2014

The Structural and Biochemical Characterization of Salmonella Host Specificity Determinant Gifsy-2 Gene E

Amanda C. Kohler

THE STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF *SALMONELLA* HOST SPECIFICITY DETERMINANT GIFSY-2 GENE E

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Amanda C. Kohler

June 2014

THE STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF *SALMONELLA* HOST SPECIFICITY DETERMINANT GIFSY-2 GENE E

Amanda C. Kohler, Ph.D.
The Rockefeller University 2014

Salmonella enterica continues to be a significant public health concern, causing an estimated 93.8 million cases of non-typhoidal salmonellosis and 21 million cases of typhoid fever worldwide each year. There are thousands of *Salmonella enterica* serovars, some with a very specific host set, and others that cause disease in rodents, birds, livestock, domestic fowl, and humans alike. In recent years, there has been much progress in the delineation of *Salmonella* infection, with the goal of understanding *Salmonella* pathogenesis at the molecular level. *Salmonella* produces many different effector proteins capable of interacting with and altering numerous biological pathways in the host – enabling host invasion and intracellular survival as well as dissemination and transmission.

We used X-ray crystallography to characterize Gifsy-2 Gene E (GtgE), an effector protease from broad-host serovars of *Salmonella*, that affords these serovars, at least in part, with the ability to maintain a diverse host repertoire. GtgE modulates vesicular trafficking of the *Salmonella*-containing vacuole by cleaving Rab GTPases, Rab29, Rab32, and Rab38, thereby preventing the delivery of antimicrobial products to the vacuole. In order to gain an understanding of GtgE's proteolytic mechanism, we determined the structure of

GtgE to 1.65Å using single-wavelength anomalous diffraction, and through structure-based mutagenesis and *in vitro* activity assays, we established the catalytic triad of GtgE, Cys45-His151-Asp169. We also examined a panel of cysteine protease inhibitors and found that N-ethylmaleimide, chymostatin, and antipain were capable of inhibiting GtgE activity *in vitro*. Furthermore, through work with the catalytically inactive mutant of GtgE (GtgE-C45A), we were able to identify the conditions necessary to form a stable complex between GtgE and Rab38, which may prove useful for further structural work and reveal the nature of GtgE's interaction with its Rab GTPase substrates.

Additionally, we investigated SipC, a *Salmonella* translocase protein with two effector domains – an N-terminal actin bundling domain and a C-terminal actin nucleation domain. We sought to define minimal constructs of these effector domains for crystallization studies, and obtained needle-like spherulites with the C-terminal domain. We also showed that mouse Exo70 is able to pull-down the C-terminal domain of SipC from cell lysate. Finally, we examined *Salmonella* effector AvrA, which has been attributed with having deubiquitinase and acetyltransferase activity, although its role in *Salmonella* pathogenesis remains poorly understood. We determined a minimal construct of AvrA that contains the proposed catalytic triad, and through a yeast two-hybrid experiment, identified ARFGEF2 as a potential interacting partner for AvrA.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. C. Erec Stebbins, for his patient guidance and support throughout my education in his laboratory. I greatly appreciate his willingness to discuss and troubleshoot with me when I encountered difficult obstacles in my work, as well as the freedom he gave me to experiment with my own ideas in the lab. Thank you to all the members of the Laboratory of Structural Microbiology for their advice and camaraderie through the years, and thank you to our collaborator on the GtgE project, Dr. Stefania Spanò, from Dr. Jorge Galán's lab at Yale University. I would also like to express my gratitude to my committee members, Dr. Vincent Fischetti, Dr. Seth Darst, and Dr. Craig Roy, for being generous with their time and offering valuable insight on my research.

Additionally, I would like to acknowledge my family and friends for all their support during my graduate studies. I am deeply thankful for the love, encouragement, and unwavering trust that my parents have always bestowed upon me, instilling in me a confidence of my own abilities. Thank you also to my dear friend, Matthew Holt, for always providing a listening ear and sound advice.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

1.1 *Salmonella enterica*

Salmonella enterica (*S. enterica*) is a rod-shaped, flagellated, Gram-negative bacterium comprised of six subspecies [1]. Of these subspecies, only one, *Salmonella enterica* subspecies *enterica*, is capable of causing disease in vertebrate animals [1]. *S. enterica* subspecies *enterica* is further divided into thousands of serovars distinguished by the antigenic polymorphisms of their flagellar, carbohydrate, and lipopolysaccharide structures [2, 3]. These serovars vary in their host specificity. *Salmonella enterica* serovar typhimurium (*S.* typhimurium) is a broad-host serovar that is capable of infecting a wide range of hosts. On the other end of the spectrum, *Salmonella enterica* serovar typhi (*S.* typhi) is a human-adapted serovar, only causing disease in higher primates [4]. Host-specificity is believed to have evolved through the acquisition of new genetic material via horizontal gene transfer and loss of functional genes through genome degradation [5]. Genetic degradation, which involves the loss of genetic information by gene deletion or pseudogene formation, is evident in strains of *S.* typhi; 210 pseudogenes have been identified in *S.* typhi, while only 39 pseudogenes are present in *S.* typhimurium [4, 6-9].

1.2 Pathogenesis

S. enterica infection can manifest itself as gastroenteritis, Typhoid (enteric) fever, bacteremia, or chronic asymptomatic carriage [2]. In humans, the severity of disease caused by *S. enterica* subspecies *enterica* varies based on host health as well as the identity of the infecting serovar [3]. Nontyphoidal *Salmonella* poses a considerable global health threat, and in 2010, there were 93.8 million cases estimated worldwide, of which approximately 155,000 resulted in death [10]. Of the more than 21 million cases of typhoid fever reported worldwide each year, upwards of 200,000 result in death, mostly in underdeveloped countries [11]. Those infected with *S.* typhi are prone to symptom relapse, and multidrug-resistant strains of *S.* typhi capable of infecting otherwise healthy individuals are widespread, making *S.* typhi a significant global-health concern [11-13].

In humans, *S. enterica* is predominantly acquired through fecal-oral transmission, typically via the ingestion of contaminated food or water. *S. enterica* is able to survive the low pH of the stomach, going on to colonize the endothelial cells of the gastrointestinal tract [14]. Other *S. enterica* targets include: macrophages, neutrophils, monocytes, dendritic cells, granulocytes, M cells, B cells, and T cells [15]. Infection of the gut epithelial cells is known as gastroenteritis, and accounts for roughly 26% of food poisoning cases in the United States [14, 16]. Gastroenteritis is characterized by the acute onset of

fever, chills, nausea, vomiting, abdominal cramping, and diarrhea [17]. Although this infection is usually self-limiting, it can cause death in infants, the elderly, and immunocompromised individuals [17].

1.2.1 *Salmonella* Pathogenicity Island-1: Intestinal Invasion

Before host cell invasion can occur, the bacterial cells adhere to the intestinal epithelial cell surface using fimbriae, which are short, proteinaceous structures extending from the bacterial surface that mediate attachment and can confer attachment specificity [3, 14]. This adherence is reversible and in order to irreversibly dock to and invade the host, *Salmonella* relies on the actions of various proteins encoded by *Salmonella* pathogenicity islands, plasmids, functional prophages, and phage remnants [18, 19]. *Salmonella* pathogenicity islands are large, virulence factor-containing segments of DNA that are inserted into chromosomal DNA, usually flanked by direct repeats and associated with tRNA loci and insertion sequence (IS) elements [3, 20]. These pathogenicity islands often contain mobility genes, such as integrases, transposases, IS elements, and origins of plasmid replication, and have a lower G-C content, roughly 37-47%, than is typical for bacterial chromosomal DNA, which has a 52% G-C content [20]. *Salmonella* pathogenicity island I (SPI-1) is activated upon docking and contains the genetic information necessary to produce and assemble a type III secretion system (T3SS1) [2].

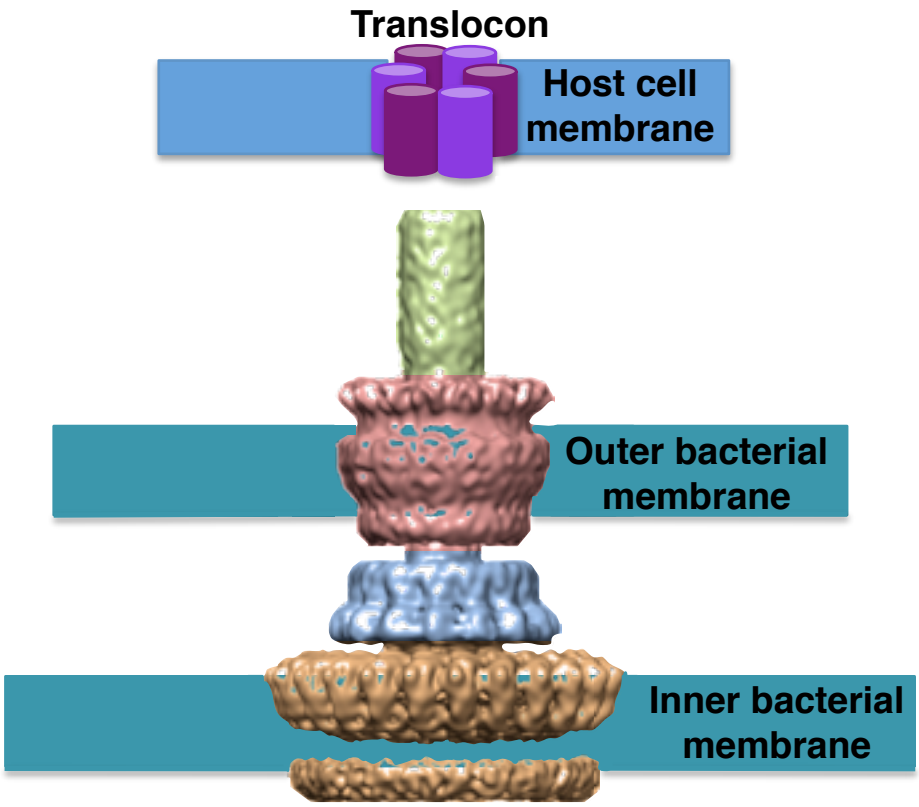


Figure 1. The type III secretion system. The T3SS apparatus is comprised of an inner (orange) and outer (red) ring that allow for passage through the inner and outer bacterial membranes, and are linked by a neck domain (blue). The needle (green) is anchored at the base of the neck via the inner rod, and extends from the surface of the bacterial cell [21]. The translocon pore (purple) is formed by an oligomerization of two bacterial effector/translocase proteins. This system creates a pathway through which bacterial effector proteins can travel from the bacterial cytoplasm into the host. EM images adapted from Schraidt, *et al.* and Kosarewicz, *et al.* [22, 23].

T3SS1 is a needle-like structure, often referred to as the injectisome, that serves as a molecular syringe with the ability to deliver bacterial effector proteins directly into the host cell (Fig. 1). T3SS1 is one of at least six different secretion systems found in Gram-negative bacteria, and is believed to be evolutionarily related to the bacterial flagellum [24]. The needle apparatus is composed of four main parts: inner rings, a membrane-spanning neck, outer rings, and an external needle complex (Fig. 1). In addition to these components, which are physically associated with the bacterial cell, T3SS1 also includes a pore complex, the translocon, that assembles within the host cell membrane [25]. Taken together, the T3SS1 components create a pathway through which effector proteins, proteins that possess the ability to modulate cellular mechanisms in the host, can travel from the bacterial cytoplasm into the host cytoplasm.

An alternative hypothesis for effector translocation, in which the T3SS effector mechanism functions similarly to the AB toxin delivery system, was recently proposed by Edgren, *et al.* [26]. The authors suggest a two-step model for translocation of effector proteins that entails: first, the secretion of translocases and effectors; and second, the translocation of effectors into the host cell upon host cell contact. Translocation is predicted to occur in an AB toxin-like fashion, where the translocase acts as the pore-forming B subunit and the effector functions as the catalytic A subunit. In this model, the needle complex serves as a sensory structure used to identify the host cell, and as a

pathway for the extracellular secretion of effectors [26]. This translocation mechanism is based on data gathered on *Yersinia*, another bacterial species utilizing a T3SS, showing that both *Yersinia* Yop translocases and effector proteins are present on the bacterial surface prior to host cell contact [27], and that a complex of translocases YopB and YopD with effector YopE occurs extracellularly, prior to host invasion [28]. Further support has come from a study by Rosqvist, *et al.* that identifies two *Yersinina* effectors, YopE and YopD, as having AB toxin activity [29]. However, it remains unclear whether the proposed AB toxin-like translocation mechanism applies only to *Yersinia* or if it can be extended to all T3SS-utilizing bacteria as well. To date, translocase homologs in *Shigella* and *Salmonella* have not been shown to have AB toxin activity, nor have extracellular intermediate translocase-effector complexes been identified in *Shigella* or *Salmonella*, although the translocase proteins do appear to associate prior to translocon formation [30, 31]. Since direct evidence for the translocon is lacking and the methods of translocase insertion and translocon formation remain unknown, additional studies will be required in order to discern the true mode of effector translocation.

Once translocated into the host, effector proteins interact with and alter host pathways in a variety of ways. The initial wave of delivered effectors performs various functions aimed at encouraging host uptake of the bacterial cell, including stimulating GTPase activation and actin polymerization in order to

induce the phagocytic engulfment of the *Salmonella* cell by the host [32]. As the bacterial cell enters the host it becomes enveloped in a membrane-bound vesicle, termed the *Salmonella*-containing vacuole (SCV), which undergoes a maturation process and serves as the site of replication for the bacterium [33] (Fig. 2).

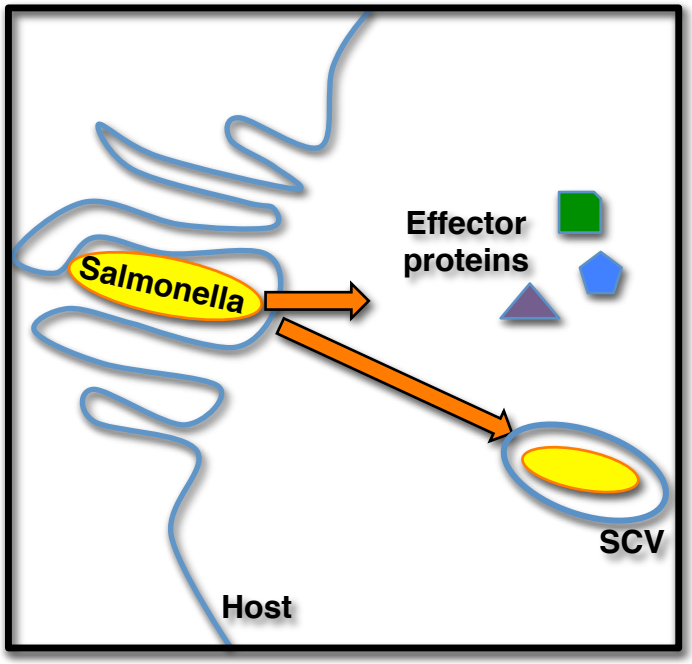


Figure 2. *Salmonella* enters the host cell. In order to invade its host, *Salmonella* injects effector proteins into the host cell that stimulate the bacterium's uptake and enable it to survive and replicate inside the *Salmonella*-containing vacuole (SCV).

