5) and 21% polyethylene glycol 3500 (Figure 8C). The structure was phased by single-wavelength anomalous dispersion with selenomethionine-substituted VopS-(75–387) protein crystals that diffracted to 2.30 Å resolution. Refinement was performed on native data to a resolution of 1.8 Å, and the model contains two VopS-(75–387) monomers (molecule A with residues 88–387 and molecule B with residues 80–387) and 514 water molecules. The R_{work} is 17.2%, and the R_{free} is 22.4%. Phasing and model refinement statistics are provided in Table 3.

The VopS-(75–387) structure is predominantly α-helical and is divided into two subdomains, an N- terminal arm and a C-terminal catalytic region (Figure 9A). The N-terminal arm region (*red*) is comprised of 9 α-helices (H1-H9) that consists of residues 75-198. The N-terminal subdomain is not conserved in other solved structures of Fic domains, and does not appear to be homologous to any existing structures. Hydrophobic side chains (depicted as *spheres*) on the N-terminal subdomain α-helices H6, H7, and H9 pack against hydrophobic side chains of a3 and a4 to stabilize and maintain Fic domain

Table 3. Data collection, phasing and refinement statistics for VopS Structure

Data collection		
Crystal	Native	SeMet[a]
Space group	$P2_1$	$P2_1$
Unit cell parameters (Å)	a = 66.67, b = 62.32, c = 75.76, b = 91.3°	a = 66.33, b = 61.96, c = 76.42, b = 91.1°
Energy (eV)	12,684	12,684
Resolution range (Å)	48.2 - 1.80 (1.87-1.80)	48.1 - 2.28 (2.36-2.28)
Unique reflections	57,468 (2,849)	24,947 (670)
Multiplicity	4.1 (3.6)	3.4 (2.0)
Data completeness (%)	99.9 (100.0)	89.9 (49.6)
R_{merge} (%)[b]	8.80 (84.0)	13.9 (57.3)
I/σ(I)	7.0 (1.7)	8.9 (2.2)
Wilson B-value (Å^2)	22.9	35.6
Phase determination		
Anomalous scatterers	21 out of 24 possible sites	
Figure of merit (30.0-2.28 Å)	0.17	
Refinement statistics		
Resolution range (Å)	48.2-1.80 (1.87-1.80)	
No. of reflections R_{work}/R_{free}	57,381/2,909 (5,363/263)	
Data completeness (%)	99.7 (99.0)	
Atoms (non-H protein/solvent)	5,178/514	
R_{work} (%)	17.2 (24.2)	
R_{free} (%)	22.4 (31.1)	
R.m.s.d. bond length (Å)	0.012	
R.m.s.d. bond angle (°)	1.089	
Mean B-value (Å^2) (protein/solvent)	33.2/36.8	
Ramachandran plot (%) (favored/disallowed)[c]	98/0.0	
Maximum likelihood coordinate error	0.28	
Missing residues	Chain A: 75-85, Chain B: 75-76, 387	
Alternate conformations	None	

Data for the outermost shell are given in parentheses.

[a]Bijvoet-pairs were kept separate for data processing

[b]$R_{merge} = 100\ \Sigma_h\Sigma_i|I_{h,i} - \langle I_h\rangle|/\Sigma_h\Sigma_i I_{h,i}$, where the outer sum (h) is over the unique reflections and the inner sum (i) is over the set of independent observations of each unique reflection

[c]As defined by the validation suite MolProbity (Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S. and Richardson, D.C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. **35**, W375-W383.).

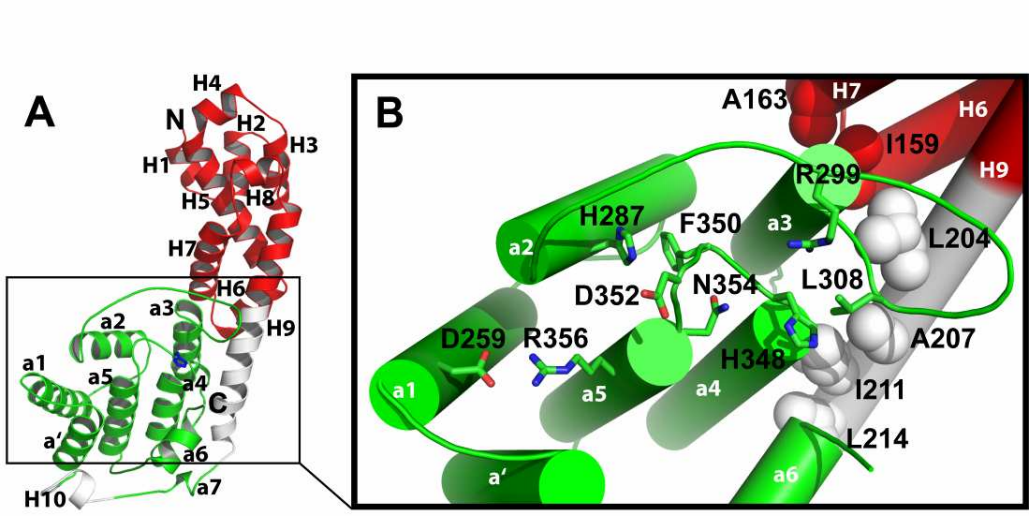

Figure 9. Crystal structure of VopS. (A) Ribbon diagram of VopS (75-387) determined at 1.8 Å resolution. The N-terminal subdomain of VopS contains helices H1-H9 (red) that are not conserved among Fic-domain-containing proteins. The C-terminal subdomain includes the later half of H9 and the pi helix H10 (white) and the structurally conserved Fic domain (green). Highly conserved histidine 348 (blue) is shown in the residues. N-terminal subdomain hydrophobic residues from H6, H7, and H9 (spheres) structurally stabilize the C-terminal Fic domain of VopS. Fic motif (HPFx[D/E]GN[G/K]R) residues and interacting residues (His287 and Asp259) are displayed as sticks. Leu308 and Arg299 are positioned in the structurally conserved hairpin loop of the Fic domain. In this figure and all that follow, nitrogen atoms are colored blue, oxygen atoms red.

tertiary structure (Figure 9B). The N-terminal subdomain is required for full activity; the removal of 150 N-terminal residues results in the loss of catalytic activity *in vitro* (Figure 10). With nine turns, the H9 helix is the longest helix and serves as a backbone that spans both N- and C-terminal subdomains (Figure 9A).

The C-terminal subdomain possesses 8 α-helices (*green*) that are structurally similar to other Fic domains (Figure 9A, *green*). The evolutionarily conserved core of the VopS Fic domain comprises two internal helices, a4 and a5, that are encircled by helices a′, a1, a2, a3, a6, and a7. The VopS Fic domain is decorated by an additional helix H10 and the latter part of the backbone helix H9.

VopS Fic Domain Comparisons to Existing Structures

Four Fic protein structures have been determined by the Protein Structure Initiative and only one of these structures have been published [22(43). These four Fic structures are from *H. pylori* [PDB ID 2F6S], *S. oneidensis* [PDB ID 3EQX], *B. thetaiotaomicron* [PDB ID 3CUC], and *N. meningitidis* [PDB ID 2GO3]. The biochemical activity and the molecular function of these four Fic structures are unknown. Structures of Fic family protein BepA from *Bartonella henselae* (*B. henselae*) were also deposited into the Protein Data Bank (44). The BepA structures from Tilman Schirmer's group were left unpublished until recently (44). The sequence identity of Fic domains is lower than 20% among the known structures, providing a wide evolutionary sampling of Fic domains for comparison.

The structures of Fic family members have a conserved core topology that is decorated by various additional secondary structure elements and domains (Figure 11) (22). Structure superpositions of Fic proteins with the VopS Fic domain (DaliLite) show a similar tertiary structure (Z-scores ranging from 3.0–9.0) (22). A circular permutation is

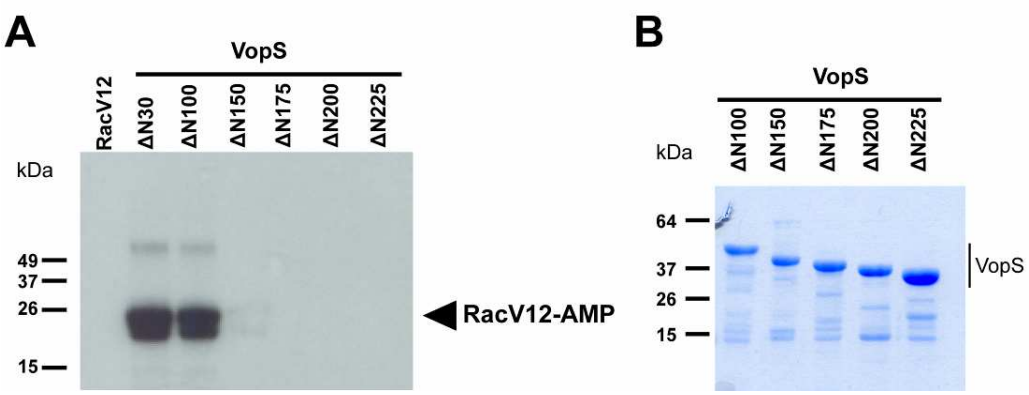

Figure 10. ***In vitro*** **AMPylation assays with VopS truncation constructs. (A)** VopS truncated protein with 30, 100, 150, 175, 200, or 225 residues removed from the N-terminus were purified and assayed for AMPylation activity. VopS is still active with 100 N-terminal residues removed but loses activity when 150 residues or more are removed. **(B)** Coomassie stain of VopS truncated constructs.

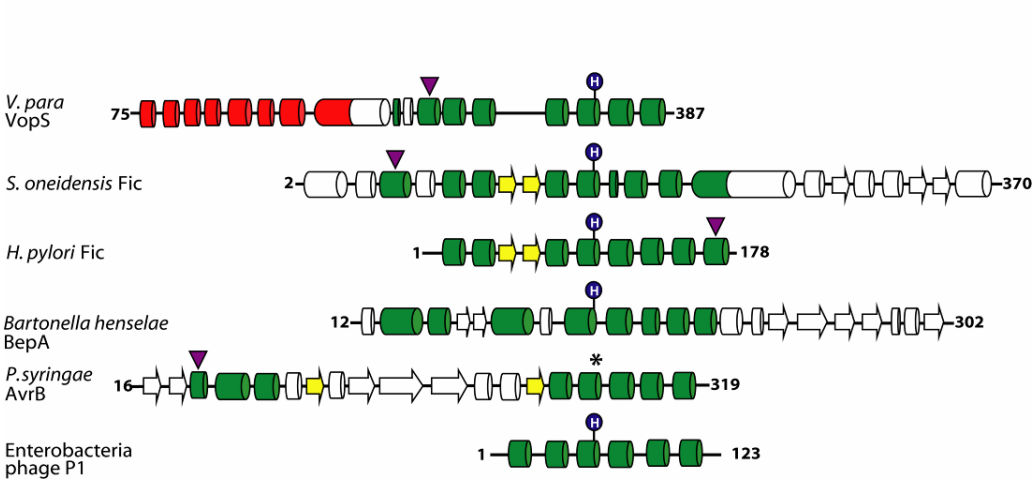

Figure 11. Schematic diagram of the secondary structure organization of known structures of Fic domain containing proteins. Shown are the VopS protein from *V. parahaemolyticus*, Fic protein from *S. oneidensis* (PDB ID 3EQX), *H. pylori* (PDB ID 2F6S), *B. henselae* (PDB ID 2JK8), and of Fido protein family member, AvrB, from *Pseudomonas syringae* (PDB ID 2NUN), and Doc domain family *Enterobacteria* phage P1 (PDB ID 3DD7). Highlighted in red and white are the structurally non-conserved N-terminus of VopS and the non-conserved secondary structures of Fido/Fic domain containing proteins, respectively. Shaded in green are the helices that form the conserved Fic domain. Purple arrowheads point to positions of the circularly permuted a' helix of Fic domains. The conserved histidine is shown for all proteins except AvrB, which lacks the conserved histidine and is thus highlighted with an asterisk.

seen in the a′ helix of Fic domains (Figure 12). The a′ helix in all Fic domains occupies the same location in the tertiary structure but not in the primary sequence; it may be located at either the N or C terminus of the protein. VopS has an N-terminal a′ helix and a second permutation corresponding to the Fic helix a7.

Crystal structures of Fic domains, such as *H. pylori* Fic protein, include a β-hairpin between helix a2 and a3 that is positioned near the Fic motif loop and has been proposed to mediate peptide substrate binding. In the VopS model, the corresponding loop does not form the hydrogen bonding pattern of a β-hairpin and is, therefore, referred to as the hairpin loop (Figure 12). The lack of a well defined β-hairpin may possibly result from the absence of an interaction with a substrate such as ATP or a Rho GTPase.

Structural Mechanism of Substrate Interaction

Superposition of VopS with BepA structures and a representative Fic domain from *H. pylori* highlights the structure differences in the hairpin loop and illustrates movement within this structural element (Figure 13). The superposition reveals four distinct positions of the hairpin element. The VopS hairpin loop adopts a position that is intermediate to the others. The BepA hairpin loop appears to be in an open conformation, whereas the *H. pylori* Fic hairpin is in a relatively closed position that is lowered toward the active site. The BepA apo- and MgPPi-bound structures have two different hairpin loop conformations, supporting the notion that ATP binding induces structural changes in the hairpin element of the enzyme.

Evidence for the β-hairpin mediating protein-substrate interaction is seen in the crystal structure of *Shewanella oneidensis* Fic protein [PDB ID: 3EQX] (Figure 13B). The N terminus of a crystallographic symmetry-related Fic domain binds edge-on to the β-hairpin to form a three-stranded β-sheet. AvrB (avirulence protein B) [PDB ID: 2NUD] from *Pseudomonas syringae* possesses a structurally similar Fic topology that includes a

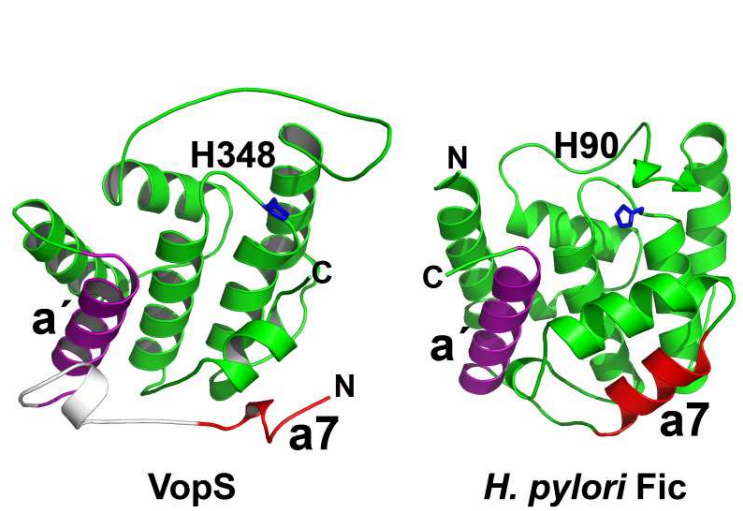

Figure 12. Structural comparison of the permuted helix of VopS Fic domain. Colored in purple is the circularly permuted a' helix of the Fic protein family. Colored in red is the second permuted helix a7. Displayed in blue is the conserved His348 for VopS and His90 for *H. pylori* Fic protein positioned near the C-terminus of the a4 helix.

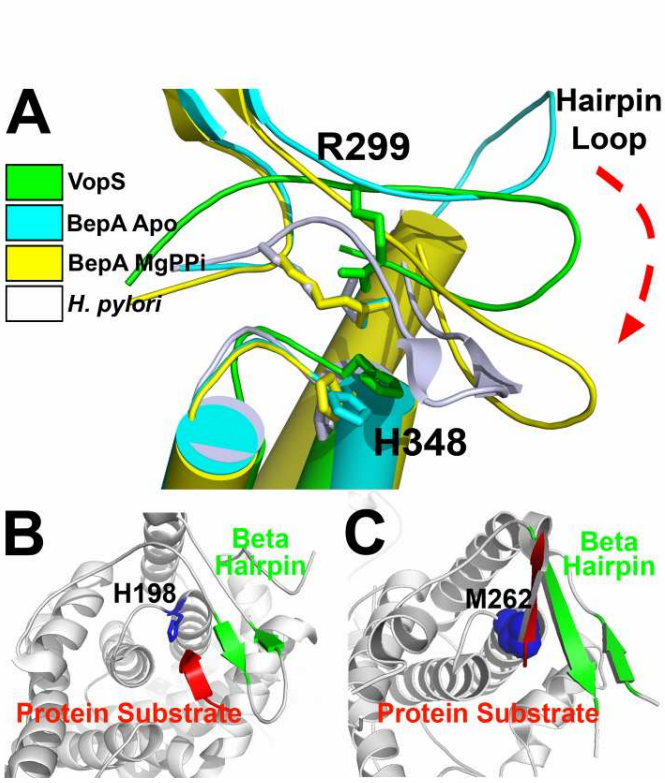

Figure 13. Structural mechanism for protein substrate binding. (A) Structural comparison of beta-hairpins and hairpin loops of Fic domains from VopS (green), BepA Apo (cyan, PDB ID 2JK8), BepA MgPPi complex (yellow, PDB ID 2JK8), and *H. pylori* (white, PDB ID 2F6S) reveal four distinct structural conformations. The BepA structures used for superposition had two molecules per asymmetric unit where one had MgPPi bound and the other BepA molecule was unbound. The hairpin loop of VopS is in an intermediate conformation. VopS residues Arg299 and His348 have altered orientations relative to other Fic structures. **(B)** The beta-hairpin (green) of *Shewanella oneidensis* Fic (PDB ID 3EQX) forms a three-stranded beta sheet interaction with the N-terminus (red) of a neighboring crystal mate. The conserved histidine of the Fic motif is depicted in blue. **(C)** AvrB (PDB ID 2NUD) forms a structurally analogous anti-parallel beta-sheet (green) formed via edge-on strand interactions with a high-affinity peptide. AvrB lacks the Fic sequence motif; highlighted in blue spheres is the Met262 residue that takes the place of the conserved histidine of the Fic motif.

similar β interaction with a bound peptide (Figure 13C). In support of this binding mechanism, a complex structure of Cdc42 with its downstream substrate, p21-activated kinase, has revealed β-strand interactions with the Cdc42 switch 1 region. Analogous to the interaction with p21-activated kinase, the Rho GTPase switch 1 region could interact with an induced Fic domain β-hairpin in VopS. Such an interaction would require a similar movement of the VopS hairpin loop illustrated in Figure 12.