1.2.2 *Salmonella* Pathogenicity Island-2: Survival and Replication

Upon internalization, *Salmonella* begins expressing SPI-2-encoded effector proteins in response to various environmental cues, such as low osmolarity, acidification of the SCV, and decreased levels of certain nutrients [34]. SPI-2 also encodes a second T3SS, T3SS2, which is structurally similar to that of T3SS1. T3SS2 functions to transport bacterial effectors across the SCV membrane and into the host cytoplasm where they can then interact with their targets [14]. *Salmonella* utilizes a variety of different effector proteins to attenuate the host's defense mechanisms and increase its chances for survival and replication inside the SCV [3].

T3SS2 also plays a role in the biogenesis of the SCV, which is a modified phagosome with characteristics of a late endosome [35]. The eukaryotic endocytic pathway is a degradation pathway that involves endosome maturation and subsequent development into a functional hydrolytic phagolysosome for cargo degradation [36]. In its early stage of biogenesis, the SCV shows the transient presence of early endocytic markers, mainly EEA1, Rab GTPase 5, and transferrin receptor [37]. These markers are replaced by the late endocytic indicators: lysosomal-associated membrane proteins (LAMP) 1 and 2, vacuolar ATPase (vATPase), and cholesterol [36, 38, 39]. LAMP1 and LAMP2 are integral membrane proteins involved in membrane fusion and the biogenesis of lysosomes, and are often present in late endosomes. V-ATPase causes the

acidification of the SCV lumen, and during this stage, the SCV moves into a juxtanuclear position proximal to the microtubule organizing center [40-42]. At this point in development, *Salmonella* begins replicating inside the SCV, and extensive membrane tubules, or *Salmonella*-induced filaments, begin to form, originating from the SCV and projecting throughout the host cell [43, 44]. The SCV is able to circumvent the degradation step of the endocytic pathway by preventing the assembly of a functional hydrolytic phagolysosome [36].

1.2.3 Dissemination and Transmission

In order to continue propagation, *S. enterica* must move from one host to the next. There are several mechanisms by which *Salmonella* can cause host cell death and enable its escape from the host: epithelial cell apoptosis, rapid T3SS1-dependent pyroptosis, and delayed T3SS2-dependent macrophage pyroptosis [45]. Apoptosis of the host is a form of programmed cell death that involves the dismantling of the cell by executioner caspases and phagocytosis of the resulting apoptotic bodies by phagocytic cells without eliciting an inflammatory response [45]. During the replication phase, *Salmonella* is able to delay apoptosis for 12-18 hours in its epithelial host, creating a more stable intracellular environment for ample proliferation, and when ready to leave the host, the bacterium triggers apoptosis in epithelial cells through the actions of T3SS1 and T3SS2 effectors [18, 46, 47].

In macrophages, *S. enterica* can cause either a rapid T3SS1-dependent or a delayed T3SS2-dependent programmed cell death known as pyroptosis [18]. Unlike apoptosis, pyroptosis is a proinflammatory pathway dependent on caspase-1 activation, which mounts an immune response that terminates in cell lysis [45, 48]. During rapid T3SS1-dependent pyroptosis, several T3SS1 effectors function together with caspase-1 to cause cell lysis within 2-3 hours post-infection [49]. In contrast, delayed T3SS2-dependent pyroptosis relies on specific T3SS2 effector proteins and caspase-1 activity to cause apoptosis-like macrophage death 18-24 hours post-infection [50]. *Salmonella* employs rapid or delayed pyroptosis depending on the given physiological conditions, tissue location, and stage of infection [49]. Rapid pyroptosis occurs during the intestinal phase of infection, and may aid in systemic spread of infection through the recruitment of phagocytes and increased inflammation [49]. Once systemic, delayed pyroptosis is utilized for intracellular spread at the site of infection by stimulating the formation of *Salmonella*-containing apoptotic bodies that are then engulfed by neighboring macrophages [49, 50]. Through these cell death mechanisms, *S. enterica* is able to spread between cells within an organism as well as propagate itself between organisms via fecal-oral transmission.

1.2.4 *S.* typhi Pathogenesis

In humans, *S.* typhi infection poses a much more serious health concern than does *S.* typhimurium, as it causes a systemic infection even in healthy individuals. This systemic infection, Typhoid fever, requires a 1-3 week incubation period, in which time the bacteria travels through the bloodstream, predominately targeting the liver, spleen, gall bladder, and bone marrow [19, 51]. The underlying mechanistic differences responsible for the dramatic contrast in *S.* typhimurium and *S.* typhi disease outcome are largely unknown. Both pathogens enter the human host orally and initially reside and replicate inside the SCV.

Unlike *S.* typhimurium, which stimulates a substantial neutrophil influx into the intestines and is accompanied by diarrhea, *S.* typhi infection does not lead to significant intestinal inflammation and only a third of cases develop diarrhea [4]. *S.* typhi expresses the Vi antigen, a capsular polysaccharide, on its surface during intestinal infection. The Vi antigen down-regulates the Toll-like receptor-mediated host response that induces neutrophil infiltration; thus, through Vi antigen expression, *S.* typhi is able to colonize deeper body tissues [52-54].

Another distinguishing feature of *S.* typhi is that it produces and secretes a typhoid toxin, which is similar to the cytolethal distending toxin (CDT) produced by many other Gram-negative bacterial pathogens [55]. Typhoid toxin, like CDT,

is an AB toxin that possesses DNase and ADP-ribosyl transferase activities and is capable of translocating to the host nucleus, making double-strand breaks in DNA and ultimately leading to cell cycle arrest and apoptosis in the host [55-57]. This toxin is expressed once *S.* typhi reaches an intracellular location, and it is then transported outside the host through a series of vesicular transport intermediates where it is able to induce apoptosis in infected and uninfected neighboring cells [56, 58, 59]. By causing apoptosis in the host tissue, CDT toxins enable bacterial colonization, persistent infection, and chronic disease [57]. *S.* typhi is able to persist in infected tissues and evade immune defenses by residing inside host cells that lack CDT receptors, and therefore, are not susceptible to typhoid toxin action [58]. This persistence allows *S.* typhi to infect humans for extended periods of time, and even for life, if the infection goes untreated.

1.3 Gifsy-2 Gene E Effector Protein

Another distinguishing factor between *S.* typhimurium and *S.* typhi is the presence of effector protease Gifsy-2 Gene E (GtgE) [56, 60]. GtgE, which is required for full virulence of *S.* typhimurium, is one of two virulent genes located on the Gifsy-2 bacteriophage and is secreted in a SPI-1 T3SS-dependent manner [56, 61]. GtgE proteolytically cleaves and inactivates its Rab GTPase substrates (Rab29, Rab32, Rab38), enabling *S.* typhimurium to evade the host's antimicrobial defenses as it carries out its lifecycle within the SCV. Expression of

GtgE in *S.* typhi allows the pathogen to overcome some of its host restriction barrier to infect mouse macrophages, which is likely due to the pathogen's ability to better elude host microbial defenses [60]. In wild-type *S.* typhi infection, the recruitment of Rab29 to the SCV aids in the formation of toxin transport intermediates, ultimately facilitating a systemic *S.* typhi infection [56].

As GtgE's biological role has only recently been discerned, little is understood as to how the protease performs its catalytic function. GtgE is a 26kD protein, 228 residues in length, with predicted C-terminal homology to known papain-like cysteine proteases as determined by HHpred [62]. This C-terminal homology does not span the entire active site of GtgE, making functional predictions about the enzyme fairly ambiguous. Through our structural and biochemical studies detailed in this work, we have determined that GtgE functions as a cysteine protease and have identified the catalytic residues necessary for GtgE's activity.

1.4 Rab GTPases

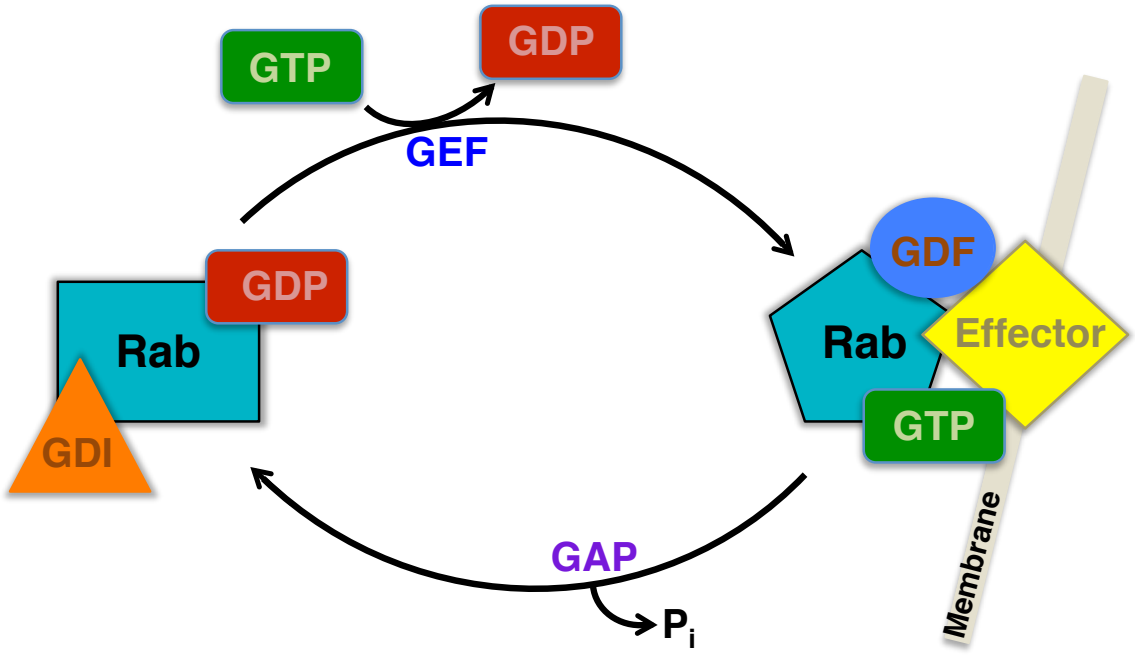


Figure 3. The activity cycle of Rab GTPases. Rab GTPases cycle between GDP- and GTP-bound states with the aid of GTPase-activating proteins (GAPs) and guanine nucleotide exchange factors (GEFs). The GTP-bound state primes the Rab for interaction with its specific, membrane-localized cellular target. Figure adapted from Stenmark, *et al.* and Ebine, *et al.* [63, 64].

In humans, the Rab GTPase family, which falls within the Ras superfamily of small G proteins, contains more than 60 members that are localized to distinct intracellular membranes based on the sequence variation and prenylation of their C-terminal domain [65]. Rab GTPases cycle between inactive and active states, depending on whether they are bound to GDP or GTP, to participate in the

regulation of membrane trafficking, cell growth, and differentiation (Fig. 3) [65, 66]. Although Rabs have intrinsic GTPase activity, they require the aid of several different interacting partners for adequate function [63]. The Rab activity cycle begins upon prenylation and consequent delivery of the Rab to its target membrane by accessory factors known as Rab escort proteins (REPs) [67]. Nucleotide exchange, GDP to GTP, and activation of Rab is facilitated by guanine nucleotide exchange factors (GEFs), which are unique to each Rab [63]. GTPase-activating proteins (GAPs) accelerate the hydrolysis of GTP by Rab, stimulating the given downstream effect [68, 69]. Following hydrolysis, the Rab is bound to GDP and the GAP is replaced by a GDP dissociation inhibitor (GDI), which displaces Rab from the membrane and sequesters it in the cytoplasm until the next transport cycle begins [67].

Rab GTPases are typically small proteins, roughly 20-25kD, and share an overall fold containing a six-stranded β sheet surrounded by five α helices [70]. The residues responsible for interacting with the guanine nucleotide and magnesium ion, an essential cofactor for nucleotide binding, reside in the loop regions connecting the β strands and α helices [64, 71]. The nucleotide-binding pocket on these Rab proteins is formed by two highly conserved regions termed Switch I and Switch II (Fig. 4A). These regions confer functional specificity to the Rabs and serve as recognition sites for effector proteins because they make dramatic conformational changes based on the identity of the bound nucleotide

[72, 73]. GtgE cleaves its Rab substrate between a glycine and valine located within the switch I region (Fig. 4B) and in doing so, presumably destroys the nucleotide-binding pocket and inactivates the GTPase.

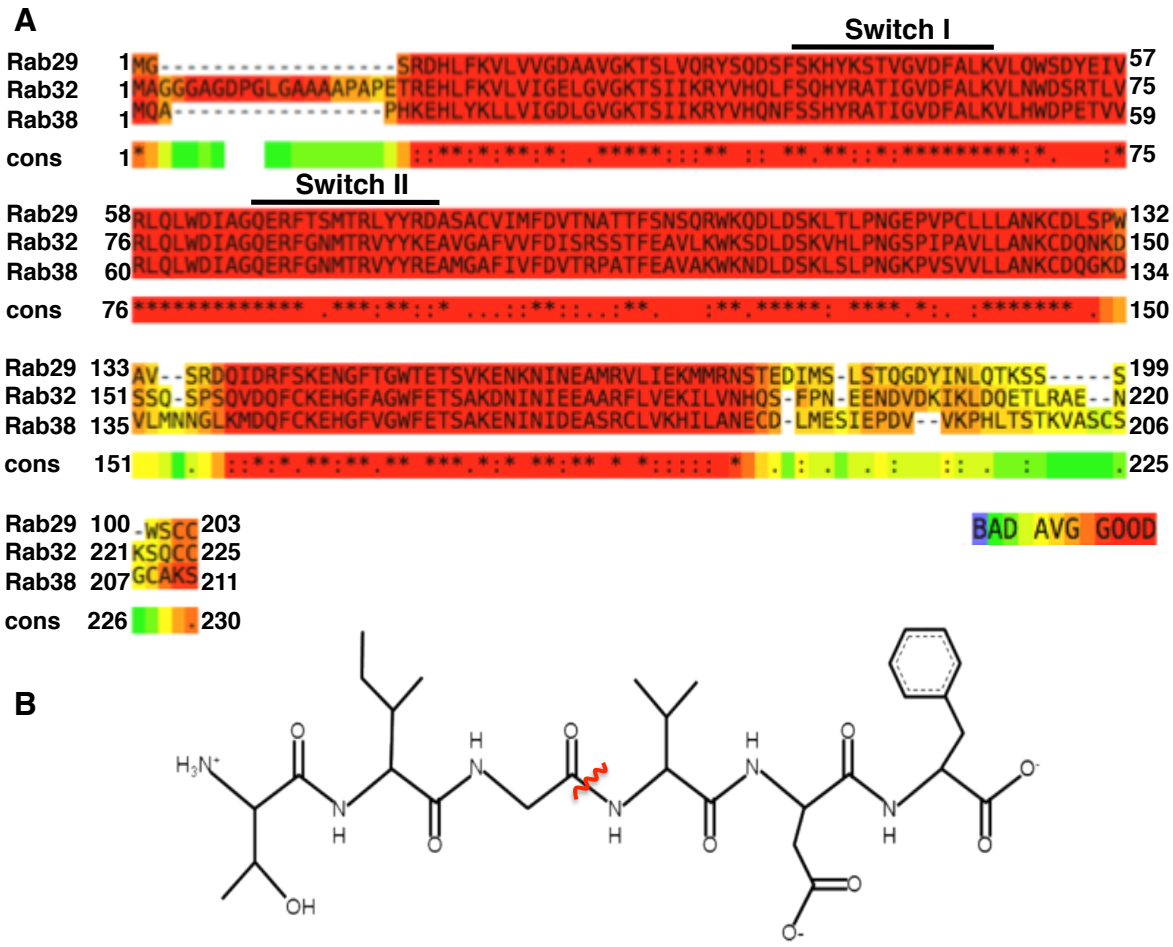


Figure 4. The Rab GTPase substrates. (A) The alignment of full-length Rab29, Rab32, and Rab38. Sequence homology is depicted on a color scale ranging from red (high homology) to blue (low homology). Key structural motifs, Switch I and Switch II, are indicated. Alignment generated in T-Coffee [74]. (B) The GtgE cleavage site of Rab38, TIGVDF, which corresponds to residues 41-46. GtgE cleaves between Gly43 and Val44.

GtgE's substrates, Rab29, Rab32, and Rab38, are closely related Rab GTPases that form their own subclass within the Rab GTPase family [60, 66]. As illustrated in Figure 4A, Rab29, Rab32, and Rab38 share high homology throughout much of their primary sequence, but differ in their C-terminal tail, which is expected because these GTPases have different cellular targets. Rab32 and Rab38 have been implicated in the biogenesis of lysosome-related organelles (i.e. melanosomes, specialized granules in platelet and T cells, and the *Salmonella*-containing vacuole) and coordinate the delivery of cargo to these organelles [60, 75-77]. In combination with BLOC-1, -2, -3, Rab32 and Rab38 allow for the delivery of antimicrobial proteins to the SCV, consequently leading to pathogen death [60, 76, 78]. Through cleavage of Rab32 and Rab38, GtgE is able to circumvent this particular host cell defense pathway. Rab29 remains uncharacterized thus far [66]. Rab29 is known to associate with the Golgi complex as well as with lengthy, dynamic tubules extending from the Golgi body, and it is has been shown that Rab29 is required for the formation of typhoid toxin transport intermediates [56].

CHAPTER 2. MATERIALS AND METHODS

2.1 Protein Expression and Purification

Salmonella typhimurium LT2 GtgE $^{1-228}$, GtgE$^{79-214}$, and all other GtgE constructs were cloned into a modified pCDF-Duet-1 vector (Novagen) between the *SalI* and *NotI* restriction sites. All GtgE$^{1-228}$ point mutants were generated by site-directed mutagenesis. GtgE$^{43-214,\Delta 74-79}$ was generated by PCR amplifying GtgE$^{43-73}$ and GtgE$^{80-214}$ and stitching the two pieces together via PCR amplification. Protein was expressed in *E. coli* BL21(DE3) cells with 0.75mM IPTG for 16 hours at 18°C. Selenomethionine-substituted GtgE$^{79-214}$ was expressed in *E. coli* 834 cells in selenomethionine-supplemented minimal media with 0.75mM IPTG at 18°C for 12 hours. Harvested cells were pelleted, resuspended in 50mM Tris-HCl, pH 8.0, 300mM NaCl, and 5mM imidazole, and lysed via high-pressure homogenization. Cleared lysate was run over Ni-NTA resin (Qiagen) and protein was eluted in 50mM Tris-HCl, pH 8.0, 300mM NaCl, and 500mM Imidazole. Protein was dialyzed against 25mM Tris-HCl, pH 8.0, 250mM NaCl, and 5mM DTT, and the N-terminal Histidine tag was cleaved off with 6xHis-rhinovirus 3C protease. The material was passed over Ni-NTA to remove the histidine tag and 3C protease. The final purification step was size exclusion chromatography using a Superdex 75, 120mL column (GE Healthcare). Native and selenomethionine-substituted GtgE$^{79-214}$ was purified into 25mM HEPES, pH 8.0, 250mM NaCl, and 5mM DTT. GtgE$^{1-228}$ constructs were purified

into 25mM Tris-HCl, pH 8.0, 250mM NaCl, and 5mM DTT. Purified protein was stable at -80°C for up to 3 months.

Human Rab38$^{1-211}$ was cloned into a modified pCDF-Duet-1 vector (Novagen) between the *SalI* and *NotI* restriction sites. Purification was carried out identically to the purification of GtgE; however, all Rab38$^{1-211}$ buffers contained 5mM $MgCl_2$ and final purification was done on a Superdex 200, 120mL column (GE Healthcare) into 25mM Tris-HCl, pH 8.0, 250mM NaCl, 5mM $MgCl_2$, 5mM DTT. Purified protein was stored at -80°C.

To form the complex of GtgE and Rab29/Rab38, GtgE and its Rab substrate were co-expressed from a modified pCDF-Duet-1 vector (Novagen). Rab29 or Rab38 was cloned between the *SalI* and *NotI* restriction sites and was N-terminally histidine-tagged. GtgE was cloned into the tag-free second cloning site between *NdeI* and *XhoI* restriction sites. Protein was expressed in *E. coli* BL21(DE3) cells with 0.75mM IPTG for 16 hours at 18°C. The complex was purified according to the purification conditions used for Rab38$^{1-211}$, and during the final Superdex 200, 120mL column (GE Healthcare) run, the material was exchanged into 25mM Tris-HCl, pH 8.0, 250mM NaCl, 5mM DTT, and 100µM $MgCl_2$. Combinations of different GtgE and Rab29/Rab38 point mutants were generated by site-directed mutagenesis and were expressed and purified

according to the protocol described for the wild-type complex. The point mutants used in this work are detailed in Table 4. Purified protein was stored at -80°C.

2.1.1 Urea Denaturation and Dialysis-based Refolding

Protein was expressed from *E. coli* BL21(DE3) cells and lysed as previously described. Lysate was cleared via centrifugation, and the resulting pellet (insoluble fraction) was resuspended in 6M urea, 0.5M NaCl, and 20mM Tris-HCl pH 8.0. Resuspended material was centrifuged at 16,000 rpm for 40 minutes, and the supernatant was applied to NiNTA resin (Qiagen) that had been equilibrated in 6M urea, 0.5M NaCl, and 20mM Tris-HCl pH 8.0. The column was washed with 6M urea, 0.5M NaCl, 30mM imidazole, and 20mM Tris-HCl pH 8.0. Protein was eluted with 6M urea, 0.5M NaCl, 500mM imidazole, and 20mM Tris-HCl pH 8.0. All work was done at room temperature.

Refolding was performed through four rounds of dialysis at 4°C in 20mM Tris-HCl pH 8.0, 250mM NaCl, 5% glycerol, 2mM EDTA, and 5mM DTT. The protein sample was diluted to approximately 0.1 mg/mL, and the protein to refolding buffer volume ratio did not exceed 1:20. The first three rounds of dialysis were 2-4 hours in length. For the final round, which was allowed to dialyze for 16 hours, glycerol and EDTA were removed from the buffer and the samples were treated with 6xHis-rhinovirus 3C protease to cleave the histidine tag. Purification then followed the protocol stated above for native proteins.

2.2 Limited Proteolysis

90µg $GtgE^{43-214}$ was treated with subtilisin protease (Sigma Aldrich) in a range of 0.55µg to 10.92µg with 5mM $CaCl_2$ for 20 minutes at 4°C. The reaction was terminated with the addition of 10mM PMSF and SDS loading buffer. The digest of the $Rab38^{V44I}/GtgE^{C45A}$ complex was run under the same conditions; however, 34.8µg of complex material was treated with subtilisin ranging from 0.00174µg to 0.261µg and the reaction allowed to proceed for 30 minutes at 4°C. Cleavage products from both reactions were separated by 15% SDS-PAGE, transferred to PVDF membrane (Millipore), and stained with SYPRO® Ruby protein stain (Sigma Aldrich). Major protein bands were cut from the membrane and sent to the Columbia University Protein Core Facility (New York, NY) for N-terminal Edman degradation sequencing.

2.3 Crystallization and Structure Determination

Native $GtgE^{79-214}$ was crystallized in a hanging drop format at a 10mg/mL drop concentration at 4°C in 0.2M Li_2SO_4, 1.75M $(NH_4)_2SO_4$, and 0.1M Tris, pH 7.0. Selenomethionine-substituted $GtgE^{79-214}$ was crystallized by two rounds of seeding with native crystals at a 10mg/mL drop concentration at 4°C in 0.2M Li_2SO_4, 1.75M $(NH_4)_2SO_4$, 0.1M Tris, pH 7.0. Crystals were cryo-protected in 0.3M $LiSO_4$, 2.5M $(NH_4)_2SO_4$, 0.1M Tris, pH 7.0 and flash-frozen in liquid nitrogen for data collection. X-ray data was collected at Brookhaven National Synchrotron

Light Source (NSLS) Beamline X29, and processed using HKL2000 [79]. Phasing and initial protein building of the selenomethione-substituted protein was done in Phenix using AutoSol [80-83]. Automated building produced a model with 133 residues, and the initial model was used subsequently with ARP/wARP to build a model with the higher resolution native data [84]. Refinement of this model was carried out with REFMAC5 [85, 86] and manual model building was done in COOT [87]. TLS refinement [88, 89] was used in the last stages of refinement to generate a model spanning residues 80-213 with an R/R_{free} of 19.31%/22.97% (Table 2). No electron density was observed for residues 79, 145, 146, 171, 193-199, and 214, so these residues were not modeled into the final structure. 98% of the residues fall into the most favored region of the Ramachandran plot with no outliers. Figures were generated using CCP4mg [90].

2.4 Gel-based Activity Assay

The activity assay was performed using Rab38 and GtgE material purified as detailed. GtgE and Rab38 were mixed in a 1:4 molar ratio in the presence of 10mM $CaCl_2$ and 10mM $MgCl_2$. End-point assays were performed at 4°C for 30 minutes and time-point assays were sampled at 4°C, as described in Figure 11C. Reactions were stopped with the addition of SDS running buffer and boiled for 10 minutes before visualization by 15% SDS-PAGE and Coomassie staining. Time point assays were run in triplicate and quantification of GtgE and Rab38 (cleaved

and uncleaved) band intensity was performed with ImageJ software [91]. These intensities were first normalized against the loading control, GtgE intensity. Then, the normalized Rab38 cleavage product intensities were expressed relative to the total Rab38 in the reaction, and these cleavage product percent values were summed to convey the product formation in terms of the total Rab38 in the reaction.

2.5 Inhibition Assay

N-ethylmaleimide, antipain, and chymostatin were solubilized in ethanol, water and DMSO, respectively. 5.75μM GtgE-WT was incubated at room temperature for 15 minutes with 10mM $MgCl_2$, 10mM $CaCl_2$, and one of the following additives: water (positive control), 2.5% ethanol/25% DMSO (delivery controls), or an inhibitor (0.25 mM N-ethylmaleimide (Sigma Aldrich), 5mM antipain (Sigma Aldrich), and 2.5mM chymostatin (Santa Cruz Biotechnology)). 5.75μM GtgE-C45A was also incubated at room temperature with water as a negative control. For the leupeptin assay, leupeptin (Sigma Aldrich) was reconstituted in water and used in the following concentrations: 10μM, 50μM, 100μM, 500μM, 1mM, 2mM, and 5mM. 25.3μM Rab38 was then added and the reaction was allowed to proceed on ice for 30 minutes. The reaction was ended by the addition of SDS running buffer and boiling. Results were visualized via 15% SDS-PAGE and Coomassie staining.

CHAPTER 3. RESULTS AND DISCUSSION

3.1 Domain Determination

Since GtgE lacks significant sequence similarity to proteins of known structure, the structural determination of GtgE was necessary in order to gain a conclusive understanding of the enzyme's active site. Our initial crystallization attempts focused on full-length GtgE; however, the full-length construct was not amenable to crystallization, so further construct refinement was done based on secondary structure predictions and limited proteolysis using subtilisin. A summary of the constructs tested for crystallization can be found in Table 1. Constructs with C-terminal truncations past residue 205 were insoluble, and were purified under denaturing conditions and refolded using a dialysis-based refolding technique (described in Materials and Methods). Solubility was not affected by the N-terminal truncations produced in this work; however, it was noted that GtgE lost activity once the N-terminal 28 residues had been removed (Fig. 5).

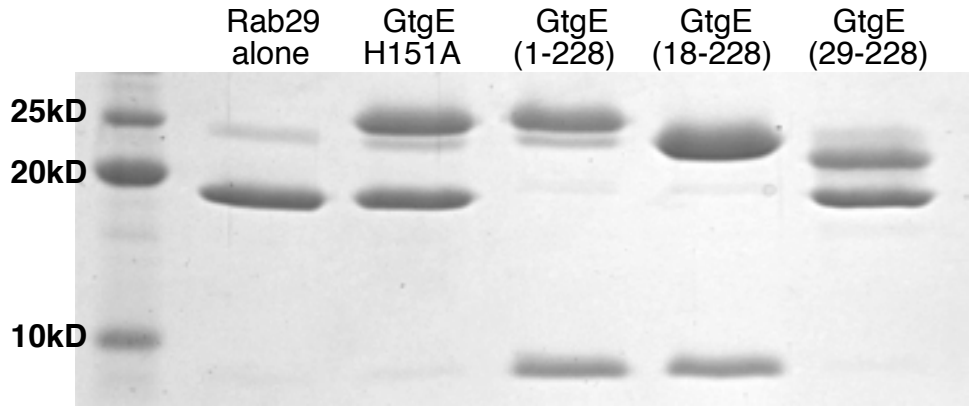


Figure 5. The effect of N-terminal truncations on GtgE activity. $Rab29^{1-177}$ is cleaved by full-length GtgE as well as by $GtgE^{18-228}$, but is not cleaved by $GtgE^{29-228}$. $Rab29^{1-177}$ that has not been treated with GtgE is in the lane to the far left. The catalytically dead mutant of GtgE, $GtgE^{H151A}$, is unable to cleave the Rab substrate.

Table 1. GtgE crystallization constructs. Yield is given in milligrams of protein obtained per liter of cell culture grown.

Residue	pI	Soluble	Yield (mg/L)	Active
1-228	4.50	Y	5.3	Y
18-228	4.31	Y	73.4	Y
29-228	4.12	Y	15.6	N
35-228	4.12	Y	14.8	
39-228	4.17	Y	12.0	
43-228	4.17	Y	15.0	
18-226	4.31	Y	7.7	
18-220	4.29	Y	65.7	Y
18-209	4.19	Y	18.6	N
18-195	4.25	N	0.9	
18-180	4.26	N	0.9	
18-175	4.30	N	1.7	
18-170	4.30	N	0.9	
29-195	4.03	N	5.2	
39-200	4.08	Y	1.0	
35-209	4.00	Y	6.3	
43-214	4.08	Y	13.1	
43-209	4.04	Y	3.0	
43-200	4.08	Y	1.4	
79-209	4.05	Y	1.6	
79-214	4.11	Y	9.1	N

Several rounds of limited proteolysis using subtilisin were performed as detailed until a minimal, crystallizable construct was determined. Subtilisin digestion of GtgE$^{43-228}$ led to the design of a minimal construct of approximately 17kD, spanning residues 79-214, which successfully crystallized (Fig. 6). A 15kD construct, roughly residues 79-200, was also identified; however, since GtgE is insoluble with C-terminal truncations past residue 205, this construct was not pursued. Edman sequencing of the bands showed that both products began at residue 79, and the C-terminal truncations were predicted based on the products' molecular weight.