VopS Active Site Fic Motif Residues Contribute to Catalysis

Fic domain family members are characterized by a consensus sequence motif HPFX(D/E)GN(G/K)R. The evolutionarily conserved histidine from the motif is seen in the central core of the VopS Fic domain, located in the loop between helix a4 and a5 (Figure 9B).

Structural superpositions of VopS with *H. pylori* and with BepA Fic domains show a conserved positioning of this loop and the corresponding side chains that define the motif (Figure 14). The VopS Phe350 side chain stacks on His287 of helix a2, whereas its main-chain carbonyl oxygen and amide nitrogen form hydrogen bonds with the main chain of another Fic motif residue, Asn354 (Figure 9B).

The structure of BepA complexed with magnesium pyrophosphate has magnesium coordinated by the acidic glutamate residue of the Fic motif (Figure 14A). This observation supports a role of the corresponding VopS residue Asp352 in magnesium coordination, as acidic residues commonly play a role in binding positively charged divalent metal ions. The BepA complex structure reveals that the conserved asparagine of the Fic motif interacts with pyrophosphate. This observation suggests that Asn354 of VopS has dual functions; that is, interaction with the β-phosphate of ATP and positioning the active site loop through main-chain interactions. One notable difference between the VopS Fic motif and those of other Fic proteins is the relative positioning of

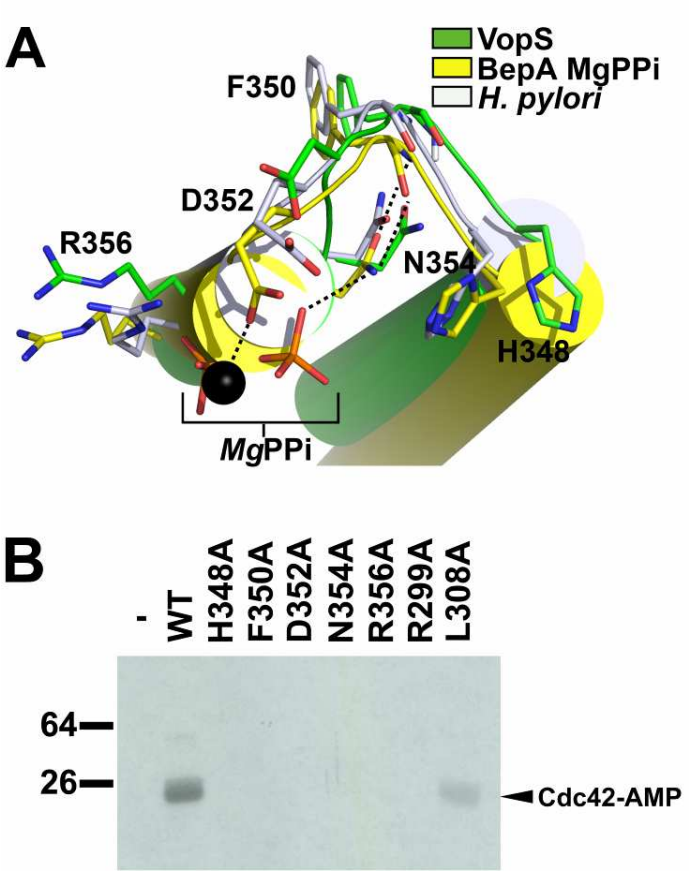

Figure 14. Residues important for function in VopS Fic domain. (A) Superposition of conserved residues of the Fic motif. Colored in green are VopS Fic secondary structural elements and carbon atoms. Superimposed onto VopS are Fic motif residues from *H. pylori* (white secondary structure and carbons, PDB ID 2F6S) and BepA from *B. henselae* (yellow secondary structure and carbons, PDB ID 2JK8). Pyrophosphate and magnesium (black sphere) are complexed with BepA. Black dashed lines represent hydrogen bonds in the MgPPi BepA complex. **(B)** *In vitro* AMPylation assay incubated with 100 μM His-Cdc42-(1-179)-(Q61L) and ^{32}P-α-ATP at 25ºC only or with 5 nM of wild-type GST-VopS-(31-387) or mutants. The assay was stopped at 90 seconds with loading buffer.

the imidizolium group of the His348 ring. In VopS His348 is rotated by ~90° relative to the position observed in other structures (Figure 14A).

Arg356 appears to have a role in protein stability because of its electrostatic interactions with Asp259 from the neighboring helix a1 (Figure 9B). Arg356 also lines a shallow pocket that includes the other active site residues, such as Asp352 and Asn354, and thus may be involved in ATP binding. Accordingly, the corresponding Fic motif arginine residue in BepA is positioned near a phosphate moiety of the pyrophosphate. Distal in sequence but conserved among Fic domains is Arg299, which points downward into the active site and is positioned above His348. Within the same hairpin loop, Leu-308 is the nearest residue to His348 at 3.5 Å (Figure 9B).

The VopS Fic motif residues appear to form a conserved active site proximal to the hairpin loop. Conserved active site residues were individually mutated to alanine, purified, and assayed for AMPylation activity using Cdc42 as protein substrate and [α-^{32}P]ATP. VopS mutant analysis shows that the Fic motif residues (His348, Phe350, Asp352, Asn354, and Arg356) and the distal but evolutionarily conserved residue, Arg299, are critical for full catalytic activity (Figure 14B). Mutation of the hairpin element residue proximal to the active site (Leu308) resulted in a slight reduction of activity, supporting a role for the hairpin loop in protein substrate binding (Figure 14B).

Crystallization Trials for VopS Complex with Substrates

Next I tried to crystallize VopS in complex with protein substrates and small molecule substrates. One of the first attempts was to co-crystallize VopS-(75-387) with Cdc42 (1-179)-(Q61L) along with 1 mM ATP, 1 mM GMPPNP, 5 mM $MgCl_2$ but there were no positive hits in the crystallization screens (Figure 15). Table 4 lists the crystallization trials attempted for VopS.

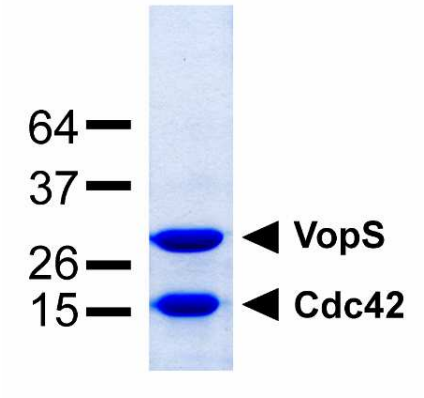

Figure 15. Coomassie stain of VopS and Cdc42 complex. VopS-(75-387) complexed with His-Cdc42-(1-179)-(Q61L) for crystallization trials.

Table 4. VopS Crystallization Trials

Date	Protein	Concentration (mg/ml)	Additives	Temp (°C)
7/2/2008	VopS-(31-387)	10	None	20
10/30/2008	VopS-(112-387)	10	None	20
11/3/2008	VopS-(112-387)	10	None	4
11/18/2008	VopS-(75-387)	12	None	20
12/19/2008	VopS-(75-387)-(H348A) + Cdc42 (1-179)-(Q61L)	12	1mM ATP, 1mM GMPPNP, 5mM MgCl2	20
12/22/2008	VopS-(75-387)-(H348A) + His-Cdc42 (1-179)-(Q61L)	20	1mM ATP, 1mM GMPPNP, 5mM MgCl2	20
1/26/2009	VopS-(75-387)-(H348A) + His-Cdc42 (1-179)-(Q61L)	20	1mM ATP, 5mM MgCl2	20
1/26/2009	VopS-(75-387)-(H348A)	12	1mM ATP, 5mM MgCl2	20
2/3/2009	VopS-(75-387) + His-Cdc42 (1-179)-(Q61L)	16	1mM AMPCPP, 1mM GMPPNP, 5mM MgCl2	20
2/3/2009	VopS-(75-387) + His-Cdc42 (1-179)-(Q61L) *	11	1mM AMPCPP, 1mM GMPPNP, 5mM MgCl2	20
4/27/2009	VopS-(75-387) + His-Cdc42 (1-179)-(Q61L)	8	0.5mM AMPCPP, 0.5mM GMPPnP, 2.5mM MgCl2	20
4/27/2009	VopS-(75-387) + His-Cdc42 (1-179)-(Q61L)	8	0.5mM AMPCPP, 0.5mM GMPPnP, 2.5mM MgCl2	4
5/8/2009	VopS-(79-387) + His-Cdc42 (1-179)-(Q61L)	6	0.5mM AMPCPP, 0.5mM GMPPnP, 2.5mM MgCl2	20
5/8/2009	VopS-(79-387) + His-Cdc42 (1-179)-(Q61L)	6	0.5mM AMPCPP, 0.5mM GMPPnP, 2.5mM MgCl2	4
5/27/2009	VopS-(31-387)-(H348A) + His-Cdc42 (1-179)-(Q61L)	7.5	1mM AMPCPP, 1mM GMPPnP, 5mM MgCl2	4
5/27/2009	VopS-(75-387)-(H348A) + His-Cdc42 (1-179)-(Q61L)	6	1mM AMPCPP, 1mM GMPPnP, 5mM MgCl2	4
9/11/2009	VopS-(79-387)	15	None	20
9/11/2009	VopS-(79-387)	15	0.5mM AMPCPP, 2.5mM MgCl2	20
9/11/2009	VopS-(79-387)-(H348A)	13	None	20
9/11/2009	VopS-(79-387)-(H348A)	13	0.5mM AMPCPP, 2.5mM MgCl2	20
9/11/2009	GST-His-Fic2-(3359-3781) + His-Rac1-(1-192)	11	None	4
9/26/2009	VopS-(79-387)	15	0.5mM PPi, 0.5mM MgCl2	20
9/26/2009	VopS-(79-387)-(H348A)	12	0.5mM PPi, 0.5mM MgCl2	20
10/5/2009	VopS-(78-387)	15	5mM AMPCPP, 5mM MgCl2	20
10/5/2009	Fic2-(3359-3781)	12	1mM PPi, 1mM MgCl2	20
10/23/2009	Fic2-(3359-3781)	12	1mM AMPCPP, 1mM MgCl2	20
10/23/2009	VopS-(79-387) + Cdc42-(1-179)-(Q61L)	6	1mM GTPgammaS, 1mM MgCl2	20
10/30/2009	VopS-(75-387) + Cdc42-(T35A)-(1-179)-(Q61L)	11	1mM GTPgammaS, 1mM MgCl2	20
11/18/2009	VopS-(75-387) + Cdc42-(T35A)-(1-179)-(Q61L)	14.7	1mM GDP, 1mM MgCl2	20
11/18/2009	VopS-(75-387) + (Thr35 AMPylated)-Cdc42-(1-179)-(Q61L)	12.1	1mM MgCl2	20
11/18/2009	(Thr35 AMPylated)-Cdc42-(1-179)-(Q61L)	8	None	20
11/25/2009	VopS-(79-387)	10	4mM AMP, 5mM MgCl2	20
11/25/2009	(Thr35 AMPylated)-Cdc42-(1-179)-(Q61L)	11	5mM GMPCPP, 5mM MgCl2	20
1/8/2010	VopS-(75-387)-(R299A)	8	2mM ATP, 2mM MgCl2	20
1/8/2010	VopS-(75-387)	13	2mM Switch1 Rac1 Peptide EYIPTVFD	20
3/8/2010	VopS-(75-387) + His-Cdc42-(1-179)-(Q61L)	13.4	2mM PPi, 2mM MgCl2	20
3/18/2010	VopS-(75-387) + His-Cdc42-(1-179)-(Q61L)	63	3mM PPi, 3mM MgCl2	20
4/6/2010	VopS-(75-387) + GMPPCP loaded Rac1 (1-176)	25	1mM PP, 1mM AMP	20

Adenosine monophosphate (AMP), pyrophosphate (PPi), α,β-Methyleneadenosine 5′-triphosphate (AMPCPP) were tested for their ability to inhibit VopS-mediated AMPylation to help determine proper concentration for crystallization trials to cocrystallize with VopS (Figure 16). AMPCPP is a non-hydrolyzable analogue of ATP. Assays were performed in duplicates under initial velocity conditions. Increasing concentrations of AMP (0 µM, 100 µM, 200 µM, 500 µM, 1 mM) were tested at constant concentrations of ATP (160 µM) and VopS-(31-387) (5nM). Inhibition studies with AMPCPP (0 µM, 100 µM, 200 µM, 500 µM, 1 mM) were carried out under the same conditions. AMP and AMPCPP appear to be poor inhibitors even at high concentrations (Figure 16). Increasing concentrations of PPi (0 µM, 150 µM, 400 µM, 900 µM) were then tested at constant concentration of VopS-(31-387) (5nM) and ATP (1mM). Different His-Cdc42-(1-179)-(Q61L) concentrations were used (60 µM, 100 µM, 150 µM, 200 µM, 500 µM). PPi is a more potent competitive inhibitor compared to AMPCPP and AMP. ATP can adopt multiple conformations that allow it to adapt and bind to different enzymes. The “C” (carbon) in AMPCPP may restrict the confirmations the analogue can form, hindering efficient binding to enzymes. AMP was unable to efficiently inhibit VopS possibly due to the lack of beta and gamma phosphate groups that are coordinated by magnesium, mediating ATP binding to VopS (Figure 14). This is consistent with the fact that PPi, an analog of beta and gamma phosphate of ATP, inhibits VopS enzyme activity at a greater potency. However, crystallization trials with PPi and VopS were not successful.

Discussion

Summary of VopS Structure

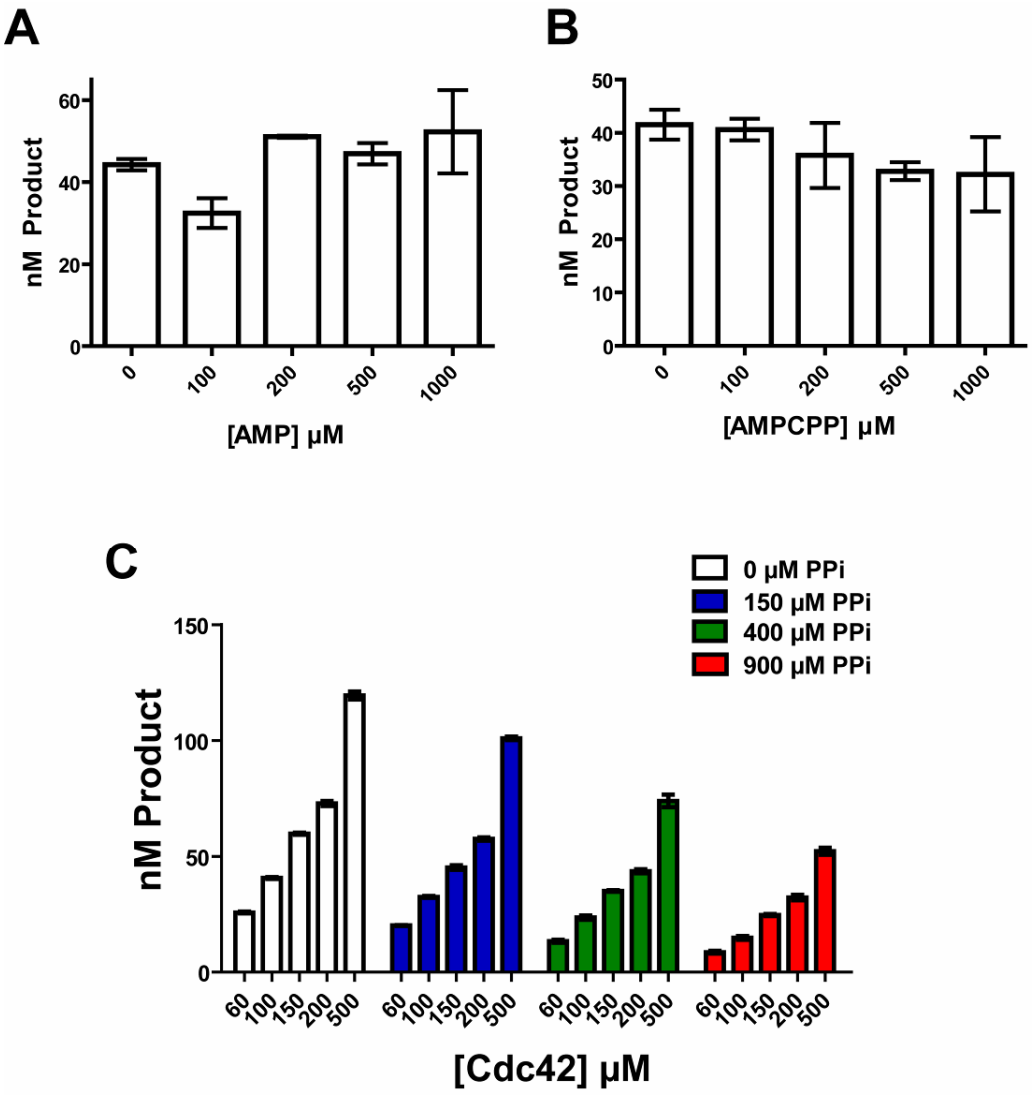

Figure 16.. **Small molecule inhibition studies. (A)** *In vitro* filter binding AMPylation assays under steady-state conditions with 5nM VopS, 160 µM ATP and 300 µM Cdc42 were incubated with different concentrations of AMP. Assays were stopped at 75 seconds. **(B)** *In vitro* filter binding AMPylation assays under steady-state conditions with 5nM VopS,160 µM ATP and 300 µM Cdc42 were incubated with different concentrations of AMPCPP. Assays were stopped at 75 seconds. **(C)** *In vitro* filter binding AMPylation assays under steady-state conditions with 5nM VopS, 1mM ATP, at various Cdc42 (60 µM, 100 µM, 150 µM, 200 µM, 500 µM), and PPi (0 µM, 150 µM, 400 µM, 900 µM) concentrations.