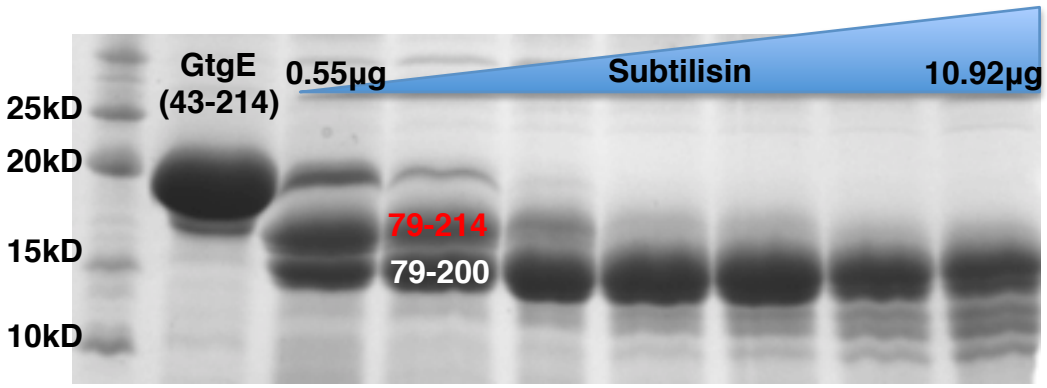


Figure 6. Subtilisin digestion of GtgE$^{43-214}$. Subtilisin concentration shown increasing from 0.55µg to 10.92µg from left to right. Undigested GtgE$^{43-214}$ is shown in the lane at the far left. The digestion produced two cleavage products: GtgE$^{79-214}$ and GtgE$^{79-200}$. The crystallized construct is shown in red.

3.2 Crystallization and Data Collection

GtgE$^{79-214}$ was expressed and purified as described. When purified in 25mM Tris-HCl, pH 8.0, 250mM NaCl, and 5mM DTT, and at concentrations above 30mg/mL, GtgE$^{79-214}$ exists in both a monomeric and dimeric state. Purification conditions that yielded stable, monomeric GtgE$^{79-214}$ were determined (25mM HEPES, pH 8.0, 250mM NaCl, 5mM DTT; 20mg/mL) and used for the purification of the crystallized protein product (Fig. 7A).

Ideal crystallization conditions were identified in a 96-well screen using a sitting drop format, and crystallization was scaled up into a 15-well hanging drop format to obtain well-diffracting crystals for structural determination. Drops contained a 1:1 ratio of reservoir buffer to protein, which was at a 20mg/mL concentration. Crystals grew at 4°C in 0.2M Li_2SO_4, 1.75M $(NH_4)_2SO_4$, and 0.1M Tris pH 7.0, appearing after 3 days and reaching full size in 1-2 weeks (Fig. 7B). The resulting crystals had a bi-pyramidal morphology and were mildly birefringent. Crystals were cryoprotected with 0.3M $LiSO_4$, 2.5M $(NH_4)_2SO_4$, 0.1M Tris, pH 7.0 and frozen in liquid nitrogen for diffraction data collection.

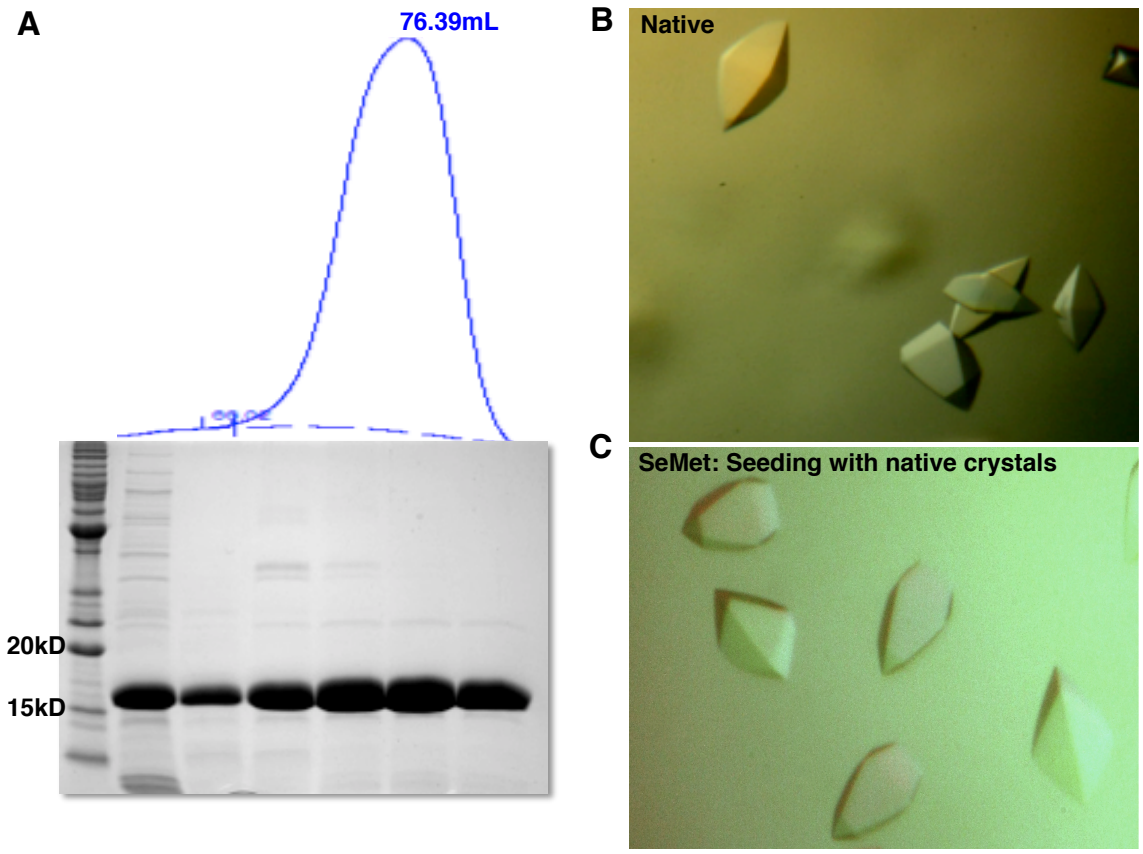


Figure 7. The purification and crystallization of GtgE$^{79-214}$. (A) Purification of GtgE$^{79-214}$ on a Superdex75, 120mL column. The construct behaves as a monomer under the detailed conditions, eluting at 76.39mL. Lane 1 (far left) contains the molecular weight marker, lane 2 is the column load material, and the remaining lanes are the fractions corresponding to the elution peak. (B) Crystals of native GtgE$^{79-214}$ grown at 4°C in 0.2M Li_2SO_4, 1.75M $(NH_4)_2SO_4$, 0.1M Tris pH 7.0 with an initial set-up protein concentration of 20mg/mL. (C) Crystals of selenomethionine-substituted GtgE$^{79-214}$ were grown under the same conditions used for native crystal growth.

GtgE's amino acid sequence does exhibit significant homology to any proteins of known structure; therefore, phase determination was carried out experimentally via single-wavelength anomalous diffraction (SAD) using selenomethionine. Selenomethionine contains selenium, a heavy atom, in place of the sulfur normally present in methionine, and can usually be incorporated into a given protein with very little effect on the overall protein structure. Heavy atoms, such as selenium, are able to absorb X-rays and reemit the absorbed radiation with an altered phase, producing a measurable difference in intensity between two Friedel pair reflections. This scattering effect, known as anomalous scattering, is most prominent when the X-ray wavelength is near the absorption edge of the heavy atom, which for selenium is 0.980Å. SAD enables the location of heavy atoms, and when combined with density modification, allows for the determination of phases for the entire structure.

For measurable anomalous scattering and subsequent successful phase determination, roughly 1 selenium atom per 100 amino acids is required in the crystallized protein [92, 93]. $GtgE^{79-214}$ is 135 residues in length and contains three methionines, which equates to 1 selenium atom for every 45 residues, making GtgE an ideal candidate for phase determination with selenium as the anomalous scatterer. Selenomethionine-substituted $GtgE^{79-214}$ was produced as detailed. Crystals of selenomethionine-substituted $GtgE^{79-214}$ were grown under the same conditions as the native crystals through successive rounds of seeding,

starting with native crystals. These crystals had the same morphology as the native crystals, but were smaller in size, only growing to roughly two-thirds the size of the native crystals (Fig. 7C). The selenomethionine-substituted crystals were cryoprotected using the same cryoprotectant as was used for the native crystals and were also flash frozen in liquid nitrogen prior to data collection.

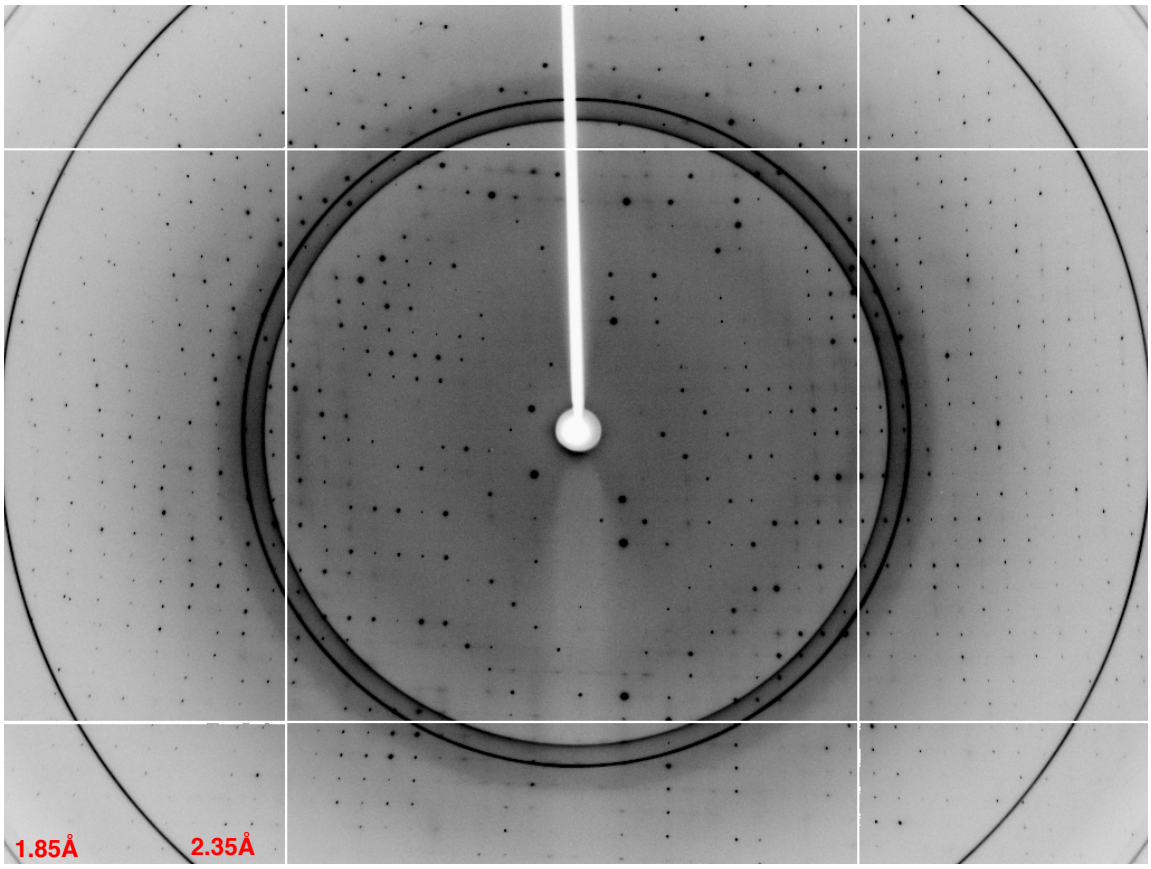


Figure 8. Diffraction pattern from native GtgE$^{79-214}$ crystals. Diffraction data was collected at NSLS Beamline X29. Native crystals diffracted to 1.65Å and selenomethionine-substituted crystals diffracted to 2.38Å.

Table 2. Crystallographic Statistics for GtgE$^{80-213}$.

	Native	Selenomethionine-substituted
Data Collection		
Space group	$P4_1\,2_1\,2$	$P4_1\,2_1\,2$
Cell Dimensions		
a, b, c (Å)	56.181, 56.181, 125.124	56.126, 56.126, 125.033
α, β, γ (°)	90, 90, 90	90, 90, 90
Wavelength (Å)	1.0750	0.9790
Resolution (Å)	50.00 - 1.65 (1.71 - 1.65)	50.00 – 2.38 (2.47 – 2.38)
No. of reflections	420901	425153
No. of unique reflections	23155 (2428)	15090 (1550)
R-merge†	0.072 (0.894)	0.162 (0.705)
Mean I/σ(I)	30.13 (3.2)	16.45 (4.67)
Completeness (%)	92.6 (100.0)	99.0 (100.0)
Redundancy	11.2 (11.7)	15.3 (15.5)
Refinement		
Resolution (Å)	19.63 - 1.65 (1.693 - 1.65)	
No. of reflections	21891 (1778)	
R-factor‡	0.1931 (0.258)	
R-free‡	0.2297 (0.250)	
Number of atoms	1129	
macromolecules	957	
ligands	5	
water	167	
Amino acid residues	124	
RMS bonds (Å)	0.017	
RMS angles (°)	1.75	
Average B-factor	30.40	
Ramachandran favored (%)	98	
Ramachandran outliers (%)	0	

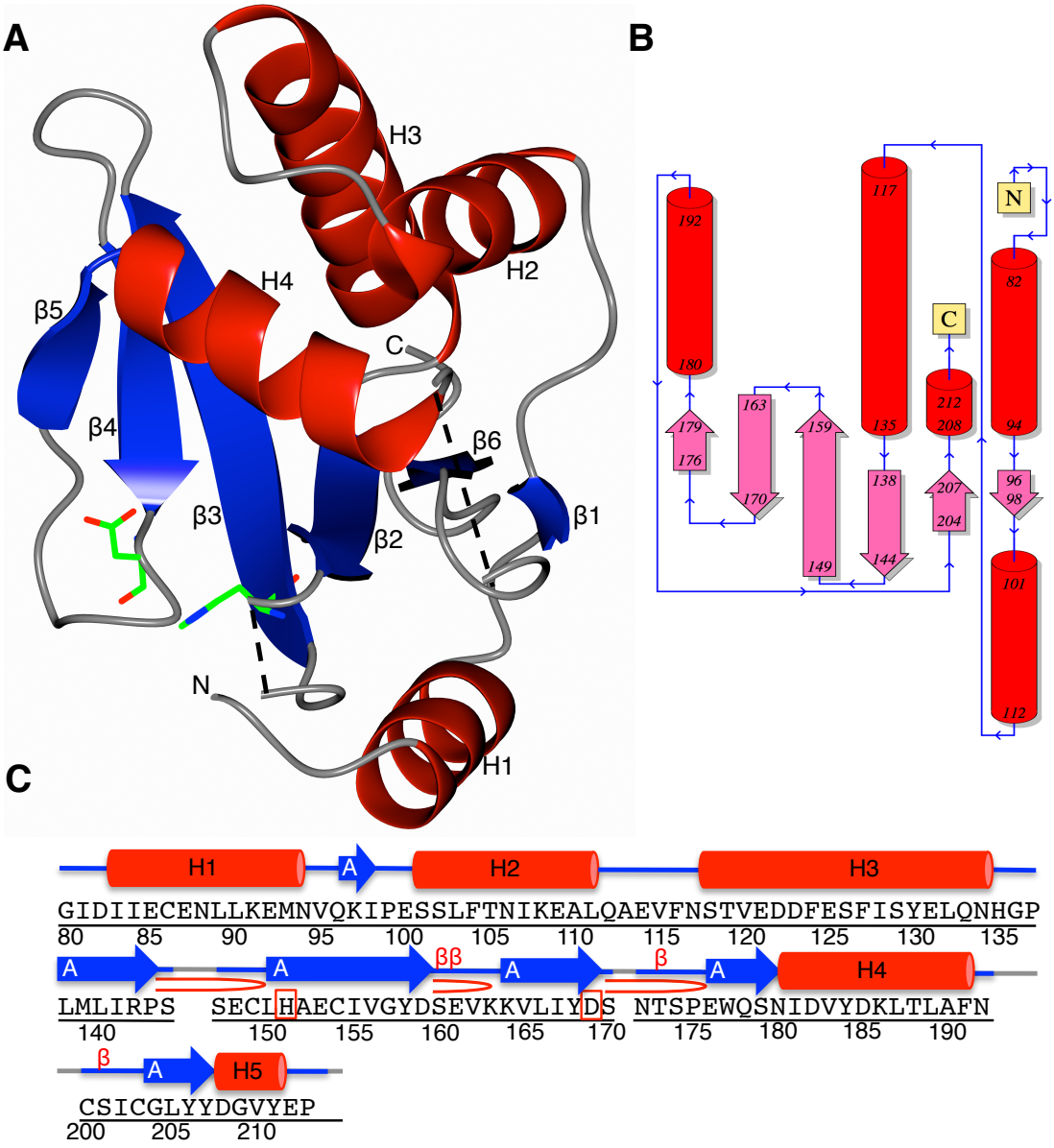
Statistics for the highest-resolution shell are shown in parentheses.
R-free value test set size: 5%
† As defined and calculated by HKL2000
‡ As defined and calculated by Refmac5

3.3 Overall Structure

The structure of GtgE, residues 80-213, was solved to 1.65Å. Diffraction data were collected at Beamline X29 at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory (BNL) from selenomethionine-substituted and native protein crystals (Fig. 8). Both native and selenomethionine crystals belonged to the $P4_12_12$ space group, diffracting to 2.38Å and 1.65Å, respectively. The model was built automatically into the SAD data with Phenix AutoSol [80, 81], and the initial model was used subsequently with ARP/wARP [84] to build a model with the native data (Table 2).

The final structure of $GtgE^{80-213}$ is comprised of a six-stranded beta-sheet that is sandwiched between three helices on one side of the sheet and one helix on the other (Fig. 9). The beta strands are arranged in an antiparallel beta-sheet adopting the topology depicted in Figure 9B. There are three beta hairpins connecting beta strand 2 (β2) to β3, β3 to β4, and β4 to β5, and a beta-turn located between helix 4 and β6. The structure contains 167 water molecules and 124 amino acids, spanning residues 80-213, with an R/R_{free} = 0.1931/0.2297 (Table 2). No electron density was observed for residues 79, 145, 146, 171, 193-199, and 214, so these residues were not modeled into the final structure.

Figure 9. The structure of GtgE$^{80-213}$. (A) The overall fold of GtgE, residues 80-213. Helices shown in red, beta strands shown in blue, and areas without observable electron density represented with a dashed line. His151 and Asp169 are depicted as sticks in green. (B) Topology diagram of GtgE, residues 80-213. Generated by *PDBSum* [94]. (C) The secondary structure sequence alignment for GtgE, residues 80-213. There are five helices, numbered and shown in red; six beta strands depicted in blue and labeled by their sheets; and active residues, His 151 and Asp169, are indicated with a red box. β-turns are represented by β, beta hairpins by ⊃ . Originally generated in *PDBSum* [94].


A
H3
H2
H4
β5
C
β4
β6
β3
β2
β1
N
H1
B
192
180
179
176
163
170
159
149
117
135
138
144
212
208
207
204
N
82
94
96
98
C
101
112
C
H1
A
H2
H3
GIDIIECENLLKEMNVQKIPESSLFTNIKEALQAEVFNSTVEDDFESFISYELQNHGP
80 85 90 95 100 105 110 115 120 125 130 135
A
A
ββ
A
β
A
H4
LMLIRPS SECIHAECIVGYDSEVKKVLIYDS NTSPEWQSNIDVYDKLTLAFN
140 150 155 160 165 170 175 180 185 190
β
A
H5
CSICGLYYDGVYEP
200 205 210

3.3.1 Structural Homologs

Structural homologs were identified using the DaliLite v. 3 [95]. Z-scores above 2.0 are considered significant, usually corresponding to the presence of similar folds [96]. There were 740 hits, largely cysteine proteases, with z-scores above 2.0, and of those hits, 23 had a Z-score above 5.0. The top 20 structure hits are summarized in Table 3. Our structure reveals significant similarity to cysteine proteases of Clan CA, namely families C1, C2, and C39 [95]. Clans denote evidence of a common ancestry, thus proteins within a clan share a common structural fold. Clan CA proteases share a papain-like catalytic fold comprised of 6 beta strands sandwiched between a variable number of alpha helices, and function using a catalytic triad. Two key active residues are a cysteine positioned at the beginning of an alpha helix and a histidine located at the beginning of beta strand 3. The third triad member (Asn/Glu for C1; Asn for C2; Asp/Glu for C39) is located at the end of beta strand 4, and is responsible for the proper orientation of the catalytic histidine [97, 98]. The catalytic core also contains a glutamine residue that resides a short distance N-terminal to the active cysteine, aiding in the stabilization of the reaction intermediate [97].

Proteins are further categorized into families based on their function and/or catalytic residue organization. In eukaryotes, members of the C1 family primarily serve as proteolytic enzymes in the lysosomal pathway [98]. C2 family members include the calcium-dependent calpains, which function in a variety of

cellular processes, such as signal transduction, apoptosis, and cytoskeletal remodeling [98]. The C39 endopeptidases are bacterial proteins that are responsible for the maturation of bacteriocin, a secreted bacterial antibiotic protein that inhibits the growth of other bacterial species [99].

Table 3. Structural homologs of GtgE$^{80-213}$. Homologs identified by the DaliLite v. 3 Homology Server [95]. Z-scores above 2.0 are considered significant. There were 740 structure hits with a Z-score of 2.0 or higher.

PDB ID	Z-Score	RMSD	No. of Residues	Length of Alignment	Sequence Identity (%)	Description
3zua-A	6.3	2.5	142	83	12	Alpha-hemolysin translocation ATP-binding protein
1df0-A	6.1	3.3	624	85	14	M-Calpain
1u5i-A	6.0	3.2	625	85	14	Calpain 2, large subunit precursor
3b79-A	5.9	2.4	125	77	8	Toxin secretion ATP-binding protein
3k8u-A	5.5	2.7	137	81	10	Putative ABC transporter, ATP-binding protein
4d8e-A	5.4	3.7	254	89	15	Streptopain
2bu3-B	5.3	3.9	204	97	15	ALR0975 protein
1mdw-A	5.3	3.2	319	82	15	Calpain II, catyalytic subunit
1pvj-C	5.3	3.2	339	88	15	Pyrogenic exotoxin B
1pvj-B	5.3	3.2	339	88	15	Pyrogenic exotoxin B
1pvj-D	5.3	3.2	339	88	15	Pyrogenic exotoxin B
2bu3-A	5.2	3.9	200	96	16	ALR0975 protein
1cv8-A	5.2	3.3	173	84	11	Staphopain
3bba-A	5.2	3.3	246	88	15	Interpain A
2uzj-B	5.2	3.6	249	88	15	Streptopain
3bba-B	5.1	3.3	250	88	15	Interpain A
2uzj-B	5.1	3.8	247	89	15	Streptopain
2ftd-A	5.1	3.4	215	80	16	Cathepsin K
2nga-B	5.0	3.5	318	84	14	Calpain 8
2btw-B	5.0	3.9	210	95	16	ALR0975 protein

3.4 Identification of the Catalytic Triad

GtgE's overall fold and placement of its active histidine (His151) align best to homologs of the cysteine protease superfamily Clan CA (Fig. 10). His151 was identified as an active residue by Spanò, *et al.* and confirmed through mutagenesis, His151Ala, and homolog structural alignment in this work (Fig. 10, Fig. 11A).

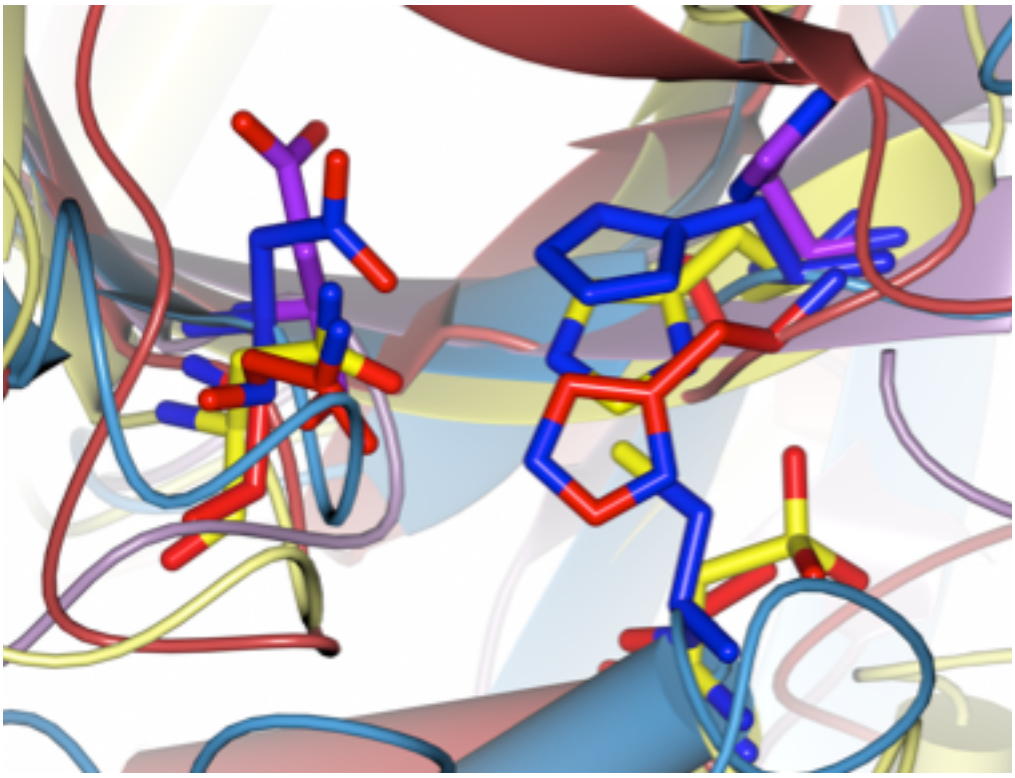


Figure 10. The active site alignment of GtgE$^{80\text{-}213}$ with its structural homologs. GtgE$^{80\text{-}213}$ (purple) with active residues His151 and Asp169; C39 family member, ComA (PDB: 3K8U, blue) with active residues Cys17, His96, Asp112 shown; C2 family member Calpain II (PDB: 1MDW, red) with active residues Cys105Ser, His262, Asn286; and C1 family member papain (PDB: 9PAP, yellow) with active residues Cys(OCS)25, His159, and Asn175.

3.4.1 Active Cysteine

Our crystal structure did not contain a cysteine in the canonical position, nor did it contain any cysteines in proximity to the other active site residues. In addition, the cysteine containing helix present in members of this superfamily was not present in the active site. Since the crystal structure lacks the first 78 residues of GtgE, we hypothesized that the active site cysteine must be located in the missing N-terminal domain. To confirm this hypothesis, we made point mutations (Cys to Ala) of each of the eight cysteine residues in GtgE and tested the enzyme's ability to cleave Rab38. Cys45 was the only Cys-to-Ala mutation that resulted in a loss of activity, and therefore, is the active cysteine of GtgE's catalytic triad (Fig. 11B). Although Cys45 is missing from our structure, we predict, based on the conserved homolog structure, that it resides at the beginning of a helix formed by the N-terminal residues not included in the crystallized construct.

3.4.2 Active Aspartic Acid

Based on our structure of $GtgE^{80-213}$ we were able to narrow down the possible identity of the third triad member to two residues: aspartic acid 169 and asparagine 172. Both residues are situated near the end of beta strand 4; however, no electron density was observed for the side chain of Asn172, so it could not be determined, from the structure alone, which residue was correctly

placed to participate in the active site. Through point mutation, Asp169Ala and Asn172Ala, Asp169 was identified as the third triad member, capable of significantly decreasing GtgE's activity; whereas, Asn172Ala does not appear to have any effect on enzyme activity (Fig. 11A).

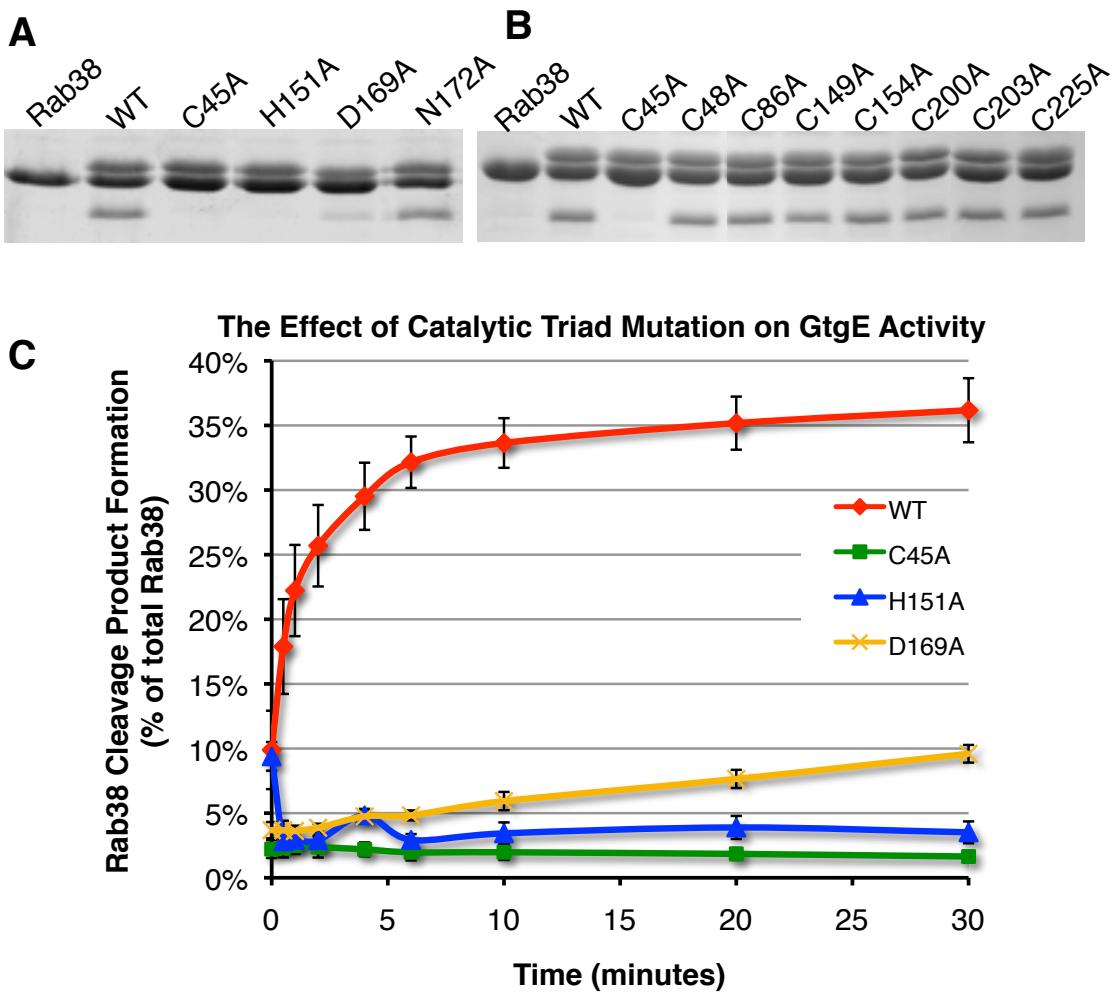


Figure 11. The catalytic triad of GtgE. (A) Mutations (C45A, H151A, D169A) to the catalytic triad of GtgE greatly reduce enzyme activity *in vitro*. N172A did not noticeably affect activity. (B) The Cys-to-Ala mutation activity profile for GtgE. Of all eight cysteines, only C45A shows a loss of function, indicating that it is the active cysteine of GtgE. (C) Catalytic triad residues (Cys45, His151, and Asp169) were mutated to alanines and their cleavage of Rab38 was charted as percent of total Rab38 cleaved over time, in minutes. Standard error of the mean indicated with black bars for each time point.