The overall structure of VopS and comparisons of known Fic structures provided structural understanding of the mechanism of AMPylation. (1) The study provided the first structure of a biochemically characterized threonine AMPylator. (2) The functions of conserved Fic motif residues were characterized in *in vitro* AMPylation assays. (3) The residues potentially responsible for binding to magnesium and ATP were identified through comparisons with other Fic structures. (4) Lastly, based on the structure, we proposed the structural basis for enzyme-substrate interaction mediated by the conserved hairpin loop in the VopS Fic domain.

Structural Comparisons of VopS with the Structure of IbpA-Cdc42-AMP Complex

The crystal structure of a Fic domain from IbpA *H. somni* called IbpAFic2 was solved in native form at 1.8 Å [PDB ID 3N3U] and as a complexed with Cdc42 at 2.3 Å resolution [PDB ID 3N3V] (45). The IbpAFic2 structure in the native and complex forms is mostly alpha helical with an N-terminal arm region and a C-terminal Fic domain similar to VopS (Figure 17A). For the complex structure, a catalytically inactive form of IbpAFic2 [IbpAFic2-(H3713A)] and AMP modified form of Cdc42 (on tyrosine 32) were used. The structure of IbpAFic2-(H3717A)-Cdc42-(AMPylated) is an end-product complex that mimics the GDI-bound state of the Rho GTPase. The beta/hairpin loop of IbpAFic2 is disordered in the native structure, but in the complex, the hairpin loop transitions to an ordered state upon protein substrate binding. Additional details on the catalytic mechanism used by IbpA and VopS to AMPylate Rho GTPases are discussed in Chapter 4, wherein steady-state kinetic studies on VopS-catalyzed AMPylation are described.

Sequences of IbpAFic2 and VopS fragments used for crystallization were aligned to determine sequence identity and similarity. IbpAFic2 and VopS share less than 16 % sequence identity and about 26 % similarity. Superposition of VopS to IbpAFic2 native

and complex structures using the DaliLite pairwise structure alignment program show an RMSD of 3.4 Å - 3.6 Å. IbpAFic2 native structure has a disordered hairpin loop but becomes ordered in complex with Cdc42 (Figure 17A). AMP was modeled into the VopS and IbpAFic2 native structures using the DaliLite structure superposition. Surface representations of VopS (green) and IbpAFic2 native structures (red) display relatively exposed AMP binding pockets (Figure 17B). AMP is buried in the IbpAFic2 complex structure (blue).

Analysis of the AMP binding pocket of IbpAFic2 reveals extensive hydrophobic and hydrogen bond interactions with AMP. Residues involved in AMP binding either through main chain or side chain interactions include Lys3670, Glu3671, Asn3672, Ala3673, Phe3675, Ile3714, Gly3722, Asn3723, Gly3724, Arg3728, Pro3752, Ile3755, and Gln3757. IbpAFic2 native structure and VopS were superimposed onto the IbpAFic2 complex structure, and AMP-coordinating residues were compared (Figure 17C). Residues in the conserved Fic motif of IbpAFic2, Gly3722, Asn3723, and Gly3724 are important for AMP binding. The corresponding residues in VopS are Gly353, Asn354 and Gly 355. Seven residues (Lys3670, Glu3671, Asn3672, Ala3673, Phe3675, Pro 3752 and Gln3757) that coordinate AMP binding in IbpAFic2 are absent in VopS and are designated with asterisks (Figure 17C). The lack of these AMP-coordinating residues in VopS is probably due to the open hairpin loop confirmation that orients away from the AMP binding pocket. The hairpin loop acts as a lid that buries AMP in IbpAFic2 (Figure 17B). Point mutation of hydrophobic residues A3673 and Phe3675 disrupted IbpAFic2's AMPylation activity [32]. Point mutation of hydrophobic residue Ile3755 that is involved in AMP-binding also disrupted AMPylation activity in IbpAFic2 [32].

The highly conserved Arg3728 of IbpAFic2 (Arg359 for VopS), outside but close to the Fic motif, coordinates the ribose ring of AMP. This residue is present among many Fic domains (22). Interestingly, this arginine is not present in the Fic protein of

AnkX from *Legionella pneumophila* (discussed in Chapter 7). Mutation of the ribose coordinating arginine disrupts the AMPylation activity of IbpAFic2. Overall the IbpAFic2 structures, native and in complex with Cdc42-AMP, revealed two states of Fic domains, providing great insights into how AMPylation enzymes bind to ATP and protein substrates.

The switch 1 (red) and switch 2 (purple) regions of Cdc42 have important roles in enzyme-substrate interaction (Figure 18A & 18D). Switch 1 region mediates strand-to-strand interaction with the hairpin loop of the Fic domain (blue) of IbpAFic2 as previously proposed for VopS (Figure 13 & 18A & 18B). VopS (green) was superimposed onto IbpAFic2 complex structure using DaliLite. The hairpin loop of VopS is in a relatively open conformation clashing (dashed circle) with the switch 2 region of Cdc42. The hairpin loop of VopS would perhaps assume a closed conformation upon protein substrate binding.

The switch 2 region of Cdc42 interacts with the N-terminal arm region (white) of IbpAFic2 (Figure 18C & 18D). Analysis of the N-terminal arm region of IbpAFic2 and VopS suggests a similar protein substrate binding mechanism between the enzymes and the switch 2 region of Cdc42 (Figure 18C). Residues in the switch 2 region that interact with IbpAFic2 are Asp63, Tyr64, Arg66, Leu67, Leu70, and Pro73. IbpAFic2 and VopS could not AMPylate Rho family GTPases when the hydrophobic residues in the switch 2 region of Cdc42, such as Tyr74, Leu67 and Leu70, and analogous residues for Rac1 and RhoA, were mutated. Mutations of opposing residues that mediate substrate interaction in the N-terminal arm region of IbpAFic2 also disrupted its AMPylation activity (45). Consistent with a role for the VopS N-terminus in substrate binding, N-terminal truncation mutations of VopS are defective in AMPylation assays *in vitro* (Figure 10).

The structure of IbpAFic2 complexed with AMPylated Cdc42 suggests that future complex crystallization attempts could use an AMP modified GTPase to complex

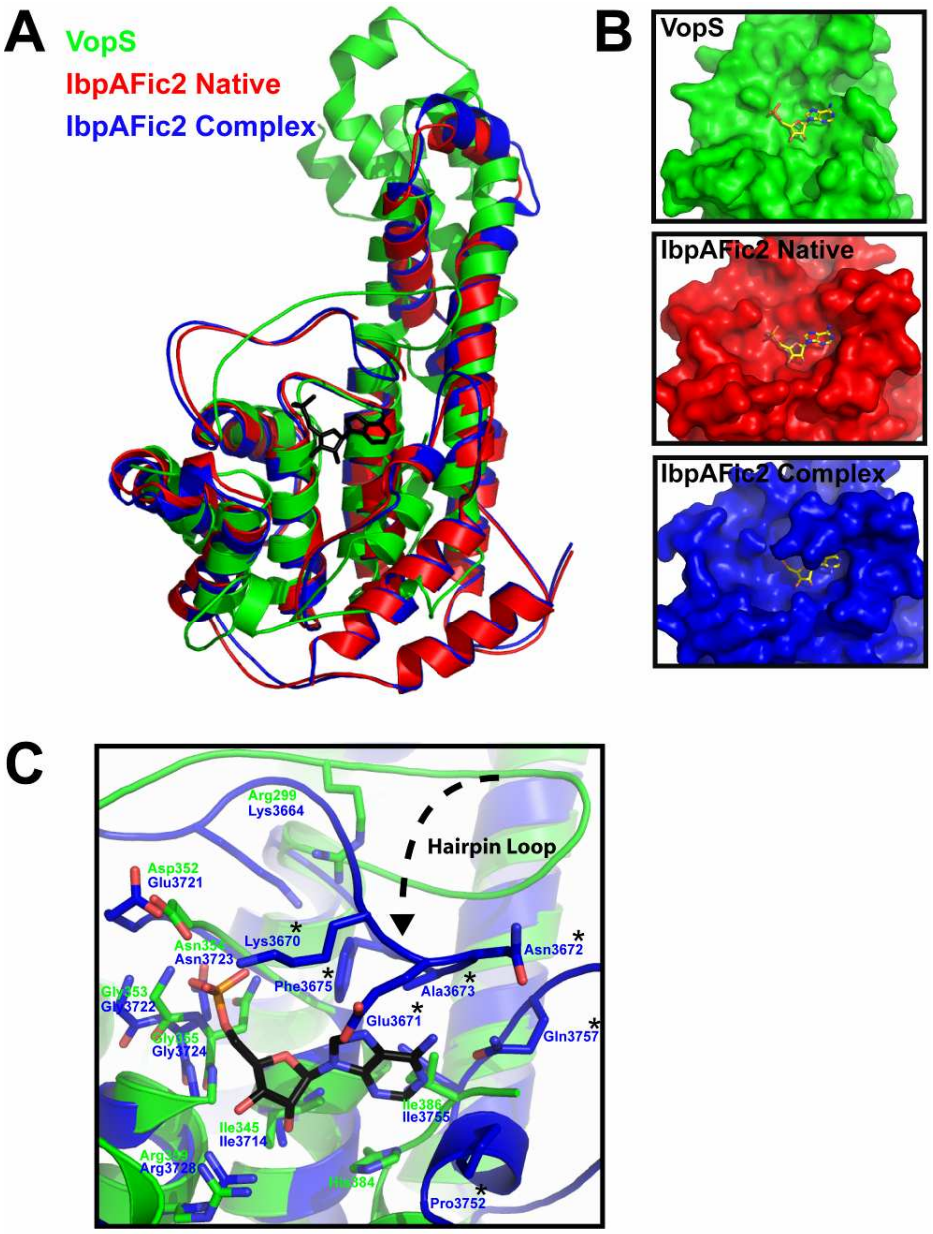

Figure 17. Analysis of AMP nucleotide binding with IbpAFic2 and VopS structures. **(A)** Structure superposition of VopS (green), IbpAFic2 native (red), and IbpAFic2 complex (blue). AMP (black) is shown. **(B)** Surface representation of VopS (green), IbpAFic2 native (red), and IbpAFic2 complex (blue). After structure superposition with DaliLite the AMP nucleotide was modeled in for VopS and IbpAFic2 native structure. The nucleotide AMP is more buried in the IbpAFic2 complex structure when compared to VopS and IbpAFic2 native protein. **(C)** Detailed look at residues important for AMP binding. VopS (green) was superimposed onto the IbpAFic2 complex (blue) structure. Asterisk represents residues in the IbpAFic2 complex important for binding AMP that are not present in the VopS structure. The hairpin loop of VopS is in an open position compared to the IbpAFic2 complex.

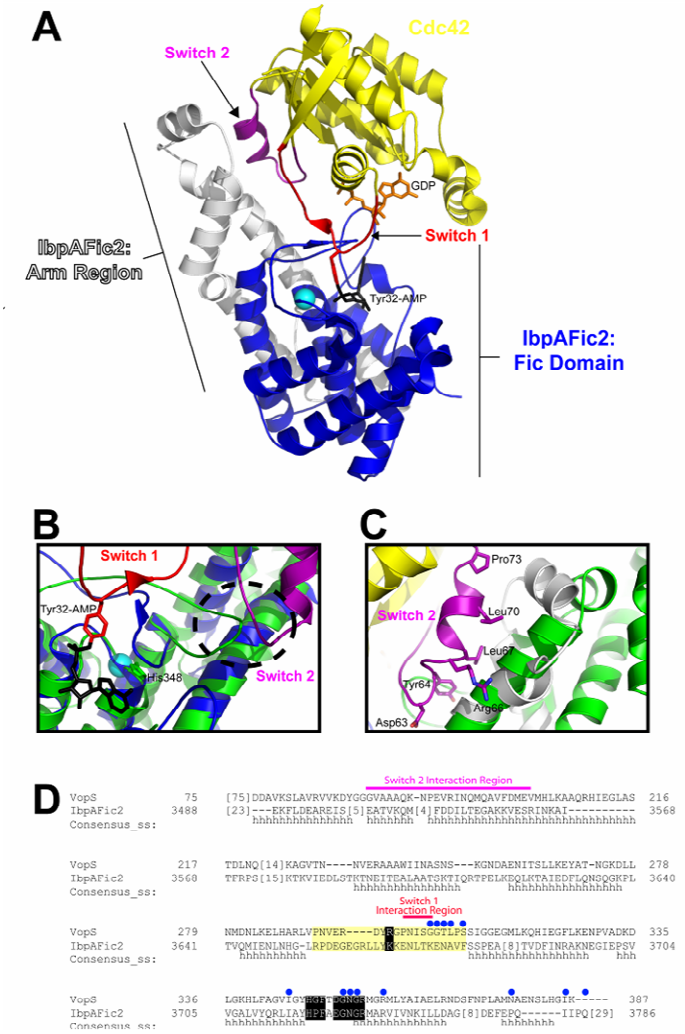

Figure 18. The protein substrate binding surface of IbpAFic2-(H3717A)-Cdc42 AMP complex with comparisons to VopS. (A) Structure of IbpAFic2 in complex with Cdc42 (yellow). Cdc42 is modified with AMP on tyrosine 32. Cdc42 switch 1 (red) region interacts strand-to-strand fashion with the hairpin loop of the conserved Fic domain (blue) of IbpA. The nonconserved arm region (white) of IbpAFic2 interacts with the switch 2 (purple) region of Cdc42. GDP is colored orange and AMP is black. Residue H3717A is denoted with a sphere (cyan). **(B)** The structure of VopS (green) and IbpAFic2 (blue) from the complex structure were superposed with DaliLite. Dashed circle highlights clashing of VopS hairpin loop with the switch 2 (purple) of Cdc42. Histidine 348 of VopS is revealed. **(C)** A closer look at the arm region of IbpAFic2 (white) superimposed with the arm region of VopS (green). Cdc42 switch 2 (purple) residues that interact with IbpAFic2 arm region (white) are shown. **(D)** Sequence alignment of VopS and IbpAFic2 that was based on their structural superposition to each other. VopS residues 75-160 did not align structurally to IbpAFic2 residues 3488-3520 and so are not shown in sequence alignment. Switch 2 interaction region (purple) and switch 1 interaction region (red) observed in IbpAFic2 complex structure are shown. Blue circles designate residues in IbpAFic2 with roles in AMP binding. Highlighted in yellow is the hairpin loop of Fic domains and shown in black are conserved residues.

with a catalytically inactive VopS mutant. The GTPase should be coexpressed with active VopS to increase the labeling efficiency and homogeneity of AMPylated GTPase for structure studies. Other crystallization strategies may want to employ catalytically inactive mutants that do not increase the apparent affinity of the substrate. One example would be to use the VopS R299A mutant that is catalytically disrupted but the K_m for ATP does not change (will be discussed in Chapter 4).