3.4.3 Quantitative Analysis of Catalytic Triad Mutations

To further investigate the effects of alanine mutations to the catalytic triad members, GtgE's cleavage of Rab38 was plotted in terms of the percent of total Rab38 cleaved over time (Fig. 11C). Under the reaction conditions described, wild-type GtgE cleaved 36% of Rab38 over the course of 30 minutes with the reaction reaching completion after 10 minutes of activity. Alanine mutations to Cys45, His151, and Asp169 resulted in the reduction of GtgE activity by 94%, 89%, and 72%, respectively, indicating that these residues are key components of GtgE's active site.

3.4.4 Proposed Catalytic Mechanism

Through this work, we have determined that GtgE is a cysteine endopeptidase, functioning with a catalytic triad composed of Cys45, His151, and Asp169. Together, these residues are capable of breaking the peptide bond between two highly conserved glycine and valine residues located in the Switch I region of Rab29, Rab32, and Rab38. A proposed mechanism for this reaction is diagramed in Figure 12.

The reaction is initiated once the Rab substrate enters the active site, creating an environment in which the nitrogen (ND1) of His151 depronates the thiol of Cys45. This depronation leads to the nucleophilic attack of the Cys45

sulfur on the carbonyl carbon of the glycine. A thioester linkage between the sulfur and the N-terminal section of the Rab substrate is then formed, and the nitrogen of the valine depronates the pronated His151 nitrogen (ND1), releasing the C-terminal Rab cleavage product and returning His151 to its initial state. A water molecule hydrolyzes the thioester linkage, which liberates the N-terminal Rab cleavage product and regenerates Cys45. Asp169 is not directly involved in the proton transfer necessary for catalysis; however, it is required for the proper positioning of His151.

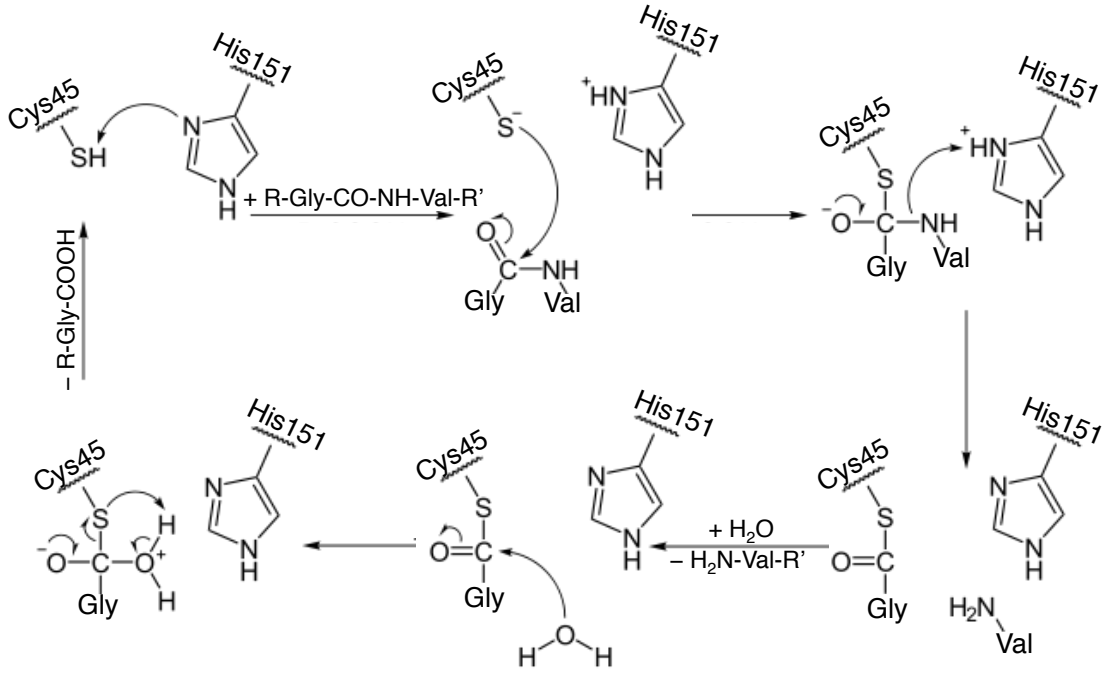


Figure 12. The proposed reaction mechanism for GtgE. GtgE functions as a cysteine protease to cleave its Rab substrate (Rab29, Rab32, and Rab38) between a critical glycine and valine in the Switch I region of these GTPases. GtgE uses a catalytic triad of Cys45, His151, and Asp169 to perform this cleavage reaction.

3.4.5 Complete Active Site Construct Design

Since our crystallized construct does not contain the active cysteine, we sought to crystallize a larger construct that would encompass the entire active site. Previous crystallization screens on larger domains of GtgE proved unsuccessful; therefore, we engineered GtgE$^{43-214}$, the largest minimal domain containing all three active residues, based on information obtained from our structure of GtgE$^{80-213}$ and its structural homologs.

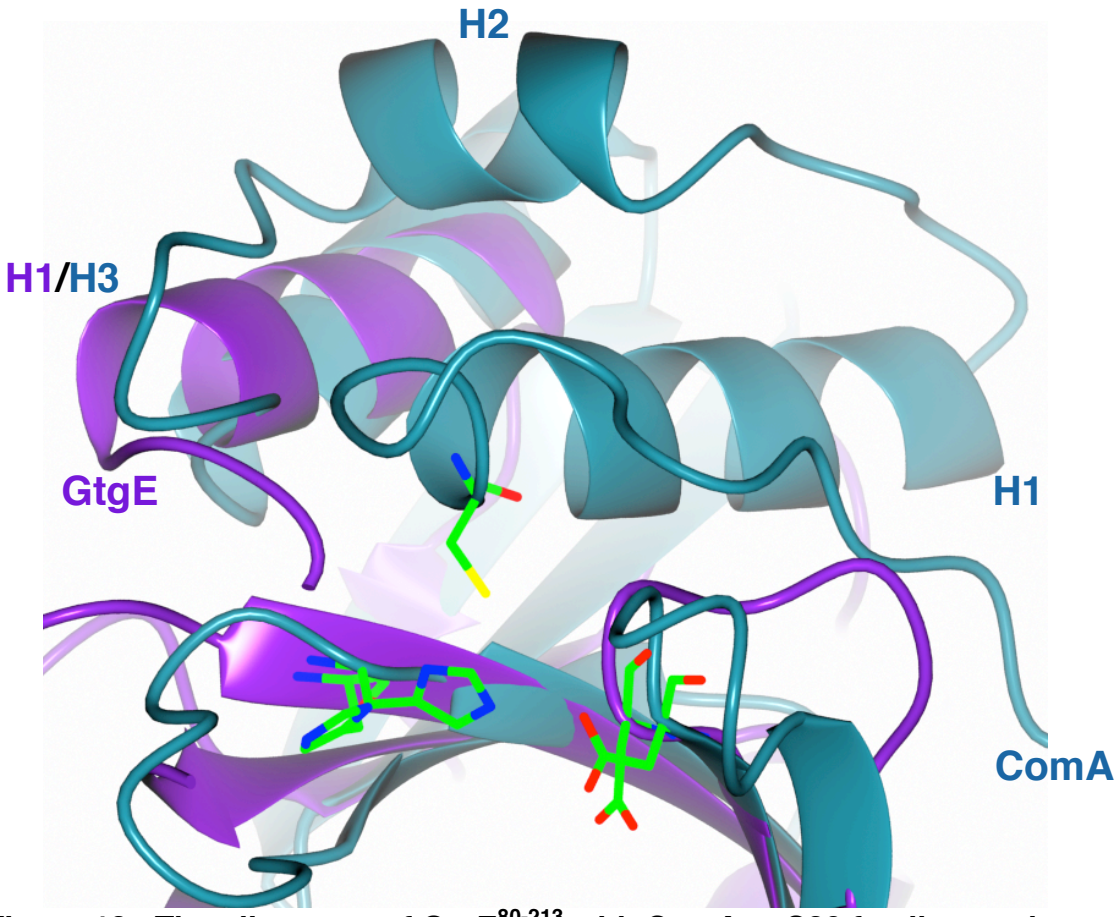


Figure 13. The alignment of GtgE$^{80-213}$ with ComA, a C39 family member. ComA (PBD: 3K8U, cyan) and GtgE$^{80-213}$ (purple) with helices labeled (H1, H2, etc.) for each and color-coded accordingly. Active residues for both structures are illustrated in green.

There were two disordered loops in the GtgE$^{80-213}$ structure, so to create a construct with a higher crystallization potential, we mutated residues in each loop to decrease the surface entropy of the protein and encourage protein-protein interactions. In disordered loop 1, spanning residues 145 to 147, Leu145, Ser147, and Glu148 were mutated to alanine. For disordered loop 2, which spans residues 193 to 199, Lys194, Lys196, and Glu198 were mutated to alanine. GtgE$^{43-214}$ constructs with mutations to loop 1 or loop 2 were screened for crystallization separately as well as together, in a GtgE$^{43-214}$ construct that contained mutations in both loop 1 and loop 2.

Through comparison and modeling work between GtgE$^{80-213}$ and its C39 family structural homologs, it was evident that GtgE contains a larger loop region located directly N-terminal to the start of GtgE$^{80-213}$. In the homologs, this loop connects an alpha helix to the helix corresponding to residues 83-93 (H1) in GtgE$^{80-213}$ (Fig. 13). We believe GtgE has a six-residue insertion in this loop, roughly equivalent to residues 74-79. As further evidence that this region is loopy and surface exposed, subtilisin digests of GtgE$^{43-214}$ led to cleavage immediately N-terminal to residue 79 (Fig. 6). A GtgE construct was designed with a deletion of residues 74 to 79, and this was tested for crystallization both alone and in combination with the loop 1 and loop 2 mutations. These crystallization screens did not yield crystals.

3.5 GtgE Inhibition

The inhibition of GtgE may serve as a means to alter the host specificity of broad-host *Salmonellae*. Equipped with an understanding of GtgE's function at the mechanistic level, we were able to identify three small molecules from a panel of cysteine protease inhibitors that are capable of inhibiting GtgE function *in vitro*: N-ethylmaleimide, antipain, and chymostatin (Fig. 14A-C). N-ethylmaleimide (NEM) covalently modifies cysteine residues, making it the least specific, although most potent, of the three inhibitors, showing inhibition at a 1:43 molar ratio of GtgE to NEM (Fig. 14B).

Antipain and chymostatin are both microbial-derived small peptide inhibitors which contain a C-terminal aldehyde that inhibits cysteine proteases by forming a hemiacetal adduct with the active thiol [100]. Antipain is bulky in structure (Arg-Val-Arg-Phe), and broadly targets cysteine and serine proteases (Fig. 14A). Chymostatin is composed of two phenylalanines, capreomycidine, a large residue unique to microbes, and a variable hydrophobic residue (Fig. 14A). Exhibiting greater specificity than antipain, chymostatin inhibits papain, chymotrypsin, and cathepsins A/B/C/H/L. Chymostatin was two-fold more potent of an inhibitor for GtgE than was antipain, inhibiting GtgE at a molar ratio of 1:435 GtgE to chymostatin, compared to a molar ratio of 1:870 GtgE to antipain (Fig. 14B). Conversely, leupeptin, another microbial-derived inhibitor comprised of

residues with relatively small side chains (Leu-Leu-Arg) (Fig. 14A), does not inhibit GtgE at the highest tested molar ratio (1:870 GtgE to leupeptin) (Fig. 14C).

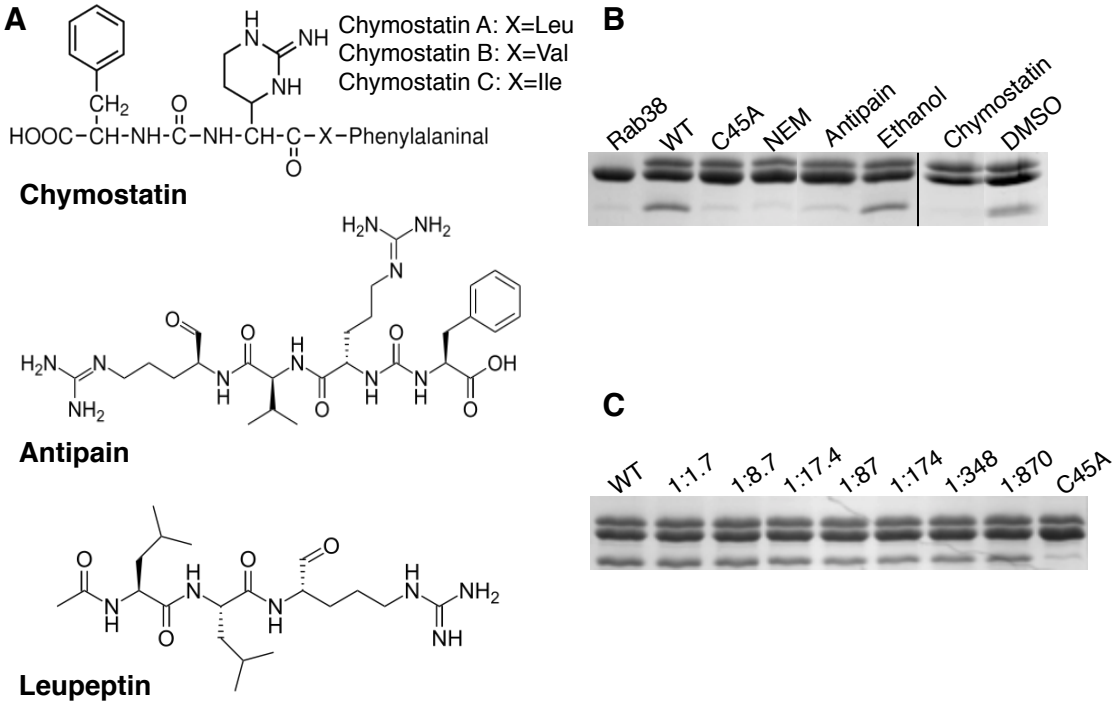


Figure 14. Inhibition of GtgE. (A) The structures of cysteine protease microbial-derived small peptide inhibitors. (B) GtgE is inhibited by NEM, chymostatin, and antipain. The reactions with chymostatin and its delivery control, DMSO, were performed separately using identical conditions to that of the reactions with NEM and antipain; thus, this is indicated with a black line. (C) Leupeptin does not inhibit GtgE at the highest concentration that could be tested given leupeptin's solubility. At a molar ratio of 1:870, GtgE to leupeptin, leupeptin did not inhibit GtgE's ability to cleave Rab38. The activity of GtgE-C45A is shown as a control.