DrrA, an ATase Domain Protein and AMPylator of Rab1 GTPase

After the publication of the VopS structure, an AMPylator containing an ATase domain, a secreted effector protein DrrA from *Legionella pneumophila* (*L. pneumophila)*, was discovered. The N-terminal region of *L. pneumophila* DrrA (residues 9-218) is structurally similar to GS-ATase (46). *In vitro*, DrrA was able to AMPylate tyrosine-77 in the switch 2 region of Rab1b, a GTPase involved in membrane trafficking. Rab1b AMPylation by DrrA disrupted vesicular trafficking in cells. Both DrrA and GS-ATase contain the conserved catalytic motif G-X_{11}-D-X-D, with each aspartate coordinating a magnesium ion. Mutations of the conserved aspartates in DrrA (D110 and D112) to alanines abrogated its AMPylation activity (46). Concurrently, a structure of tyrosine-77-AMPylated Rab1b in GppNHp-bound form was also solved (46). So far, structures of the DrrA-Rab1b complex have not been solved. Structural comparisons between modified Rab1b with a closely related GTPase, Rab3A, suggests that AMP modification does not induce structural rearrangements, but instead inhibits downstream signaling by sterically hindering the binding of substrates (46).

CHAPTER 4

STEADY-STATE KINETIC STUDIES ON VOPS MEDIATED PROTEIN AMPYLATION

Introduction

Fic domains are widely distributed across various organisms, from archaea, bacteria, to human, suggesting that protein AMPylation may be an important mechanism for regulating protein functions. Threonine AMPylation by Fic domains has not been kinetically characterized, and the kinetic contribution of conserved residues in the Fic motif HPF*X*(D/E)GN(G/K)R during catalysis is unknown. Studying the steady-state kinetics of VopS-mediated AMPylation should provide a mechanistic understanding of AMPylation and a foundation for future studies.

Results

Steady-state Substrate Measurements with WT and Mutant VopS Constructs

We determined the apparent steady-state kinetic constants for ATP using a dominant active form His-Cdc42-(1–179)-(Q61L), where glutamine 61 was mutated to leucine to disrupt intrinsic GTPase activity and lock it into a GTP-bound state. The protein substrate had 12 amino acids deleted from its C terminus to eliminate variable C-terminal proteolysis of purified substrate. The concentrations of VopS (5 nM) and His-Cdc42-(1–179)-(Q61L) (600 μM) were kept constant while varying ATP concentrations from 50–2000 μM. The apparent K_m for ATP was 160 ± 18 μM with a k_{cat} of 26 s^{-1} ± 1.0 (Figure 19A; Table 5). The catalytic efficiency for ATP was 1.60 × 10^5 s^{-1} M^{-1}. The kinetic constants for His-Cdc42-(1–179)-(Q61L) were then determined with constant enzyme (5 nM) and ATP (2000 μM)

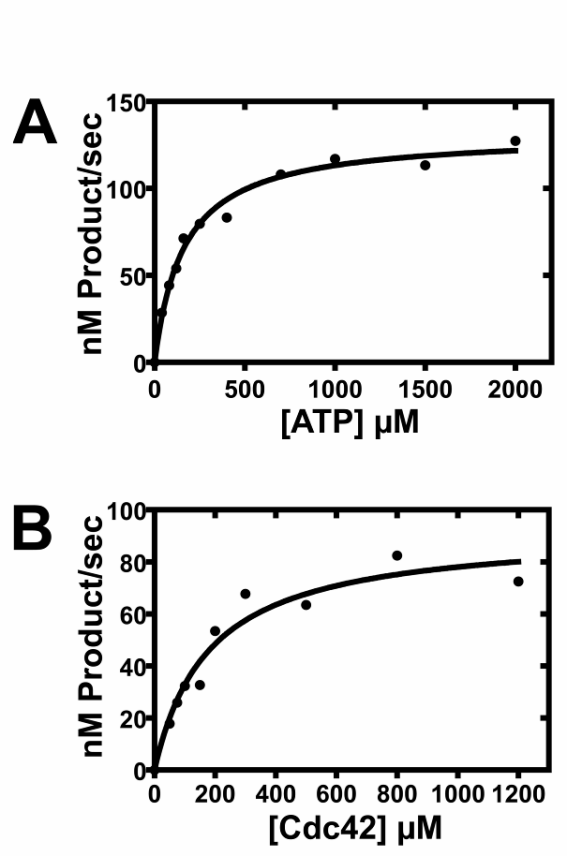

Figure 19. Apparent steady-state kinetic measurements for ATP and Cdc42. (A) Initial-velocity measurements for ATP were obtained using a constant concentration of His_6-Cdc42-(1-179)-(Q61L) at 600 µM while varying ATP concentrations (40, 80, 120, 160, 250, 400, 700, 1000, 1500, 2000 µM). **(B)** Initial velocity measurements for Cdc42 were obtained with a constant concentration of ATP at 2 mM while varying the concentration of His_6-Cdc42-(1-179)-(Q61L) (50, 75, 100, 200, 500 µM). Assays were performed in triplicate with VopS (31-387) at 5 nM. In both parts, the individual data points are depicted as circles, and the line represents the fit to these data using the Michaelis-Menten equation (Eq.1).

Table 5. Apparent Kinetic Constant Measurements for ATP

	k_{cat} [a]	K_m [a]	k_{cat}/K_m
	s^{-1}	μM	s^{-1}/M^{-1}
Wild Type	26 ± 1.0	160 ± 18	1.6×10^{5}
H348A[b]	ND[c]	ND	3.9×10^{-4}
F350A	3.0 ± 0.32	1600 ± 290	1.9×10^{3}
D352A	0.098 ± 0.0070	580 ± 100	1.7×10^{2}
N354A	0.85 ± 0.11	2600 ± 480	3.3×10^{2}
R356A	0.035 ± 0.039	1700 ± 330	2.0×10^{1}
R299A	0.36 ± 0.0019	220 ± 40	1.6×10^{2}
L308A	2.8 ± 0.28	590 ± 150	4.8×10^{3}

[a] Mean of three assays ± S.E.
[b] H348A VopS mutant was deemed severely inactive. H348A k_{cat}/K_m (s^{-1}/M^{-1}) was calculated from an endpoint measurement (performed in triplicate) at 10μM enzyme 2mM ATP, 600μMCdc42.
[c] Not determined

Table 6. Apparent Kinetic Constant Measurements for Cdc42

	k_{cat} [a]	K_m [a]	k_{cat}/K_m
	s^{-1}	μM	s^{-1}/M^{-1}
Wild Type	18 ± 1.5	180 ± 40	1.0×10^5
R299A	0.027 ± 0.0036	660 ± 170	4.1×10^1

[a] Mean of three assays ± S.E.

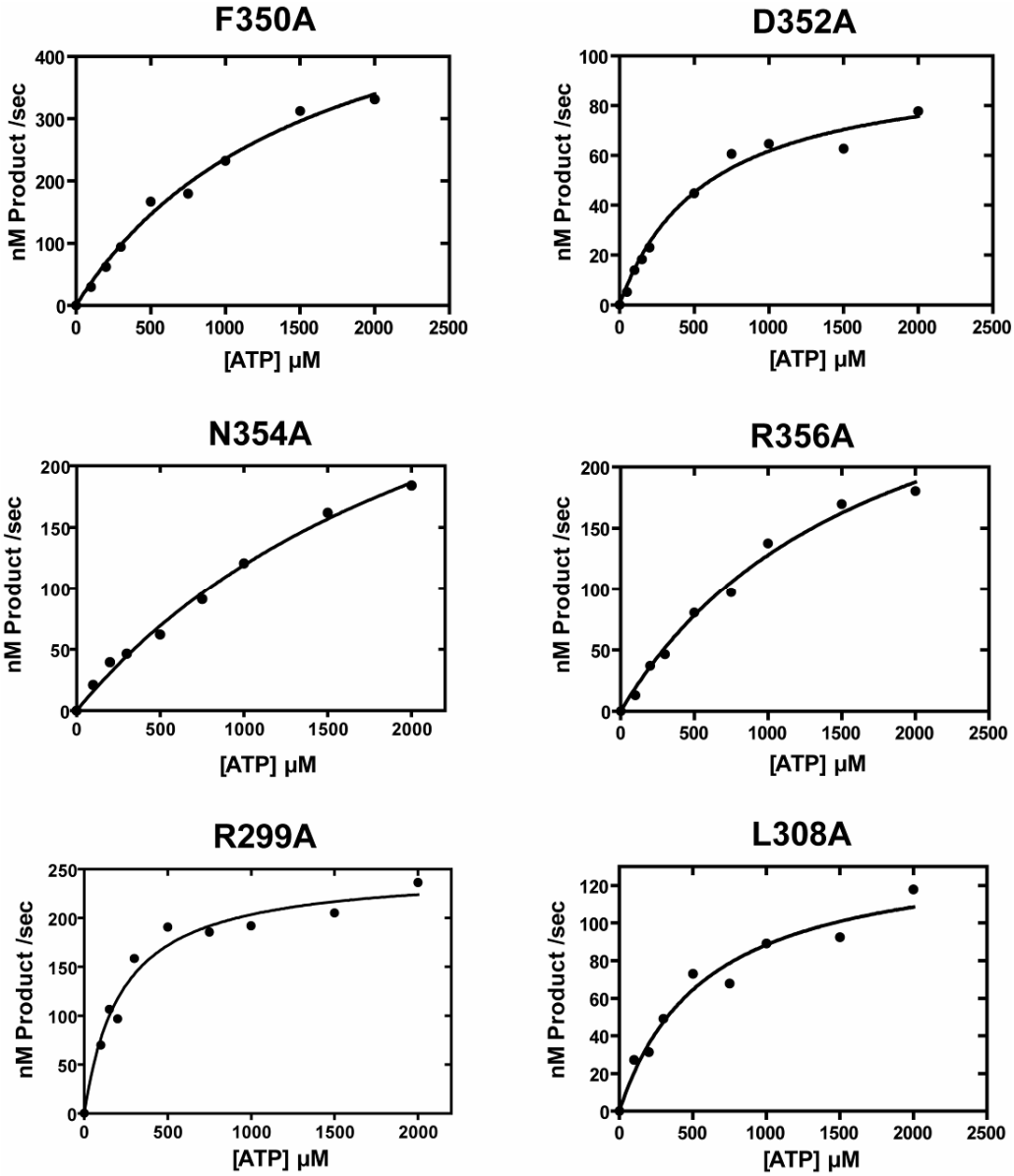

Figure 20. Apparent steady-state kinetic measurements for ATP with VopS mutant constructs. Initial-velocity measurements for ATP with VopS wild-type and mutants were obtained using constant enzyme concentration and constant concentration of His-Cdc42-(1-179)-(Q61L) at 600 µM while varying ATP concentrations from 40 µM to 2000 µM). The individual data points are depicted as circles, and the line represents the fit to these data using the Michaelis-Menten equation (Eq.1).

concentrations. Varying Cdc42 concentrations from 50 to 1200 μM were used to determine apparent values. The K_m was 180 ± 42 μM and the k_{cat} was 18 ± 1.5 s^{-1} with a catalytic efficiency of 1.0×10^5 s^{-1} m^{-1} (Figure 19B; Table 6).

The Fic motif (HPFX(D/E)GN(G/K)R) and the distal amino acid residue Arg299 were demonstrated to be important for full catalytic activity (Figure 14B). To address the question of which residue(s) of the Fic domain is critical for catalysis and to understand the function of each residue in protein AMPylation, we set out to determine the steady-state kinetic parameters for the small molecule substrate ATP for different VopS mutants. The most deleterious mutation was H348A; VopS harboring this mutation had a catalytic efficiency of less than 4.0×10^{-4} s^{-1} m^{-1}, which is 9 orders of magnitude lower than wild-type protein and 5 orders of magnitude lower than the second worst mutation R356A with a value of 2.0×10^1 s^{-1} m^{-1} (Figure 20; Table 5). Although the Michaelis-Menten constant K_m is not strictly equivalent to K_D, the value still provides a useful means to determine relative affinity when comparing the wild-type enzyme to its mutant form. The N354A mutation decreased the apparent k_{cat} and had the most drastic increase in apparent K_m for ATP: more than 2.5 mM, which is about 16-fold higher than that of the wild-type enzyme. This result further supports the hypothesis that asparagine 354 plays a role in ATP binding (Figure 20; Table 5).

VopS in which the hairpin residue Arg299 had been mutated to an alanine had a low k_{cat} value (0.036 s^{-1}), but its apparent K_m for ATP was statistically indistinguishable from the wild-type enzyme (220 μM) (Table 5). This phenomenon suggests that although catalysis is compromised, ATP binding is unaffected in the R299A mutant. To ascertain if the R299A mutant displays impaired binding of protein substrate, the K_m with respect to Cdc42 was determined. The apparent K_m was 3-fold higher than that of the wild-type enzyme (Table 6), implying that Arg299 plays a minor role in protein substrate binding.

The Arg299 residue of VopS adopts a different side-chain rotamer relative to arginines in other Fic domains (Figure 13A). Kinetic studies with the R299A mutant show an increased K_m for Rho GTPase but not for ATP, supporting a functional role for this residue in protein-substrate binding (Table 5 & 6). Comparison of VopS Arg299 and the analogous arginines of several Fic domain structures, such as the Fic protein of *H. pylori*, reveal differences in polar contacts. R299A resides in a loop just N-terminal to the VopS β-hairpin element (Figure 21). The Arg299 guanidinium group forms hydrogen bonds to a backbone carbonyl oxygen atom at the C terminus of the hairpin loop, thus anchoring the loop at both ends. VopS Arg299 anchors the hairpin loop by forming hydrogen bonds (main chain and side chain) with its C terminus. The analogous conserved arginine *H. pylori*, Arg50, interact with its β hairpin through hydrogen bonds between the guanidinium group and main-chain carbonyl oxygen atoms at its N terminus and C terminus (Figure 21). However, the *H. pylori* Arg50 guanidinium group also forms hydrogen bonds to main-chain carbonyl oxygen atoms of the active site histidine. The evolutionarily conserved Arg299 may function as an anchor that coordinates the hairpin loop toward the active site in addition to properly positioning the catalytic loop for efficient catalysis. The very low k_{cat} value for Arg299 supports a catalytic function (Figure 20, Table 5). Similar to the arginine finger observed in GTPases, Arg299 may be involved in neutralizing the negative charge of the α-phosphate in the transition state.

Kinetic Analysis of VopS Supports a Sequential Mechanism

The kinetic mechanisms of bisubstrate enzymes such as VopS can be distinguished with initial velocity studies analyzed using double-reciprocal plots. In such analyses, a set of intersecting lines supports a sequential mechanism wherein the enzyme and both substrates form a ternary complex during the reaction. In a sequential (ternary

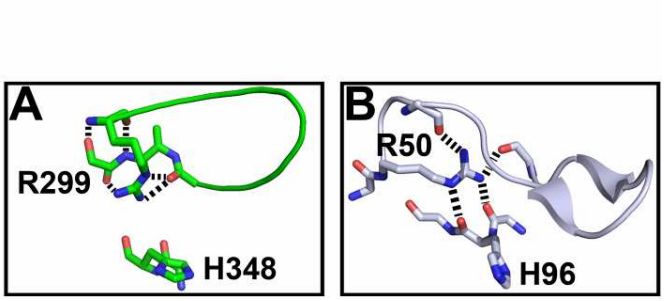

Figure 21. The molecular anchor: possible role of arginine 299. (A) VopS Arg299 interactions are mainly within the hairpin. **(B)** In *H. pylori* Fic the analogous arginine, Arg50, forms main chain-side-chain interactions that function as an anchor that links and holds the hairpin loop in an appropriate conformation, and anchors the catalytic histidine in the active site. Black dashes represent hydrogen bonds.

complex) mechanism, VopS would interact with both ATP and the GTPase to allow transfer of the AMP group to the threonine residue. In contrast, a set of parallel lines supports a ping-pong reaction mechanism in which the enzyme forms a high energy covalent intermediate that leads to subsequent modification of the second substrate. In a ping-pong mechanism, VopS could potentially form a covalent phosphoramidate intermediate where the AMP group of ATP is transferred to the conserved His348 of the Fic motif followed by AMP transfer to the GTPase. Initial velocity studies with VopS and its substrates, Cdc42 and ATP, reveal a set of intersecting lines that supports a sequential reaction mechanism. Kinetic analysis comparing a random (r^2 = 98.6) to an ordered (r^2 = 97.3) rapid-equilibrium model were statistically indistinguishable, where r^2 is the square of the correlation coefficient (Figure 22). The enzymatic analysis support a ternary complex mechanism is used for VopS-mediated AMPylation. H348A mutant is essentially an inactive enzyme, which support the model that histidine 348 functions as a general base during catalysis (Figure 22 B). The pK_a of histidine is 6.0, and so the predicted optimal activity for the VopS enzyme would be at a pH greater than 6.0 where the histidine would be more basic (Figure 23). Consistent with our kinetic results, the pH profile of VopS WT activity demonstrate the enzyme is less active at lower pH conditions (Figure 23).

Discussion

Summary of steady-state kinetic studies

Steady-state studies reveal that the conserved His348 is indispensable for AMPylation activity. The conserved residue Arg299 is also required for catalytic activity. This residue is not involved in ATP binding; instead, it probably mediates the formation of a transition state intermediate.