The GtgE cleavage site on its Rab29/32/38 substrates is highly conserved among these three Rab GTPases, consisting predominantly of hydrophobic residues and several residues with bulkier side chains, such as aspartic acid and phenylalanine. We theorize that chymostatin and antipain are more suited to inhibit GtgE because they both contain aromatic residues, whereas leupeptin does not. Therefore, these inhibitors are better able to mimic the Rab29/32/38 cleavage site, and are best suited to provide the foundation for more targeted inhibitor design.

3.6 The GtgE/Rab GTPase Complex

To investigate the nature of GtgE's interaction with its Rab GTPase substrates, we determined the necessary conditions for stable complex formation *in vitro* for the purposes of crystallization and structure determination. We initially examined complex formation, as verified by an elution peak shift on the size exclusion Superdex 200, 120mL column (GE Healthcare), between the catalytically inactive $GtgE^{C45A}$ and Rab29, 32, and 38. Complex stability is defined here as the ability of the complex to remain in solution at 4°C for at least several days post-purification, and for the appearance of lower molecular weight Rab cleavage bands to be minimal.

Rab29 was found to be the least stable *in vitro*; however, it was found that Rab29 could be expressed in larger quantities when expressed with a C-terminal

truncation of the last 26 residues, thereby making it possible to perform subsequent binding assays using this substrate. Rab38 was determined to be the most ideal of the three candidates in forming a stable complex with $GtgE^{C45A}$. Unlike Rab29 and Rab32, Rab38 did not precipitate during purification and was even stable for up to 18 hours at room temperature. A cleavage band at approximately 14kD did appear, as it did during the purification of the other Rabs, and could be due to cleavage by a contaminating protease or by residual activity of $GtgE^{C45A}$ itself. The addition of a protease inhibitor cocktail did not prevent this cleavage.

Various mutations in both GtgE and Rab29/Rab38 were screened in an effort to identify a combination of mutations that would result in a stable complex with minimal residual Rab38 cleavage. These mutations, their ability to form a complex, and the long-term stability of the complex are summarized in Table 4. The GtgE mutations were initially tested in complex with $Rab29^{1-177}$, and the Rab29 and Rab38 mutations were complexed with $GtgE^{C45A}$.

Table 4. Identifying complex stabilizing mutations in GtgE and Rab29/38.

Mutation	Complex Formation	Long-term Stability
GtgE(C45A)	Y	Y
GtgE(C45A/C48A)	Y	N
GtgE(H151A)	N	N
GtgE(C45A/H151A)	Y	N
Rab29(V40L)	Y	N
Rab29(G41A)	N	N
Rab29(G41V)	N	N
Rab29(G41S)	N	N
Rab38(V44I)	Y	Y
Rab38(V44L)	Y	N
Rab38(F46Y)	Y	N
Rab38(Q69L)	Y	Y
Rab38(S35N)	Y	Y

Mutations in GtgE's catalytic core were focused on C45A, H151A, double mutant C45A and H151A, and double mutant C45A and C48A. Of these four, only the single C45A mutation yielded a stable complex over time. GtgE-C45A/C48A formed a complex with $Rab29^{1-177}$, but a significant cleavage band appeared during purification. The double catalytic mutant C45A/H151A also formed a complex with $Rab29^{1-177}$, but this complex dissociated during purification, possibly because the double mutation disturbed the active site in such a way as to decrease GtgE's ability to tightly grasp its Rab substrate. $GtgE^{H151A}$ did not form a complex with $Rab29^{1-177}$.

To examine GtgE's ability to recognize Rab38 in its GTP-bound ("on") state or its GDP-bound ("off") state, constitutively active (Q69L) and constitutively inactive (S35N) mutations on Rab38 were produced. The constitutively active glutamine to leucine mutation prevents the binding of the GTPase Activating Protein (GAP) so that there is no external stimulation of the Rab's GTPase activity [101]. The glutamine residue also plays an important role in GTP hydrolysis, forming a hydrogen bond with a water molecule positioned to attack the phosphoryl bond of GTP [102]. Since leucine cannot participate in such an interaction, the Q69L mutation also inhibits the Rab's intrinsic rate of GTP hydrolysis [101, 103].

The constitutively inactive mutant, serine to asparagine, increases the Rab's affinity for GDP [104]. The inactive mutant retains its ability to bind Guanine Nucleotide Exchange Factors (GEFs), but since exchange for GTP does not occur, the mutant Rab forms a dead-end complex with its GEF, and therefore, is trapped in its 'off' state [105]. $GtgE^{C45A}$ binds both $Rab38^{Q69L}$ and $Rab38^{S35N}$ with equal affinity, as judged by the ratio of GtgE to mutant and wild-type Rab38 purified in complex by size exclusion chromatography. Therefore, based on this work, GtgE does not show preference for either Rab conformation, and is able to recognize and bind stably to both the constitutively active and constitutively inactive Rab38 mutants.

We also made various mutations around the GtgE cleavage site on Rab29 and Rab38 to identify a mutation that led to a decrease in the appearance of lower molecular weight Rab products. Mutations in Gly41 in Rab29 prevented complex formation with $GtgE^{C45A}$, and only one mutation in Rab38, V44I, made a significant reduction in the formation of the 14kD cleavage band.

The purification of the $GtgE^{C45A}$/$Rab38^{V44I}$ complex on the Superdex 200, 120mL column (GE Healthcare) is shown in Figure 15A. $GtgE^{C45A}$ elutes at 89.23mL (Fig. 15B), $Rab38^{V44I}$ elutes at 93.52mL (Fig. 15C), and the complex elutes at 88.44mL (Fig. 15A). Through work with the wild-type complex (Rab38/$GtgE^{C45A}$), it was determined that the complex is most stable at a relatively neutral pH (pH 7.5-8.0), a salt concentration of 100mM-250mM sodium chloride, 5mM of reducing agent dithiothreitol, and a minimum of 100µM magnesium chloride. The Rab GTPases are not stable without the addition of magnesium chloride, as the magnesium ion is necessary for the protein's stable interaction with GTP/GDP. Complex preparations were performed with an excess of magnesium chloride, 5mM $MgCl_2$, and material used for crystallization was exchanged into buffer containing 100µM $MgCl_2$. Additionally, complex formation was affected by the method of expression used for the two components. A stable complex was only obtained when Rab38 and $GtgE^{C45A}$ were co-expressed on the same plasmid with Rab38 N-terminally His-tagged and $GtgE^{C45A}$ untagged.

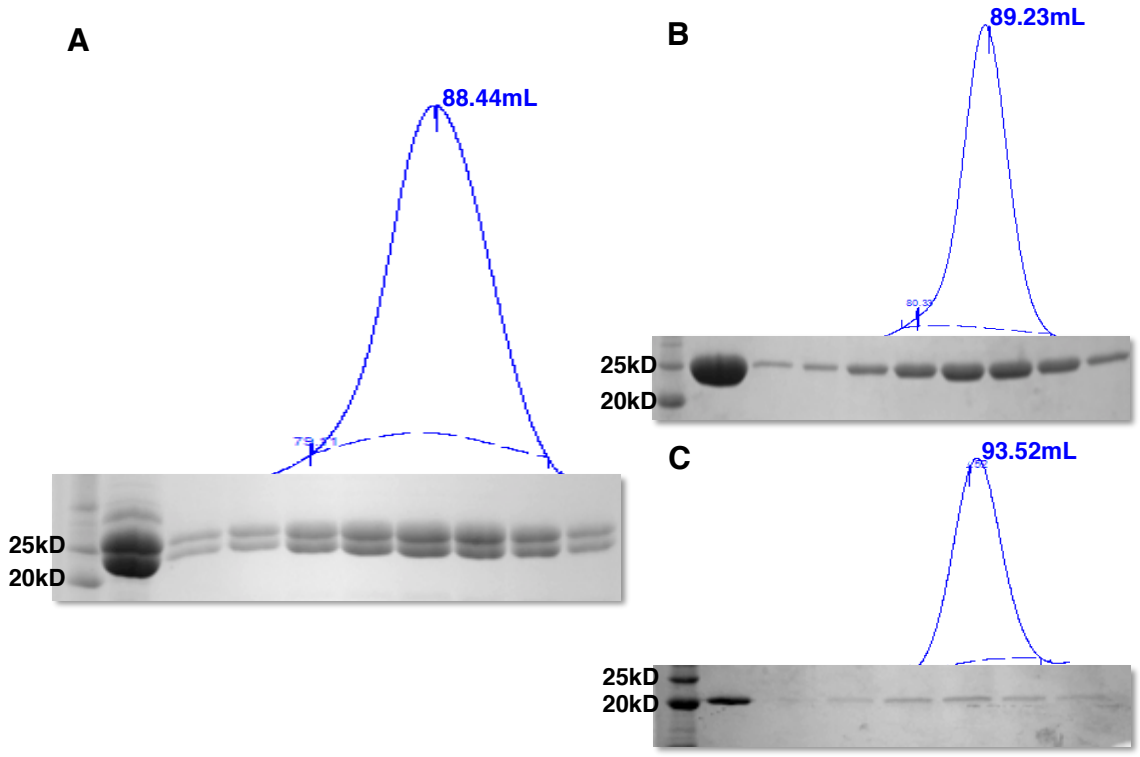


Figure 15. The purification of full-length Rab38^{V44I} and full-length GtgE^{C45A}. (A) GtgE^{C45A} and Rab38^{V44I}, in complex, elutes at 88.44mL on a Superdex200, 120mL column (GE Healthcare). (B) GtgE^{C45A} elutes at 89.23mL. (C) Rab38^{V44I} elutes at 93.52mL. Lane 1 (far left) on gels is the molecular weight marker, lane 2 is the column load material, and the remaining lanes are fractions corresponding to the elution peak.

Crystallization screens were carried out with each stable variation of the complex that was produced. The Rab38^{V44I}/GtgE^{C45A} material was subjected to limited proteolysis by subtilisin and trypsin to identify minimal constructs for crystallization. The subtilisin digest is shown in Figure 16. The complex was treated with subtilisin, ranging from 1.7ng to 0.261µg, and the reaction was

incubated for 30 minutes at 4°C. There were two main cleavage products at approximately 23kD and 19kD. Edman sequencing of these bands revealed that the 23kD product corresponds to full-length Rab38^{V44I} and GtgE$^{18-228,\ C45A}$, and the 19kD band corresponds to GtgE$^{57-228,\ C45A}$. A robust co-expression of Rab38^{V44I} with GtgE$^{18-228,\ C45A}$ was not possible, yielding low expression levels of mostly insoluble GtgE$^{18-228,\ C45A}$. However, a complex between Rab38 and GtgE$^{23-228,\ C45A}$ was successfully purified, but was unsuccessful in crystallization screens.

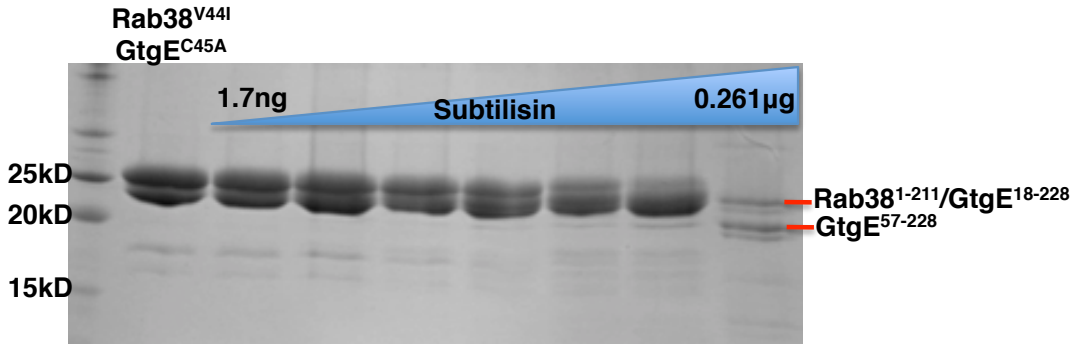


Figure 16. Limited proteolysis of the Rab38^{V44I} /GtgE^{C45A} complex. The complex was digested with 1.7ng to 0.261µg of subtilisin protease. The reaction was performed at 4°C for 30 minutes. Untreated complex material is shown in the far left lane.

CHAPTER 4. CONCLUSIONS

Members of the cysteine protease superfamily participate in a diverse array of biological pathways. The conserved active site triad is remarkably adaptable to seemingly disparate chemical reactions, functioning within enzymes as different as proteases, acetyltransferases, transglutamases, deamidases, and deubiqutinases. The catalytic activity of these enzymes is driven by a nucleophilic cysteine thiol and an electron-accepting histidine, and often requires a third residue for the proper orientation of this cysteine-histidine pair. Superfamily members share a conserved overall fold, but differ in the placement of catalytic triad residues within their canonical active site.

Interestingly, a sizeable portion of characterized bacterial T3SS effector proteins belong to the cysteine protease superfamily, utilizing this cysteine-powered catalytic core to manipulate the host's cellular processes in a variety of ways. For example, *Escherichia coli* encoded Cif, functions as a deamidase, using a Cys-His-Gln catalytic core to deamidate a critical glutamine in the ubiquitin-like NEDD8 protein [106]. Through binding to and deamidating NEDD8, Cif effectively prevents the E3 ligase activity of neddylated CRL complexes, resulting in cell cycle arrest, formation of stress fibers, and host apoptosis [107-112]. Another cysteine protease effector, *Yersinia pestis* Yop J, employs its His-Glu-Cys triad to acetylate serine and threonine residues on MAPK kinases and the IκB kinase complex, inhibiting both MAPK signaling and activation of the

NFκB pathway, respectively [113, 114]. YopJ's catalytic activity ultimately leads to the inhibition of innate and adaptive immunity responses and to the induction of cell death [113, 115-117]. In addition, *S.* typhimurium produces a deubiquitinase, SseL, which functions using a His-Asn-Cys triad to remove K63-linked ubiquitin chains from SCV-associated aggregates that are targeted for autophagic degradation [118, 119]. In this manner, SseL decreases the autophagic flux within the host, consequently contributing to down modulation of NF-κB-dependent cytokine production and macrophage-delayed cytotoxicity [118-121].

GtgE is the newest addition to this growing list of bacterial effector cysteine proteases. Produced by *S.* typhimurium, GtgE cleaves its Rab GTPase substrates, Rab29, Rab32, and Rab38, preventing the delivery of antimicrobial agents to the SCV; thereby, subverting one facet of the host's defense mechanism. In this manner, GtgE contributes to the ability of broad-host strains of *Salmonella* to maintain a diverse repertoire of host species. Our crystal structure of $GtgE^{80-213}$ firmly establishes that GtgE is a cysteine protease of Clan CA utilizing a catalytic triad of Cys45, His151, and Asp169. Biochemical studies with mutations in the active site residues show dramatic loss of activity against the Rab GTPases.

The inhibition of GtgE may serve as a means to alter the host specificity of broad-host *Salmonellae*. Equipped with an understanding of GtgE's function at the mechanistic level, we were able to identify three small molecules from a panel of cysteine protease inhibitors that are capable of inhibiting GtgE function *in vitro*: N-ethylmaleimide, antipain, and chymostatin. All of these compounds target the active thiol; however, due to their more specific inhibition, antipain and chymostatin are best suited to provide the foundation for more targeted inhibitor design. Chymostatin and antipain are bulkier microbial-derived small peptide inhibitors, which more closely mimic the cleavage site of GtgE's Rab GTPase substrates. Conversely, leupeptin, a microbial-derived inhibitor comprised of residues with relatively small side chains, targets the active thiol in the same manner as chymostatin and antipain, but does not inhibit GtgE.