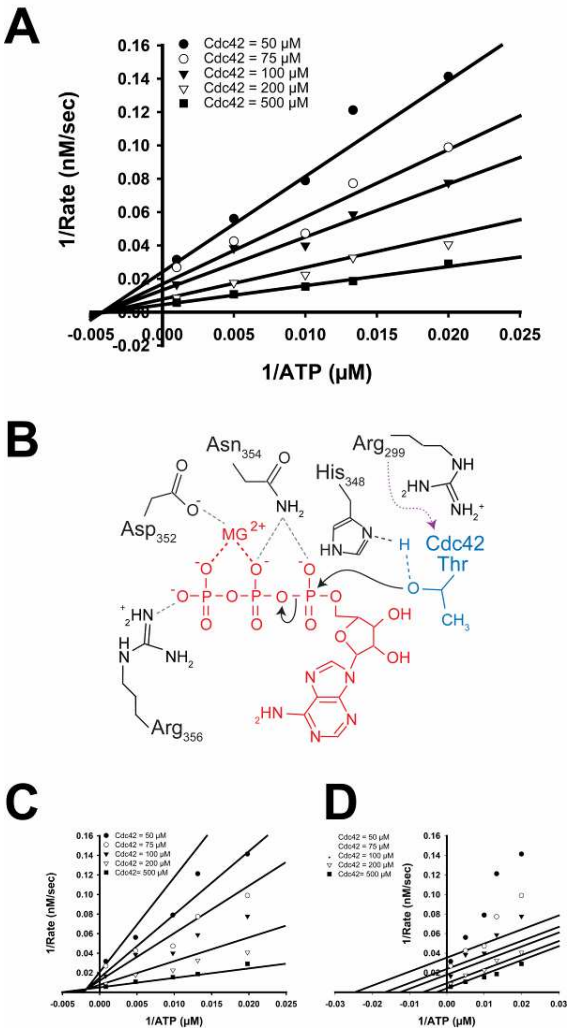

Figure 22. Ternary complex is required for catalysis. (A) Double-reciprocal plot of 1/rate vs. 1/[ATP]. Assays were performed at fixed concentrations of His-Cdc42-(1-179)-(Q61L) 50 µM (●), 75 µM (○), 100 µM (▼), 200 µM (▽), 500 µM (■) while varying ATP concentrations (50, 75, 100, 200, 1000 µM). The data were fitted to a rapid-equilibrium random model (Eq. 2) using Sigma Plot. The best-fit lines are shown. The square of the correlation coefficient (r^2) for the global fit data is 98.6. Kinetic constants from the global fit have k_{cat} = 89 s^{-1} ± 13, K_{ATP} = 280 µM ± 90, K_{Cdc42} = 490 µM ± 120. **(B)** Proposed model of catalytic mechanism. H348 functions as a general base to abstract a proton from the hydroxyl of the bound Rho GTPase threonine. The activated threonine of Cdc42 performs a nucleophilic attack on the alpha phosphate of ATP. N354 interacts with the beta phosphate of ATP. The magnesium, coordinated by D352, interacts with the beta and gamma phosphate of ATP. R356 coordinates the gamma phosphate of ATP. R299 positioned in the hairpin has a role in the transition state. Dashed lines indicate proposed hydrogen bonds. **(C)** Global fit of the data for ordered rapid-equilibrium mechanism. The square of the correlation coefficient (r^2) for the global fit data is 97.3. **(D)** Global fit of the data for the ping-pong mechanism. The square of the correlation coefficient (r^2) for the global fit data is 94.7. Global fits were performed with Sigma Plot.

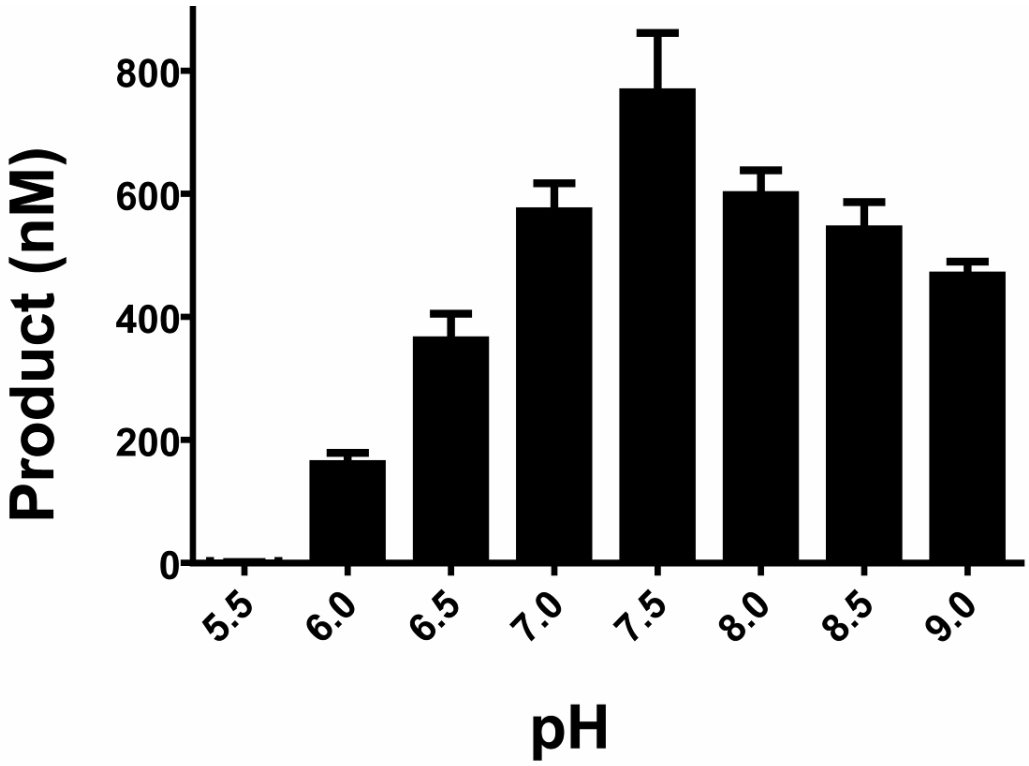

Figure 23. pH profile of VopS wild-type activity. Assays were performed in triplicate at different pH ranges. Reactions were incubated with 5nM VopS-(31-387), 100uM His-Cdc42-(1-179)-(Q61L), and 200 μM ^{32}P-α-ATP for 75 seconds and spotted onto P81 Whatman filters. Error bars represent standard error of the mean.

The kinetic studies support a sequential reaction mechanism whereby a ternary complex is formed between VopS, ATP, and Cdc42. In such a complex, the threonine residue of the Rho GTPase is positioned near the imidazole side chain of the catalytic histidine (His348) of VopS (Figure 22B).

Total intracellular concentrations of ATP have been reported at about 3 mM (47). Low millimolar ATP concentrations in normal host cells suggest that VopS would be saturated with ATP since the apparent K_m for ATP was 160 ± 18 μM. On the other hand the apparent K_m for Cdc42 was 180 ± 40 μM. Intracellular concentrations of RhoA, Rac1, and Cdc42 have been estimated to be 3 μM, 7 μM, and 2.5 μM, respectively (48,49). This implies that the activity of VopS inside host cells may be limited to the availability of protein substrate unless other signaling mechanisms are used to compensate for the high K_m. Signaling molecules exhibit multiple mechanisms to speed up catalysis in order to enhance catalytic efficiency. These mechanisms include compartmentalization, membrane localization, and protein interaction with adaptor and scaffolding proteins (49,50). Proteins that are membrane localized or brought into close proximity by adaptor or scaffolding proteins would provide a local and high protein concentration environment that can promote activity of proteins with normally poor K_m or k_{cat}. Studies performed by Melanie Yarbrough to determine VopS localization using biochemical subcellular fractionation and confocal microscopy suggest localization to the cytosol (unpublished results).

The VopS-(31-387) construct was used for kinetics assays. The removal of the first 30 amino acids may be a reason for the higher than expected K_m levels for Cdc42. The first 30 amino acids of VopS is the secretion signal that promotes the translocation of the effector through the Type III needle complex. It is not known if the first 30 amino acids of VopS is required for full activity or provide another function during the time course of cellular infection.

The order of substrate binding

The structure of IbpAFic2-(H3717A)-Cdc42 AMP supports a ternary complex mechanism that is consistent with VopS kinetic studies [32]. The IbpAFic2 complex structure suggests an order of substrate binding. The IbpAFic2-(H3717A)-Cdc42 AMP structure display in atomic detail indicate that the position of the catalytic histidine of IbpAFic2 may act as a base to abstract a proton from tyrosine 32 to promote the attack of the alpha position of ATP. The nucleotide binding pocket is open and exposed in the IbpAFic2 native structure but in the IbpAFic2 complex structure the nucleotide binding pocket is buried by the hairpin loop (Figure 17). This suggests that ATP binds first and the GTPase substrate binds second.

Comparison of AMPylation and other protein modifications

Similar to phosphorylation, VopS uses the abundant high energy metabolite, ATP, to modify its substrates. Comparing the turnover rates of VopS-mediated AMPylation with other posttranslational modifications mediated by various enzymes, i.e. phosphorylation, acetylation, and methylation, suggests that VopS is a relatively fast enzyme with the apparent k_{cat} = 18 s^{-1} (Table 1). Phosphorylation by mitogen-activated protein kinase p38 has a k_{cat} of 22.6 s^{-1} (51). Histone acetyltransferase GCN5 has a k_{cat} of 1.7 s^{-1}, and the G9a histone methyltransferase has a k_{cat} of 0.0012 s^{-1} (52,53). The AMPylation turnover values are well within the range of other established posttranslational modifications. Therefore, AMPylation is a potential regulatory mechanism that could mediate eukaryotic signaling processes.

CHAPTER 5

ACTIVITY BASED PROBE TO STUDY PROTEIN AMPYLATION

Introduction

The discovery of protein AMPylation mediated by Fic and ATase domain-containing enzymes motivated us to develop tools to study this posttranslational modification. We first developed an anti-tyrosine-AMP and anti threonine-AMP antibody that can detect AMP modified proteins (42). We next sought a different strategy for detection and purification of AMPylated proteins. In collaboration with Howard Hang's group at Rockefeller an ATP analogue, N^6pATP, was designed. The AMP analogous group of N^6pATP, after being transferred onto AMPylation substrates, can be modified by additional tags through a copper-catalyzed azide cycloaddition mechanism (also referred to as "click" reaction). Such tags facilitate the subsequent detection and purification of AMP-modified substrates (54) . We characterized N^6pATP in AMPylation reactions and assessed its applicability as an activity-based AMPylation probe. First, N^6pATP can be used as a bona fide non-radioactive substrate for AMPylation reaction mediated by several AMPylators. Second, modified substrates can be labeled by an affinity purification tag to facilitate the purification and identification of novel AMPylation substrates.

Results

General Reaction Scheme with N^6pATP

Copper catalyzed azide cycloaddition (CuAAC) has been used to synthesize activity based alkynyl chemical probes for detection and purification of posttranslational modifications (also referred to as “click” reaction) (54). An ATP analogue called N^6pATP was synthesized by Markus Grammel and Howard Hang at The Rockefeller University. The N6 position of adenine has a propargyl alkyne that is chemically reactive to azide groups under CuAAC conditions. The azide can be attached to a fluorescent group or biotin to allow for the detection or purification of proteins modified by N^6pATP (Figure 24). Recent structures of Fic and ATase domains suggested the modification of a propargyl group at the N6 position of adenine could be tolerated by these enzymes. Figure 24 describes the general scheme for using N^6pATP to study AMPylation. An AMPylation reaction with AMPylation enzyme (AMPylator), protein substrate, and N^6pATP, generates an AMPylated product. Subsequently, the AMPylated product can be covalently modified by an azide-probe, such as, azido-rhodamine for in gel detection by fluorescence, and azido-azo-biotin to allow for the purification of modified substrates by streptavidin beads (55,56).

N6pATP works in vitro with Fic Domains and ATPase AMPylators

N^6pATP is tested in *in vitro* AMPylation assays of GTPases, catalyzed by various AMPylators. Recombinant proteins used in these studies are shown in Figure 25.

In the presence of N6pATP and VopS, Cdc42 can be modified and subsequently labeled with azido-rhodamine, as detected by a fluorescent band in the SDS-PAGE gel corresponding to modified Cdc42 (Figure 26). Importantly, incubation of N^6pATP with Cdc42 without VopS does not result in labeling. Cdc42-T35A mutant cannot be labeled, suggesting that threonine 35 is required for labeling by N^6pATP. A catalytically inactive mutant VopS-H348A cannot label Cdc42. Labeling with N^6pATP is also seen with other VopS substrates, RacV12 and RhoA. Incubation of Rav12-(T35S), where threonine 35

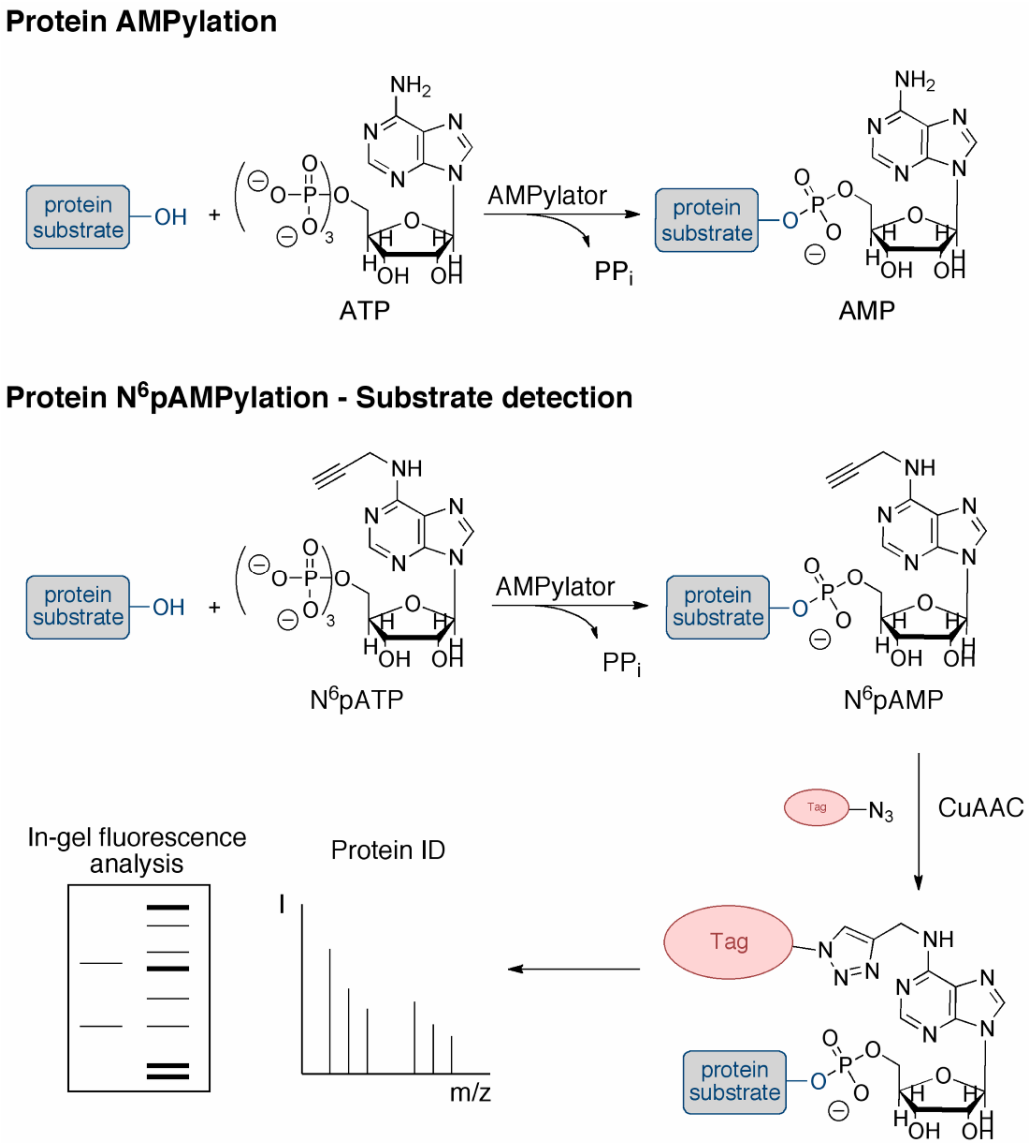

Figure 24. Labeling scheme for detection and identification of substrates. N^6pATP – N6-propargyl adenosine-5'-triphosphate; N6pAMP – N6-propargyl adenosine-5'-monophosphat; PPi – pyrophosphate; CuAAC – Cu(I)-catalyzed azide-alkyne cycloaddition; Tag – rhodamine fluorescence dye or cleavable biotin enrichment tag. Figure adapted from (42).

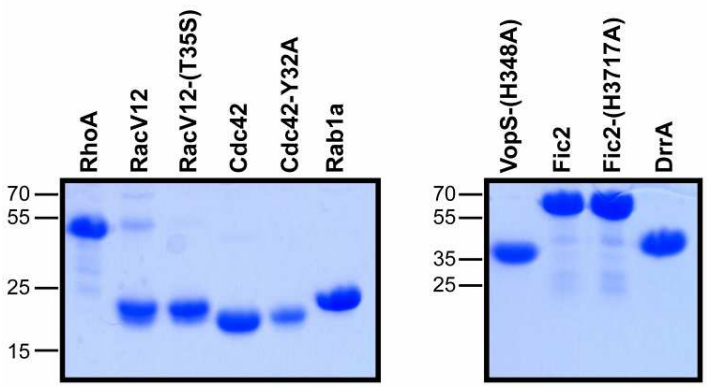

Figure 25. Coomassie stain of protein constructs. Protein were loaded onto SDS-PAGE gels and stained by Coomassie. Coomassie stain of active VopS protein purity can be seen in Figure 8A.

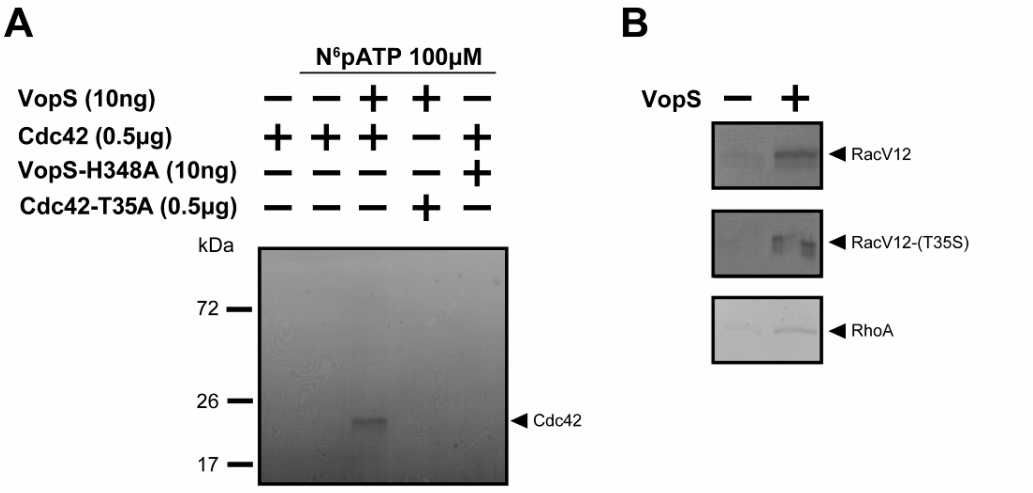

Figure 26. VopS modifies Cdc42 with N^6pATP. (A) Labeling is seen with VopS and Cdc42 in the presence of N^6pATP, but not in the Cdc42-(T35A) mutant. Recombinant proteins were incubated for 30 minutes at 30°C with or without 100 µM N^6pATP. Proteins were then chloroform/methanol precipitated and labeled with azido-rhodamine. Proteins were loaded onto SDS-PAGE gels and scanned at excitation 532 nm/emission 580 nm. **(B**) Click labeling of RacV12 (1 µg), RacV12-(T35S) (1 µg), and RhoA (1 µg) with VopS (0.1 ng).

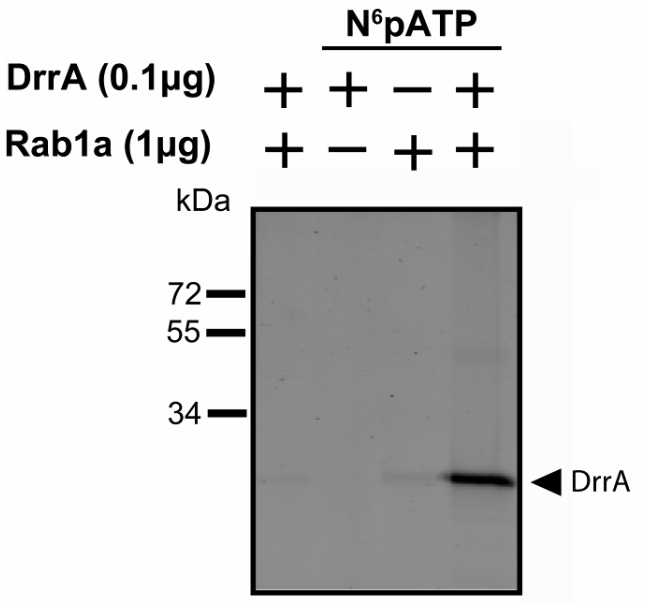

Figure 27 DrrA modifies Rab1a with N^6pATP. Labeling is seen with DrrA and Rab1a in the presence of N^6pATP. Recombinant proteins were incubated for 30 minutes at 30°C with or without 100 µM N^6pATP. Proteins were then chloroform/methanol precipitated and labeled with azido-rhodamine. Protein were loaded onto SDS-PAGE gels and scanned at excitation 532 nm/emission 580 nm.

was changed to a serine residue, demonstrated that VopS could modify a serine residue with AMP (Figure 26 B).

N^6pATP can also be used by other AMPylators. DrrA and Rab1a were incubated with N^6pATP to determine if the chemical probe can be used by AMPylators with AT-ase domains (Figure 27). Figure 27 shows labeling of Rab1a by DrrA in the presence of N^6pATP. The Fic domain region (Fic2) of IbpA in *H. somni* was also tested in an AMPylation reaction using N^6pATP as a substrate. Grammel *et al* 2011, demonstrated that IbpAFic2 can AMP modify tyrosine-77 of Cdc42 with N^6pATP (42).

N^6pATP can detect endogenous substrates in cell lysates

To test whether N^6pATP can detect endogenous AMPylated proteins, HeLa cell lysates were incubated with recombinant VopS and N^6pATP, and subsequently labeled with azide-rhodamine. SDS-PAGE followed by fluorescence detection revealed a number of labeled bands. A band around 21 kDa may represent endogenous Rho family GTPases AMPylated by VopS (Figure 28). Incubation of lysate alone with N^6pATP does not display enriched labeling of the 21 kDa band. Grammel *et al* performed similar experiments with IbpAFic2 and DrrA, and showed that IbpAFic2 and DrrA can label proteins around 21 kDa in HeLa cell lysates (42).

A critical experiment was performed by Grammel *et al*, to test whether endogenous protein substrates can be purified after AMPylation with N^6pATP. HeLa cell lysate was incubated with VopS and N^6pATP, and subsequently labeled with azide-azo-biotin. After affinity purification with streptavidin beads, associated proteins were identified by mass spectrometry. Indeed, N6pAMP modified Cdc42 was enriched in VopS treated samples.

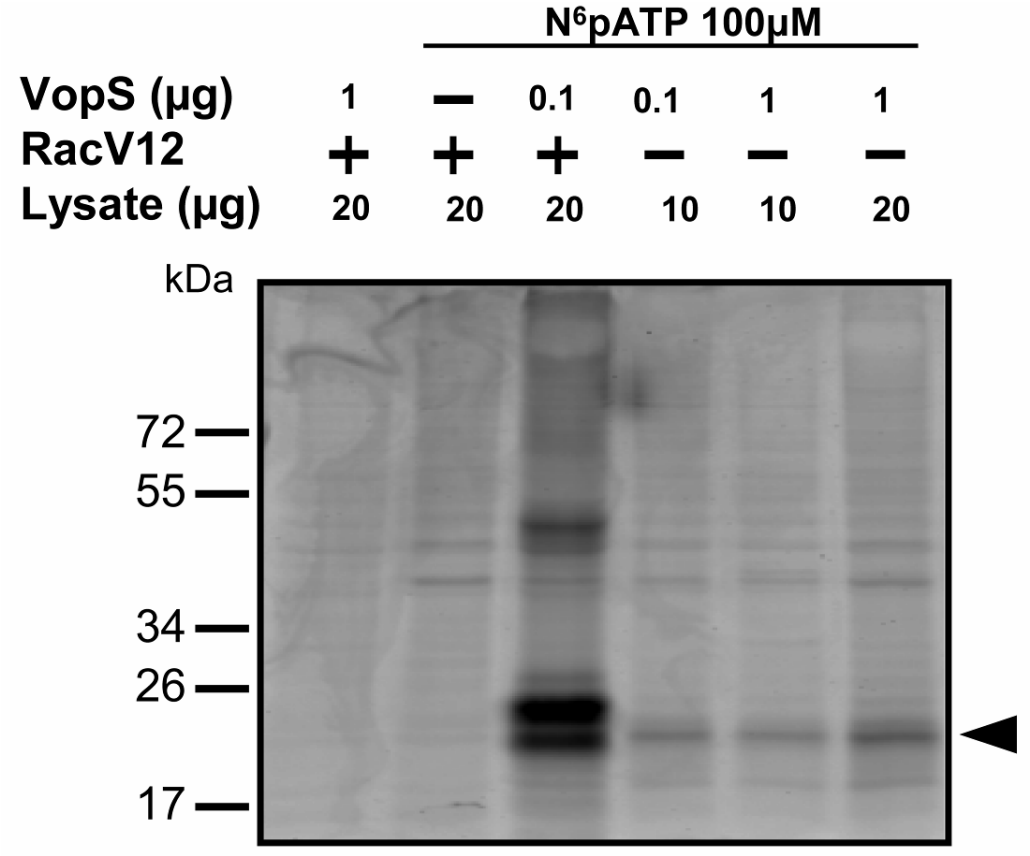

Figure 28. Endogenous labeling of VopS substrates with N^6pATP. HeLa lysates (10 µg or 20 µg) were incubated with N^6pATP (100 µM) and VopS (0.1 µg or 1 µg). Arrow indicates endogenous labeling of GTPases.

Discussion

Activity based protein profiling using click chemistry has been utilized to synthesize probes to detect various post translational modifications (42). In this study an N^6pATP probe is synthesized and demonstrated to be an effective chemical probe for protein AMPylation. N^6pATP can be used as a substrate by two different Fic domain containing enzymes that AMPylate threonine or tyrosine residue of their substrates. An ATase domain containing AMPylator, DrrA, can also use N^6pATP to modify its substrate, Rab1a. This study further demonstrates the sensitivity of the N^6pATP probe by labeling endogenous protein substrates with cell lysates and recombinant AMPylators. Importantly, when tagged with azido-azo-biotin, this probe can be used to purify endogenous AMPylation substrates.

The chemical addition of the propargyl group at the adenine N6 position was tolerated by Fic domains (VopS and IbpAFic2) and an AT-ase domain (DrrA) because both domain families were capable of utilizing N^6pATP as substrate. The AMP bound IbpAFic2 structure shows that the N6 position of adenine points out away from the nucleotide binding site that could accommodate the propargyl group. DrrA does not have a nucleotide bound structure but is structurally similar to an AT-ase domain member called kanamycin nucleotidyltransferase [PDB ID 1KNY]. The AMPCPP is bound to kanamycin nucleotidyltransferase and the N6 position of adenine is directed away from the active site. The N6 position is solvent exposed, which could explain why DrrA could tolerate a propargyl group

CHAPTER 6

PURIFICATION, EXPRESSION AND CRYSTALLIZATION OF VOPQ

Introduction

VopQ is a bacterial effector secreted by the T3SS1 of *V. parahaemolyticus* that is involved in the cytotoxicity observed in tissue culture cells (7). VopQ was previously demonstrated to be necessary and sufficient to induce autophagy in mammalian cells, the biochemical and molecular mechanisms of which were unknown (30). VopQ is not homologous to any known protein domain. To gain structural insight for its molecular function, I decided to crystallize VopQ.

Results

Domain architecture of VopQ

VopQ is a 492 amino acid protein that contains several regions. Contained in the approximately first 30 amino acids at the N-terminus is a secretion signal necessary for effectors to be delivered through the needle complex of T3SS1 (Figure 29A). Amino acids 30-100 of VopQ constitute the chaperone binding domain (CBD) critical for binding to its chaperone, VP1682 (13). C-terminal to the CBD is a region of unknown function that has no known structural similarity to any structure in the PDB database (Figure 29A). The C-terminal domain contains at least two hydrophobic regions. These hydrophobic regions may form a hydrophobic core, or interact with lipid membranes.

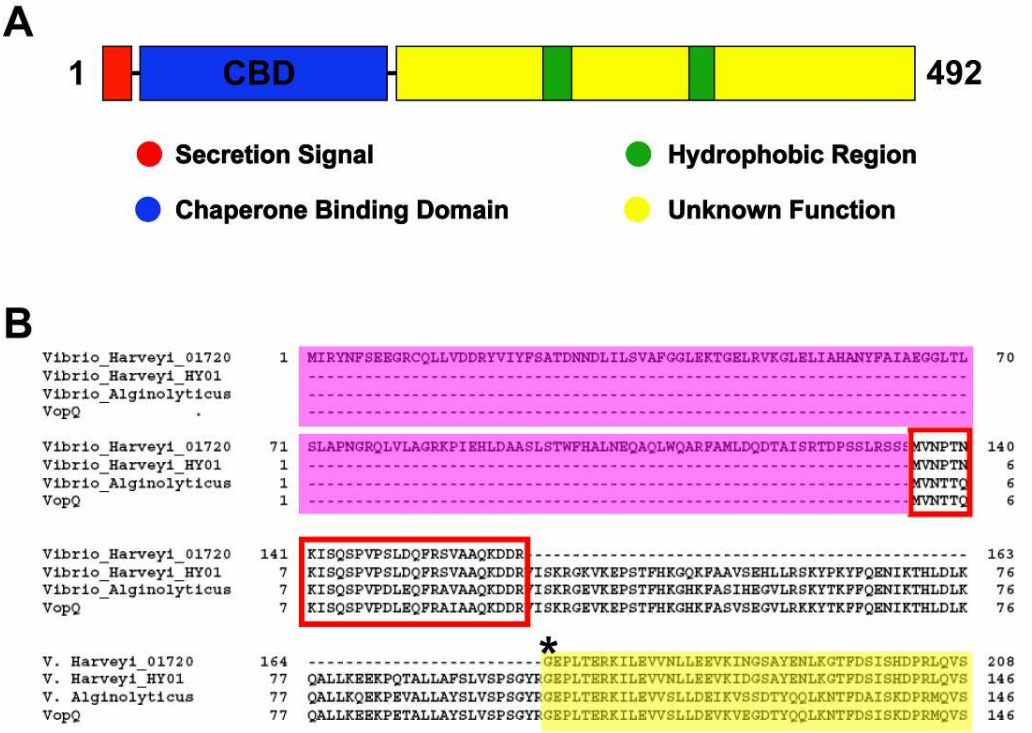

Figure 29. Protein architecture of VopQ. (A) Domain organization of VopQ. Sequence signal (red), CBD region (blue), and C-terminal region (yellow) are noted. Two predicted hydrophobic regions are highlighted as green. **(B)** Protein alignment of VopQ homologs. Highlighted in purple is the chaperone fused to *Vibrio harveyi 01720* homolog. Red box indicates the secretion signal. Highlighted in yellow begins the C-terminal region of unknown function. Star designates the beginning of C-terminal region.

The primary sequence of VopQ is conserved in *Vibrio* species. Shown in Figure 29B is a sequence alignment of VopQ homologs for the first one-third of the protein sequence. Notably, VopQ from *Vibrio harveyi* (*V. harveyi)* 01720 (also known as *Vibrio* BB120 or *Vibrio* BA-1116) is different from other homologs, including the absence of a CBD domain. Also *the V. harveyi* VopQ homolog has a natural built-in chaperone fused to the N-terminus of its secretion signal (red box). The rest of the protein, part of which is shown and highlighted in yellow (starting at residue 102 for *V. para.*), is the domain of unknown function important for VopQ's activity in cells (Figure 29A) (30).

VopQ Construct Design

Several factors were considered when designing expression constructs for VopQ. Secretion signals for effector proteins are disordered and may inhibit protein solubility and interfere with crystallization (10). In the crystal structures of effector proteins with their secretion signal attached, the secretion signal residues are disordered and not observed in the electron density (10). Generally, the CBD region of effectors are not included when designing constructs to express effectors, unless the interaction with chaperones are of particular interest (10). The stability of CBD may require the presence of chaperone, and as seen previously in Figure 2, the CBD has extended structural elements that wraps around the chaperone. The protein sequence alignment of VopQ homologs suggests a defined boundary at residue 102 from *V. para* and *Vibrio alginolyticus* (*V. alginolyticus*) and residue 164 from the *V. harveyi 017202*. VopQ fragments from *V. para* (residues 102-492), *V. alignolyticus* (residues 102-492), and *V. harveyi* 01720 strain (residues 163-554) were cloned into pPROEXHTa vector, expressed in *E. coli* BL21 cells, and purified by nickel chromatography (Figure 30). These constructs do not contain the secretion signal, CBD domain, and chaperone. VopQ from

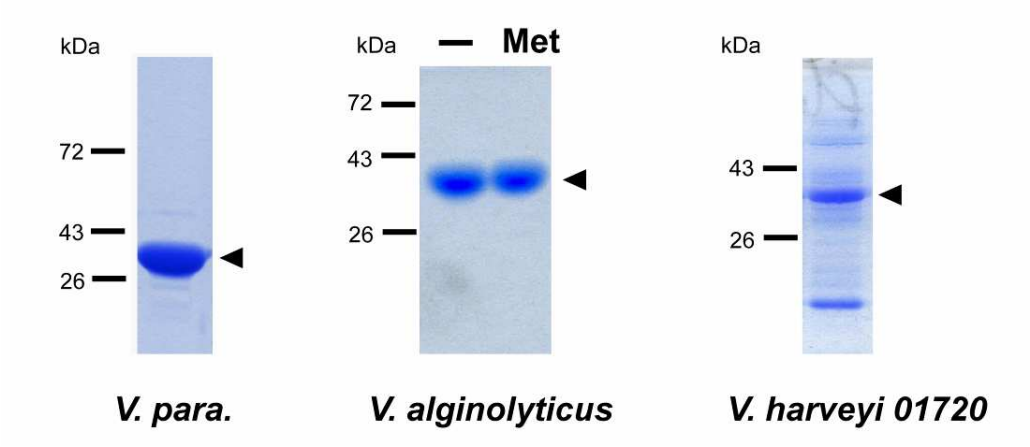

Figure 30. Purification of the C-terminal Region of VopQ. Protein from *V. para* VopQ-(102-492), *V. alignolyticus* VopQ-(102-492), *V. harveyi* 01720 VopQ-(163-554). Coomassie stained gels of VopQ protein after nickel purification.

V. alginolyticus and *V. para* did not yield crystals, and VopQ homolog from *V. harveyi* was impure. Methylation of VopQ protein from *V. alginolyticus* was carried out to change the surface properties of the protein. There were no positive crystallization hits with methylated VopQ from *V. alginolyticus*.

Limited Proteolysis of VopQ

To determine if there are globular domains in VopQ that may be appropriate for crystallization, we tried to analyze VopQ protein (from *V. para.*) with limited proteolysis. We started with full length (FL) VopQ. However, expression and purification of the His-FL-VopQ construct yielded low amounts of protein with degradation products. His-VopQ-(29-492) construct gives a higher protein yield, but most of the purified protein appears to form large aggregates as analyzed with gel filtration. The small amount of stable His-VopQ-(29-492) (non-aggregates) after gel filtration was used for limited proteolysis. Trypsin and chymotrypsin treatment did not generate clear proteolytic products (not shown). Subtilisin and papain digestion generated a relatively stable product around 26kDa. N-terminal Edman sequencing revealed the stable cleavage product starts at residue 248 (Figure 31A). Constructs were designed to express VopQ with the first 247 residues removed, with His or GST tag. However, expression of these construct could not yield soluble, stable and pure protein (Figure 31B).

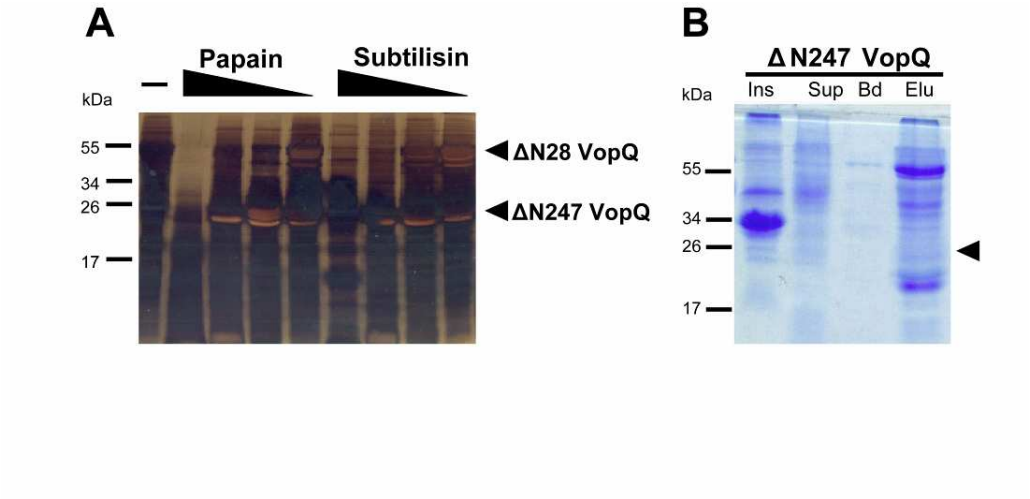

Figure 31. Limited proteolysis of VopQ protein. (A) Silver stain from limited proteolysis of His-VopQ-(29-492) with varying concentrations of papain or subtilisin protease. **(B)** Purification of His-VopQ-(248-492). Insoluble (Ins), supernatant (Sup), beads (Bd) and elute (Elu) are designated. Expected protein size of His-VopQ-(248-492) in Elu fraction is indicated with arrow.

Coexpression of VopQ and VP1682

I next attempted to crystallize the chaperone and VopQ in a complex. To date, there has been no structure of any chaperone in complex with full length effector. His-tagged VP1682 and VopQ with no tag were co-expressed in *E. coli*, and purified with nickel chromatography followed by gel filtration chromatography. VopQ and VP1682 indeed form a complex as shown by their co-purification (Figure 32A and 32B). VopQ in complex with VP1682 become resistant to papain digestion (Figure 32C), suggesting that VopQ assumes a protected compact conformation when bound to VP1682. I noticed that when His-tagged VopQ and untagged VP1682 were co-expressed, a higher amount of stable VopQ can be obtained. VopQ can also be separated from VP1682 by anionic exchange Q column, when pure and stable VopQ (not in complex) is desired.

Crystal trials with various constructs of VopQ and VP1682 complex yield reproducible crystals, but unfortunately the diffraction profiles were exceedingly poor (Figure 32D). Normally good crystals behave like glass in that they shatter when hit by a hammer. The larger crystals of the VopQ and VP1682 complex are soft, such that when poked they are squishy like donuts.

Construction of Chaperone-VopQ Fusion protein

The *V. harveyi 01720* VopQ homolog has a built-in chaperone fused to the C-terminal domain of the effector and so I decided to mimic nature and created a chaperone fused effector using VopQ from *V. para.* (Figure 29). Figure 33 shows the purification and crystallization of a synthetic construct containing chaperone VP1682 fused to VopQ-(102-492) (Figure 33). This construct has the secretion signal but not the CBD domain. Other fusion constructs were also tested (Table 7). Crystals formed within a day. Optimized crystals were allowed to grow for up to a week (Figure 33C). However, the best crystals unfortunately diffracted only to 9.5 Å.

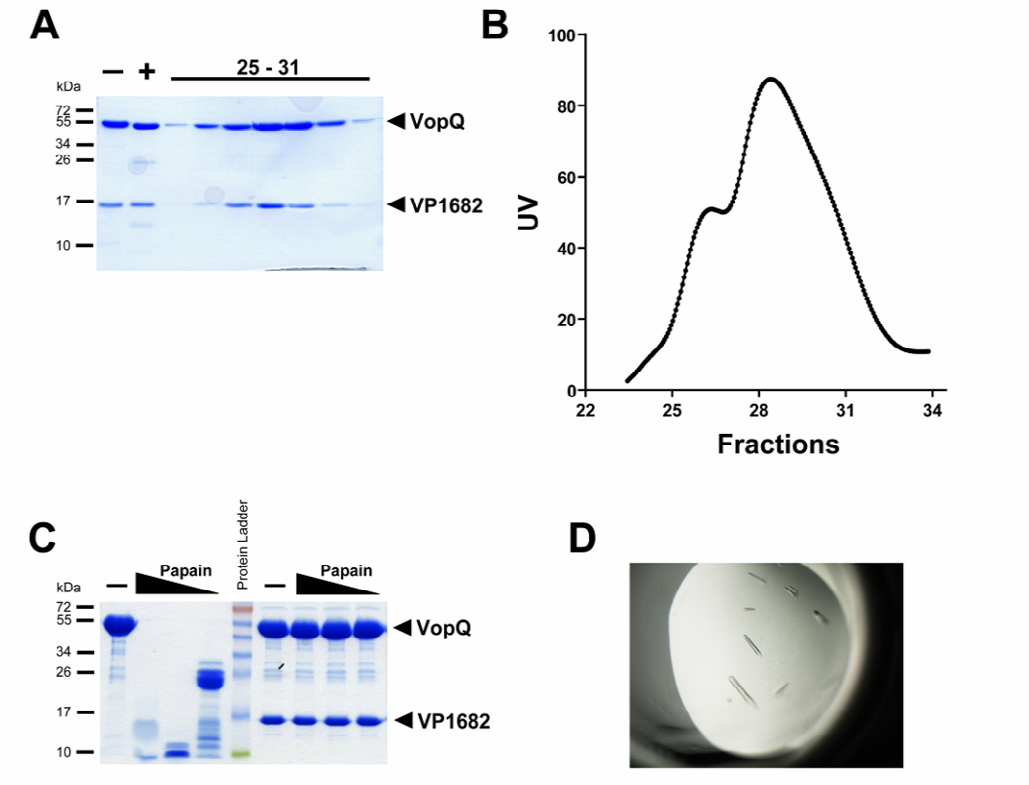

Figure 32. Purification, proteolysis, and crystallization of VopQ complexed with VP1682. (A) Coomassie staining from gel filtration run of VopQ-(1-492) complexed with VP1682. Minus (-)and plus (+) designates before and after TEV cleavage of His tag, respectively. **(B)** UV profile from gel filtration of VopQ-(1-492) complexed with VP1682. **(C)** Limited proteolysis of full length VopQ alone or in complexed with full length VP1682 using papain protease. **(D)** Crystals of full length VopQ complexed with full length VP1682. Crystals grew in 0.1M HEPES pH7.5, 0.1M KCl, 15% PEG 6000. Crystals diffracted to greater than 20Å

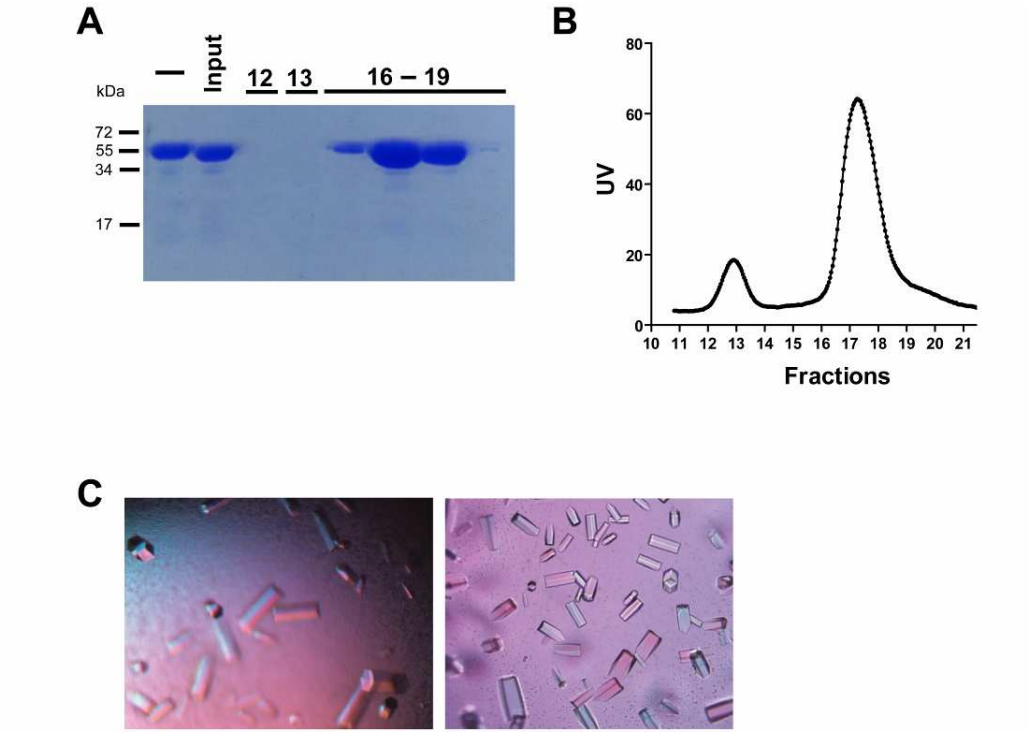

Figure 33. Purification and crystallization of VP1682-(1-142)-VopQ-(102-492) fusion protein. (A) The fusion protein eluted between fractions 16 to 18 on HiTrap QHP column. Minus (-) is fusion protein before TEV protease cleavage of His tag. Input is fusion protein after cleavage and before injection onto HiTrap QHP column**. (B)** The UV profile of fusion protein on the HiTrap QHP column. Crystals of VP1682-(1-142)-VopQ-(102-492). Crystals were grown in 0.1M MES pH6.0, 0.2M NaCl, 10% PEG 4000 (left panel) and 0.1M HEPES pH 7.5, 0.2M NaCl, 10% PEG 8000 (right panel). Crystals from left panel diffracted to about 9.5Å.

Table 7. VopQ Crystallization Trials

Date	Protein	Concentration (mg/ml)	Additives	Temp(°C)
8/14/2008	VopQ-(108-492)	12	None	20
9/4/2008	VopQ-(108-492)	12	0.5% Octylglucoside	20
8/2/2009	VopQ-(77-492)	7	None	20
9/11/2009	VopQ-(77-492)	8	0.5% Octylglucoside	20
9/11/2009	VopQ-(77-492)	8	None	4
4/27/2010	VopQ-(81-492)	8.4	None	20
8/6/2010	His-VopQ-(29-492)	36	None	20
8/6/2010	His-VopQ-(102-492)	10	None	20
8/12/2010	VopQ-(31-492) + VP1682-(1-137)	25	None	20
8/18/2010	VopQ-(102-492)	6.2	None	20
8/19/2010	VopQ-(29-492) + VP1682-(1-137)	23	None	20
9/3/2010	VopQ-(102-492)*	15.7	None	20
9/8/2010	VopQ-(16-492) + VP1682-(1-137)	14.8	None	20
9/10/2010	Methylated-VopQ-(102-492)*	9.7	None	20
9/24/2010	VopQ-(1-492) + VP1682-(1-152)	13.4	None	20
10/7/2010	VopQ-(30-492) + VP1682-(1-137)	21.2	None	20
10/24/2010	VopQ-(1-492) + VP1682-(1-137)	5.7	None	20
10/26/2010	Methylated VopQ-(1-492) + VP1682-(1-137)	15.2	None	20
10/29/2010	VopQ-(1-492) + VP1682-(1-137)	11.7	None	20
10/29/2010	VopQ-(30-492) + VP1682-(1-137)	14.7	None	20
11/4/2010	Methylated VopQ-(30-492)	11	None	20
11/24/2010	VopQ-(30-492)* + VA1682-(1-135)*	13.6	None	20
11/29/2010	VopQ-(30-492) + VP1682-(1-147)	8	None	20
12/30/2010	Fusion: VP1682-(1-142)-VopQ-(102-492)	10	None	20
1/13/2011	Fusion: VP1682-(1-137)-VopQ-(102-492)	10	None	20
2/10/2011	Fusion: VA1682-(1-137)-VopQ-(102-492)*	10	None	20
2/9/2011	Fusion: SeMet* VP1682-(1-142)-VopQ-(102-492)	10	None	20
2/9/2011	Fusion: GSTail* VP1682-(1-142)-VopQ-(102-492)	13.1	None	20
5/4/2011	Fusion: VP1682-(1-142)-VopQ-(102-492)-(K363Y)	8	None	20

Discussion

The biochemical mechanism of VopQ function in mammalian cells is still unclear. Protein sequence analysis of VopQ revealed no homology to any known protein structure suggesting VopQ may have a unique structural fold. Alternatively, VopQ may have low sequence identity and share similarity to a known structural fold. Crystallization trials for VopQ were pursued in order to gain a structural understanding of its function. Various VopQ fragments, complexes, and fusion forms were tested but with no success (Table 7). Until the substrate is discovered it may be difficult to solve the structure of VopQ. When the substrate is known I would recommend using the VopQ-(102-492) construct to complex with the substrate for crystallization trials.

CHAPTER 7

DISCUSSION AND FUTURE DIRECTIONS

Summary of Research Findings

VopS from *V. para*. is an AMPylation enzyme with a role in bacterial pathogenesis. The crystal structure of VopS revealed that it contains two sub domains. The unique N-terminal subdomain is likely involved in protein substrate binding, while the C-terminal Fic domain mediates catalysis. Fic domain contains the conserved Fic motif and a hairpin loop element. Comparisons with other Fic structures suggest that the hairpin element mediates interaction with protein substrate, and Asp352 in the Fic motif mediates binding to MgPPi. Other residues within the Fic motif required for AMPylation activity are identified. Steady-state kinetic analysis revealed that His348 is the catalytic base and the most important residue for activity. Another catalytic residue, Arg299, is required for activity likely by mediating the formation of an AMPylation transition state. Kinetic analysis with its substrates (Cdc42 and ATP) supports a ternary complex reaction mechanism. A chemical probe, N^6pATP, was developed as a tool to study AMPylation. N^6pATP can be used as a substrate for both Fic- and AT-ase-domain AMPylators to modify threonine, serine and tyrosine. N^6pATP can be used for the detection and purification of endogenous AMPylated proteins.

Attempts to solve the structure of VopQ, a Vibrio effector protein of unknown function, were not successful. A number of constructs were made for VopQ, and tested for their suitability for purification and crystallization. I was able to obtain crystals of a chaperone-effector fusion protein, VP1682-VopQ(102-492), however the crystals diffracted poorly. Future crystallization studies could be carried out with VopQ-(102-492) in complex with its cellular protein target/substrate (yet to be identified).

Current Updates in the AMPylation and Fic Domains

Recently, there have been many published articles that provide new insights into AMPylation mechanisms. As mentioned before at the end of Chapter 3, a structure of the Fic-domain-containing protein IbpA from *H. somni* in complex with Cdc42 was solved (45). The catalytic histidine of IbpA Fic domain was intentionally mutated to alanine to generate an inactive enzyme used for crystallization. Interestingly, during refinement the authors found that Cdc42 appeared to be AMP modified on tyrosine 77. The complex structure revealed that protein substrate interacts with the hairpin element of IbpA Fic domain. Protein-protein interactions were observed in the switch 2 region of Cdc42 and the N-terminal region of IbpA that is analogous to the VopS N-terminal subdomain. The complex structure also revealed how the AMP group interacts with the enzyme (45).

Another AMPylator was discovered, encoded by *Legionella pneumophila*, the causative agent for Legionnaires' disease (46). Upon infection, the intracellular pathogen *L. pneumophila* form *Legionella*-containing vacuoles (LCV) that allow the bacteria to safely replicate. In order to form and control the formation and maintenance of LCVs, *L. pneumophila* has usurped the host membrane trafficking machinery. The GTPase Rab1 is a primary target of *L. pneumophila* effectors. The effector DrrA has a C-terminal region that functions as a GEF to interfere with Rab1 function. While trying to determine the structure of the N-terminal region of DrrA, Muller *et al* 2010 discovered that DrrA's N-terminus resembles the GS-ATase domain (57). Indeed, DrrA could modify Rab1 with an AMP group to make the GTPase constitutively active (57). Another study found another *L. pneumophila* effector, SidD could remove AMP from modified Rab1, i.e. "de-AMPylate" Rab1 (58,59). The discovery of DrrA and SidD revealed that protein AMPylation is a reversible posttranslational mechanism important for *Legionella* pathogenesis. Another Fic domain protein from *L. pneumophila*, AnkX, was discovered

to have phosphocholine transferase activity using CDP-choline as substrate to modify Rab1 during infection (60). Therefore, Fic domains may represent a domain family that can catalyze a variety of phosphor-transfer reactions, including AMPylation and phosphocholination.

Future Directions

Introduction to Eukaryotic Fic proteins

Fic domains are evolutionarily conserved from prokaryotes to eukaryotes. The functions of several Fic domain proteins in bacteria have been characterized. However, little is known about the function of Fic domains in higher eukaryotes. What are the physiological functions of eukaryotic Fic proteins? What are the biochemical activities of eukaryotic Fic proteins?

There is one Fic domain containing protein in higher eukaryotic organisms such as human, fly, and worm. These eukaryotic Fic proteins contain a predicted single spanning N-terminal transmembrane region that is not present in VopS, IbpA, or AnkX, a TPR (tetratricopeptide repeat) domain that likely mediates protein-protein interaction, and a C-terminal Fic domain. Below I will discuss how I would characterize the human Fic protein, Huntingtin yeast interaction protein E (HYPE), and other potential Fic proteins.

Determine the physiological function of eukaryotic Fic proteins

One can take a genetic approach to study the functions of Fic proteins in eukaryotes. RNA interference-mediated knockdown can be carried out to specifically disrupt Fic genes in worms, flies and mammalian cells. Available strains of worms and flies carrying Fic gene mutations can also be obtained. Genetic study can also be carried

out in mice. However, one can hardly predict the phenotype of Fic gene disruption. Any physiological phenotype should be examined. Rescue experiments should be performed to confirm the specific defects caused by Fic gene disruption.

Determine the molecular function of HYPE

It has been a couple of years since the lab has initiated studies with HYPE and the "quick and easy" experiments have given a lot of negative data but some interesting positive results. In order to move forward I propose developing a stable cell line system for HYPE that will allow stable expression, stable knockdown of the HYPE gene, as well as knock-in of HYPE mutants. Development of stable cell lines will be a tool to promote cell biology studies and promote molecular studies on the function of HYPE.

Identify the small molecule substrate for HYPE

The biochemical activity for HYPE has not been conclusively determined. *In vitro* AMPylation assays with HYPE to search for substrates had been unsuccessful. In contrast, VopS and IbpA are able to label endogenous substrates in cell lysates. It has been proposed that Fic domains can catalyze a variety of phosphotransfer reactions (60).

I propose to set up experiments to identify small molecule substrates of HYPE. HYPE was shown to have weak AMPylation activity towards Cdc42 *in vitro*, but the amount of enzyme and substrate used in the assay were exceedingly high. The true substrate is probably not Cdc42 or ATP, but these substrates are similar enough to the true substrates to promote the phosphotransfer activity. GTPases share a conserved structural fold, and Fic domains have shown to target Rho GTPases and Rab family GTPases making it reasonable to think that HYPE might also target GTPases in eukaryotes. *In vitro*, VopS can also use GTP, CTP, and UTP, though much less efficiently compared to ATP (61) .

I propose to co-express HYPE and Cdc42 in eukaryotic cells (such as yeast cells, mammalian cells, etc.). Since HYPE is predicted to be in the endoplasmic reticulum (ER), a heterologous ER-localizing signal may be fused to Cdc42 to facilitate its interaction with HYPE in the ER. After coexpression, the Cdc42 substrate will be purified and its mass determined by mass spectrometry. Generic substrates such as myelin basic protein, histone, and casein have been used for kinases, methyltransferase, acetyltransferase, and ubiquitin ligases *in vitro*. Such proteins could also be tested as substrate for HYPE. HYPE was originally identified in a screen to interact with the N-terminus of huntingtin protein (62). Huntingtin could also be a possible generic or real substrate and might be helpful in determining the type of modification used by HYPE.

Characterize eukaryotic homologs that naturally do not have a transmembrane region

Biochemical studies with human Fic protein HYPE, *Caenorhabditis elegans* Fic protein, and the *Drosophila melanogaster* Fic protein have not progressed in terms of identifying the protein substrate and it is unclear if the homologs can perform AMPylation. All three homologs have a transmembrane region at their N-terminus that is predicted to localize these proteins to membrane compartments. Having the transmembrane region complicates purification and the assay development when trying to identify an activity. Various constructs have been tested from these homologs but results have been negative in terms of observing substrate labeling.

I propose that there should be a focus on working with eukaryotic splice variants that naturally exist and do not have the N-terminal transmembrane region. In mouse (*Mus musculus*) two splice variants have been identified. The long form has the transmembrane region but the short form does not. I propose to use the short soluble form for biochemical characterization studies. Also the Sumatran orangutan (*Pongo abelii*) has a long and short form with the short form lacking the transmembrane region. I also

recommend the yellow fever mosquito *Aedes aegypti* and the malaria mosquito *Anopheles gambiae* because they lack the transmembrane region and because of their influence in human infectious diseases. The water flea *Daphnia pulex* and the lancet *Branchiostoma floridae* lack N-terminal transmembrane regions and are potential candidates but they do not have strong disease relevance.

The Future of the AMPylation Field

In recent years, protein AMPylation has emerged as an important mechanism utilized by bacterial pathogens to disrupt host cell signaling during infection. Fic domains and AT-ase domains can catalyze AMPylation and also enzymes that mediate de-AMPylation have been identified (29,46,58) (59). AMPylation is a versatile modification that can activate or inactivate signaling proteins

There is only one Fic gene encoded in higher organisms such as flies, mouse and human and the molecular and physiological function of these Fic protein are unknown. It is not clear if eukaryotic Fic proteins can perform AMPylation, phosphocholination, or another type of phosphotransfer reaction. Genetic knockout models of the Fic gene will provide an important tool and foundation to study the function of Fic protein in higher eukaryotes. Thousands of genes containing Fic domains have been identified in bacteria that appear to function in bacterial signaling systems (22). The physiological and molecular function of prokaryotic Fic domains has remained unexplored.

The significance of protein AMPylation will be tested as researchers determine if AMPylation is a regulatory mechanism used in eukaryotic signaling. Also the importance of Fic domains and their impact in signal transduction pathways of prokaryotic and eukaryotic systems will require additional characterization of more Fic proteins. Fic domains appear to be versatile catalytic modules capable of performing a variety of

phosphotransfer reactions. As the field explores the function of thousands of other Fic proteins, the variety of transfer reactions mediated by Fic proteins may be expanded.

BIBLIOGRAPHY

1. Orth, K. (2007) *Microbe* **2**, 183-186
2. Galan, J. E., and Wolf-Watz, H. (2006) *Nature* **444**, 567-573
3. Nair, G. B., Ramamurthy, T., Bhattacharya, S. K., Dutta, B., Takeda, Y., and Sack, D. A. (2007) *Clin Microbiol Rev* **20**, 39-48
4. Ono, T., Park, K. S., Ueta, M., Iida, T., and Honda, T. (2006) *Infect Immun* **74**, 1032-1042
5. Troisfontaines, P., and Cornelis, G. R. (2005) *Physiology (Bethesda)* **20**, 326-339
6. Park, K. S., Ono, T., Rokuda, M., Jang, M. H., Iida, T., and Honda, T. (2004) *Microbiol Immunol* **48**, 313-318
7. Burdette, D. L., Yarbrough, M. L., Orvedahl, A., Gilpin, C. J., and Orth, K. (2008) *Proc Natl Acad Sci U S A* **105**, 12497-12502
8. Broberg, C. A., Zhang, L., Gonzalez, H., Laskowski-Arce, M. A., and Orth, K. *Science*
9. Page, A. L., and Parsot, C. (2002) *Mol Microbiol* **46**, 1-11
10. Lilic, M., Vujanac, M., and Stebbins, C. E. (2006) *Mol Cell* **21**, 653-664
11. Stebbins, C. E., and Galan, J. E. (2001) *Nature* **414**, 77-81
12. Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T., Tagomori, K., Iijima, Y., Najima, M., Nakano, M., Yamashita, A., Kubota, Y., Kimura, S., Yasunaga, T., Honda, T., Shinagawa, H., Hattori, M., and Iida, T. (2003) *Lancet* **361**, 743-749
13. Akeda, Y., Okayama, K., Kimura, T., Dryselius, R., Kodama, T., Oishi, K., Iida, T., and Honda, T. (2009) *FEMS Microbiol Lett* **296**, 18-25
14. Casselli, T., Lynch, T., Southward, C. M., Jones, B. W., and DeVinney, R. (2008) *Infect Immun* **76**, 2202-2211
15. Yarbrough, M. L., Li, Y., Kinch, L. N., Grishin, N. V., Ball, H. L., and Orth, K. (2009) *Science* **323**, 269-272
16. Worby, C. A., Mattoo, S., Kruger, R. P., Corbeil, L. B., Koller, A., Mendez, J. C., Zekarias, B., Lazar, C., and Dixon, J. E. (2009) *Mol Cell* **34**, 93-103
17. Hall, A. (1998) *Science* **279**, 509-514
18. Rossman, K. L., Der, C. J., and Sondek, J. (2005) *Nat Rev Mol Cell Biol* **6**, 167-180
19. Ellenbroek, S. I., and Collard, J. G. (2007) *Clin Exp Metastasis* **24**, 657-672
20. DerMardirossian, C., and Bokoch, G. M. (2005) *Trends Cell Biol* **15**, 356-363
21. Madaule, P., and Axel, R. (1985) *Cell* **41**, 31-40
22. Kinch, L. N., Yarbrough, M. L., Orth, K., and Grishin, N. V. (2009) *PLoS ONE* **4**, e5818
23. Luong, P., Kinch, L. N., Brautigam, C. A., Grishin, N. V., Tomchick, D. R., and Orth, K. (2010) *J Biol Chem* **285**, 20155-20163
24. Yarbrough, M. L., and Orth, K. (2009) *Nat Chem Biol* **5**, 378-379

25. Jiang, P., Pioszak, A. A., and Ninfa, A. J. (2007) *Biochemistry* **46**, 4117-4132
26. Jiang, P., and Ninfa, A. J. (2009) *Biochemistry* **48**, 415-423
27. Xu, Y., Zhang, R., Joachimiak, A., Carr, P. D., Huber, T., Vasudevan, S. G., and Ollis, D. L. (2004) *Structure* **12**, 861-869
28. Xu, Y., Carr, P. D., Vasudevan, S. G., and Ollis, D. L. *J Mol Biol* **396**, 773-784
29. Stadtman, E. R. (2001) *J Biol Chem* **276**, 44357-44364
30. Burdette, D. L., Seemann, J., and Orth, K. (2009) *Mol Microbiol* **73**, 639-649
31. Yu, L., Strandberg, L., and Lenardo, M. J. (2008) *Autophagy* **4**, 567-573
32. Levine, B., and Kroemer, G. (2008) *Cell* **132**, 27-42
33. Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) *Acta Crystallogr D Biol Crystallogr* **62**, 859-866
34. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr D Biol Crystallogr* **58**, 1772-1779
35. Otwinowski, Z. (1991) in Isomorphous Scattering and Anomalous Replacement (Wolf, W., Evans, P.R., and Leslie, A. G. W., eds) pp. 80-86. Science & Engineering Research Council, Cambridge
36. Cowtan, K., and Main, P. (1998) *Acta Crystallogr D Biol Crystallogr* **54**, 487-493
37. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat Struct Biol* **6**, 458-463
38. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard. (1991) *Acta Crystallogr A* **47 (Pt 2)**, 110-119
39. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) *Acta Crystallogr D Biol Crystallogr* **58**, 1948-1954
40. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) *Nucleic Acids Res* **35**, W375-383
41. DeLano, W. L. (2002) The PyMOL Molecular Graphics System.
42. Grammel, M., Luong, P., Orth, K., and Hang, H. C. *J Am Chem Soc*
43. Das, D., Krishna, S. S., McMullan, D., Miller, M. D., Xu, Q., Abdubek, P., Acosta, C., Astakhova, T., Axelrod, H. L., Burra, P., Carlton, D., Chiu, H. J., Clayton, T., Deller, M. C., Duan, L., Elias, Y., Elsliger, M. A., Ernst, D., Feuerhelm, J., Grzechnik, A., Grzechnik, S. K., Hale, J., Han, G. W., Jaroszewski, L., Jin, K. K., Klock, H. E., Knuth, M. W., Kozbial, P., Kumar, A., Marciano, D., Morse, A. T., Murphy, K. D., Nigoghossian, E., Okach, L., Oommachen, S., Paulsen, J., Reyes, R., Rife, C. L., Sefcovic, N., Tien, H., Trame, C. B., Trout, C. V., van den Bedem, H., Weekes, D., White, A., Hodgson, K. O., Wooley, J., Deacon, A. M., Godzik, A., Lesley, S. A., and Wilson, I. A. (2009) *Proteins* **75**, 264-271
44. Palanivelu, D. V., Goepfert, A., Meury, M., Guye, P., Dehio, C., and Schirmer, T. *Protein Sci* **20**, 492-499

45. Xiao, J., Worby, C. A., Mattoo, S., Sankaran, B., and Dixon, J. E. (2010) *Nat Struct Mol Biol* **17**, 1004-1010
46. Muller, M. P., Peters, H., Blumer, J., Blankenfeldt, W., Goody, R. S., and Itzen, A. (2010) *Science*
47. Traut, T. W. (1994) *Mol Cell Biochem* **140**, 1-22
48. Jilkine, A., Maree, A. F., and Edelstein-Keshet, L. (2007) *Bull Math Biol* **69**, 1943-1978
49. Michaelson, D., Silletti, J., Murphy, G., D'Eustachio, P., Rush, M., and Philips, M. R. (2001) *J Cell Biol* **152**, 111-126
50. Sallese, M., Pulvirenti, T., and Luini, A. (2006) *EMBO J* **25**, 2663-2673
51. Chen, G., Porter, M. D., Bristol, J. R., Fitzgibbon, M. J., and Pazhanisamy, S. (2000) *Biochemistry* **39**, 2079-2087
52. Patnaik, D., Chin, H. G., Esteve, P. O., Benner, J., Jacobsen, S. E., and Pradhan, S. (2004) *J Biol Chem* **279**, 53248-53258
53. Tanner, K. G., Langer, M. R., Kim, Y., and Denu, J. M. (2000) *J Biol Chem* **275**, 22048-22055
54. Sletten, E. M., and Bertozzi, C. R. (2009) *Angew Chem Int Ed Engl* **48**, 6974-6998
55. Charron, G., Zhang, M. M., Yount, J. S., Wilson, J., Raghavan, A. S., Shamir, E., and Hang, H. C. (2009) *J Am Chem Soc* **131**, 4967-4975
56. Yang, Y. Y., Grammel, M., Raghavan, A. S., Charron, G., and Hang, H. C. *Chem Biol* **17**, 1212-1222
57. Muller, M. P., Peters, H., Blumer, J., Blankenfeldt, W., Goody, R. S., and Itzen, A. *Science* **329**, 946-949
58. Neunuebel, M. R., Chen, Y., Gaspar, A. H., Backlund, P. S., Jr., Yergey, A., and Machner, M. P. *Science* **333**, 453-456
59. Tan, Y., and Luo, Z. Q. *Nature* **475**, 506-509
60. Mukherjee, S., Liu, X., Arasaki, K., McDonough, J., Galan, J. E., and Roy, C. R. *Nature* **477**, 103-106
61. Mattoo, S., Durrant, E., Chen, M. J., Xiao, J., Lazar, C. S., Manning, G., Dixon, J. E., and Worby, C. A. *J Biol Chem* **286**, 32834-32842
62. Faber, P. W., Barnes, G. T., Srinidhi, J., Chen, J., Gusella, J. F., and MacDonald, M. E. (1998) *Hum Mol Genet* **7**, 1463-1474