Our work with the GtgE/Rab GTPase complex has led to the identification of *in vitro* conditions that maintain long-term stability of the complex. Of GtgE's substrates, we have shown that Rab38 is the most amenable to *in vitro* work and with additional purification optimization, the GtgE/Rab38 complex may prove to be the best option for crystallization of the complex. Through obtaining a structure of GtgE in complex with its substrate, we hope to characterize GtgE's interaction and possible recognition of its substrate, providing a more thorough understanding of how GtgE functions *in vivo*.

Since Rab GTPases undergo significant structural changes depending on whether they are bound to GDP or GTP, it is imperative to crystallization success that the Rab protein under study adopts a homogenous conformation. We attempted to address this by making constitutively active, GTP-bound conformation, and constitutively inactive, GDP-bound conformation, mutations to Rab38. We did not find a measurable difference in complex formation or stability between GtgE and the constitutively active or inactive mutant of Rab38, and neither complex yielded crystals. Another method that may prove more useful in creating Rab homogeneity is to chemically induce nucleotide loading of either GDP or a non-hydrolyzable analog of GTP, such as Gpp(NH)p. This method may better mimic the GDP/GTP-bound conformations of Rab *in vitro*, perhaps leading to a more tightly bound, stable complex and potentially aiding in crystallization.

APPENDIX I. *SALMONELLA* INVASION PROTEIN C

AI.1 Introduction

Salmonella Invasion Protein C (SipC/SspC) is encoded on *Salmonella* pathogenicity island 1 and is secreted by the T3SS1 [122]. Expression and secretion of SipC is controlled by its chaperone, *Salmonella* invasion chaperone A (SicA) [123, 124]. SipC plays a critical role in *Salmonella* infection, serving as both a translocase and an effector protein, and is necessary for host invasion [125, 126]. SipC is a 42kD protein with one predicted transmembrane domain and two effector domains (Fig. 17).

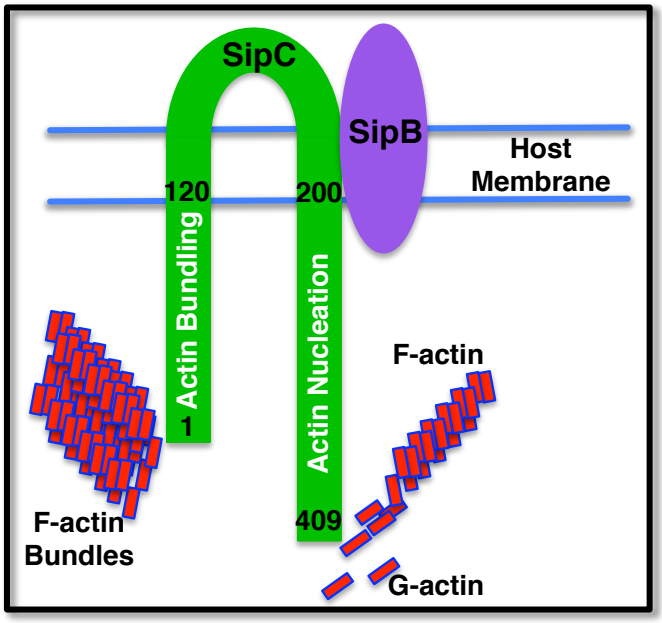


Figure 17. SipC and host cell invasion. SipC is a 409 residue effector and translocase protein, and is predicted to oligomerize with SipB to form the translocon pore in the host cell membrane. The N-terminal 120 residues of SipC extend into the host cytosol where they function in actin bundling, while the C-terminal 209 residues nucleate actin polymerization.

Shigella flexneri invasion plasmid antigen C (IpaC) is highly homologous to SipC, sharing 34% amino acid identity and 54% similarity with SipC [127]. IpaC and SipC share a significant sequence homology in their respective hydrophobic domains and C-terminal effector domain, and both are known to interact with actin [128]. A key difference in the way *Shigella* and *Salmonella* infect cells is that *Salmonella* remains within the SVC after host entry, while *Shigella* escapes into the host cytoplasm [128]. SipC and IpaC appear to be important players in this difference in infection behavior, although the exact cause of this difference remains unknown [128].

AI.1.1 The Translocon

SipC oligomerizes with *Salmonella* Invasion Protein B (SipB), another translocase protein, to form the translocon pore in the host cell membrane. The two proteins have been shown to associate *in vitro* [30], but it is not understood how they are inserted into, and associate within, the membrane. SipC is not detected in the host cell membrane in the absence of SipB, suggesting that SipC is dependent on SipB for membrane insertion [25, 127]. It is unclear whether SipB and SipC oligomerize into a ring structure prior to membrane insertion, but complex formation between SipC and SipB is essential for their insertion into the membrane, and a domain of SipC, residues 340 to 409, is required for its interaction with SipB [129]. Once inserted into the membrane, SipC is hypothesized to assume a hairpin-like formation, such that its central

hydrophobic domain resides within the plasma membrane while its N- and C-terminal effector domains extend into the host cell cytoplasm (Fig. 17) [25].

AI.1.2 SipC: N-terminal Domain

Functioning as a *Salmonella* effector protein, SipC participates in the rearrangement of the host's actin cytoskeleton, aiding in *Salmonella*'s initial uptake by the host. Actin rearrangement by SipC appears to be enhanced by the action of another *Salmonella* invasion protein, SipA, suggesting that SipC cooperates with SipA *in vivo* [130]. Both SipC and SipA appear to bind directly to actin and influence filament dynamics independently of host cell components [30]. The N-terminal domain of SipC, spanning residues 1 to 120, bundles actin filaments in a 1:1 SipC to actin filament molar ratio *in vitro* [30].

Actin is a 42kD globular protein that serves as the building block for microfilament formation in eukaryotic cells. There is a large variety of eukaryotic actin binding proteins that help to tightly regulate microfilament dynamics inside the cell. Monomeric actin, known as G-actin, polymerizes to form long, filamentous, double-helical polymers, called F-actin. Proteins responsible for actin bundling, such as the N-terminal domain of SipC, are able to group together F-actin strands into bundles (Fig. 17). These bundles function primarily in maintaining the overall cell structure; allow for force generation; and are also

important for phagocytosis [131, 132]. Through actin bundling, the N-terminal domain of SipC stimulates host phagocytosis and uptake of the *Salmonella* cell.

AI.1.3 SipC: C-terminal Domain

The C-terminal domain of SipC, residues 200 to 409, nucleates actin polymerization *in vitro* (Fig. 17) [30]. Since the G-actin concentration inside the cell is kept below the critical concentration, the concentration above which G-actin polymerizes, the action of actin-nucleating proteins, such as SipC, is necessary for actin polymerization [33]. Actin polymerization follows three distinct phases: lag phase, filament growth, and steady state (Fig. 18). Nucleation begins during the lag phase and continues until approximately 45% of the actin is polymerized [133]. During filament growth, filaments polymerize from nucleating centers, and a steady state is achieved once the concentration of monomeric actin drops to the critical concentration causing polymerization to cease [133]. Proteins that nucleate actin polymerization increase the number of nucleating centers; therefore, decreasing the length of the lag phase, enabling polymerization to begin sooner, and steady state to be achieved faster, than with actin alone. By nucleating actin polymerization, the C-terminal domain of SipC enables *Salmonella* to exploit host actin, ultimately leading to the internalization of the bacterial cell by the host.

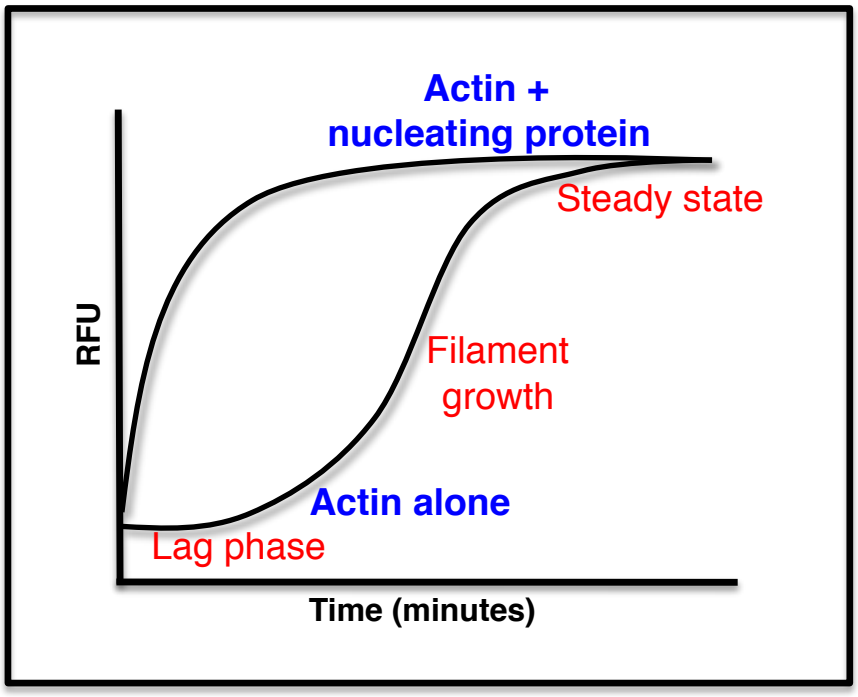


Figure 18. Actin polymerization. The polymerization of actin occurs in three stages: lag phase, filament growth, and steady state. When a nucleating protein is added to monomeric actin, the lag phase is eliminated and filament formation occurs logarithmically.

SipC has an activity range similar to that of the Arp 2/3 complex, which is considered to be a relatively weak nucleator *in vitro* [30, 134]. Thus far, several studies have investigated the actin polymerization ability of SipC's C-terminal domain, and have found that, in order to function *in vitro*, this domain multimerizes via contact points located between residues 201 to 220 [135]. Through in-frame deletion/insertion mutations, residues 201 to 220 were also found to be essential for actin nucleation, while residues 321 to 409 were required for translocation of effectors [25, 136].

Two eukaryotic proteins, cytokeratin 8 and Exo70, are believed to interact with the C-terminal domain of SipC. Yeast two-hybrid experiments with HeLa cell cDNA identified cytokeratin 8, a 54kD intermediate filament protein, as a potential binding partner of SipC [25]. Cytokeratin 8 is comprised of three domains: a non-helical N-terminal domain, an alpha helical rod domain, and a non-helical tail [25]. In this yeast two-hybrid screen, SipC was found to interact with a C-terminal segment of cytokeratin 8, spanning roughly two-thirds of the second coil in the rod domain in addition to the entire tail domain [25]. The central alpha helical domain forms heterodimers with cytokeratin 18 to assemble into dynamic filaments that are approximately 10nm in diameter and extend from the nucleus to the plasma membrane [137]. Mutations in the C-terminal effector domain of SipC disrupted the interaction between SipC and cytokeratin 8, suggesting that SipC directly interacts with this filament protein and may utilize this interaction to influence the cytoskeletal structure of the host [25].

A study by Nichols, *et al.* found that the C-terminal domain of SipC is sufficient to cause the co-precipitation of endogenous Exocyst Complex Component 7 (Exo70, EXOC7) from HeLa cell lysate [138]. Exo70 is one of eight exocyst complex proteins that, together, mediate the docking and fusion of exocytic vesicles with the plasma membrane. Exo70 is an 84kD protein composed of three helical domains, termed the N, M, and C domains [139]. Exo70 interacts with several Sec proteins (Sec6, Sec8, Sec10, and Sec15), as

well as Arpc1 and numerous small GTPases to target and tether exocytic vesicles to the plasma membrane in preparation for SNARE-mediated fusion [139, 140]. Through its interplay with Arpc1, a subunit of the Arp2/3 complex, Exo70 stabilizes actin reorganization at the plasma membrane [141]. SipC is capable of binding Exo70 in its unbound state as well as when it is assembled into the exocyst complex [138]. SipC's interaction with Exo70 serves as another means for *Salmonella* to gain access to the host actin cytoskeleton, enabling the bacteria to encourage its own internalization through host phagocytosis.

As an effector protein, SipC is able to interact with host actin both directly, through its N- and C-terminal domains, and indirectly, via its interaction with cytokeratin 8 and Exo70. Despite this functional characterization, it is difficult to predict how SipC carries out actin rearrangement because the effector lacks primary sequence homology to other known eukaryotic actin binding proteins [130]. Our aim was to provide structural insight into how this key component of *Salmonella* invasion performs its various effector roles by obtaining the X-ray structure of SipC.

AI.2 Materials and Methods

AI.2.1 Protein Expression and Purification

SipC$^{1-120}$, SipC$^{1-120(7+)}$, and SipC$^{5-120}$ were cloned between the *SalI* and *NotI* restriction sites in a modified pGEX4 vector (GE Healthcare), containing a rhinovirus 3C protease cleavage site C-terminal to GST, and protein was expressed in *E.* coli BL21(DE3) cells for 16 hours at 21°C with 1mM IPTG. Cells were lysed by high-pressure homogenization into lysis buffer (50mM Tris-HCl, pH 8.0, 200mM NaCl, and 10mM DTT). Lysate was centrifuged for 40 minutes at 16,000rpm and supernatant was first passed over a Q-sepharose column (GE Healthcare) that was pre-equilibrated in lysis buffer. Q-sepharose flow-through was then run over a glutathione-sepharose column (GE Healthcare) pre-equilibrated with lysis buffer. The column was washed with 50mM Tris-HCl, pH 8.0, 500mM NaCl, and 10mM DTT, and protein was eluted from the column with 50mM Tris-HCl, pH 8.0, 200mM NaCl, 10mM DTT, and 10mM glutathione. Eluted protein was dialyzed for 16 hours at 4°C with in-house prepared 6xHis-rhinovirus 3C protease into 50mM Tris-HCl, pH 8.0, 150mM NaCl, and 5mM DTT. After dialysis, the protein was passed over a NiNTA column (Qiagen) to remove the histidine-tagged 3C protease. In order to separate free GST from untagged SipC$^{1-120}$, the material was diluted with 50mM Tris-HCl, pH 8.0, 5mM DTT to a NaCl concentration of 50mM NaCl, and was then loaded onto a Source Q column (GE Healthcare) that was pre-equilibrated in 50mM Tris, pH 8.0, 5mM

DTT. The Source Q column was then washed with 50mM Tris-HCl, pH 8.0, 5mM DTT and the bound proteins, free GST and SipC$^{1-120}$, were eluted from the column separately using a salt gradient of 0-100% NaCl, run over 160 minutes. Free GST eluted with about 100mM NaCl and SipC$^{1-120}$ eluted at approximately 300mM NaCl. SipC$^{1-120}$ was then run over a 120mL Superdex 75 column (GE Healthcare) into 50mM Tris-HCl, pH 8.0, 150mM NaCl, and 5mM DTT. Peak fractions were collected and concentrated using a 3kD Amicon Ultra centrifugal concentrator (Millipore), and loaded onto a 25mL Superdex 200 column (GE Healthcare) into a final buffer containing 25mM Tris-HCl, pH 8.0, 150mM NaCl, and 2mM DTT.

SipC$^{1-120}$ was also cloned into a modified pCDF-Duet-1 vector (Novagen) between the *SalI* and *NotI* restriction sites. Protein was expressed in *E.* coli BL21(DE3) cells at 21°C with 1mM IPTG for 16 hours. Cells were lysed with high-pressure homogenization into 50mM Tris-HCl, pH 8.0, 300mM NaCl, and lysate was centrifuged at 16,000rpm for 40 minutes. The resulting pellet was resuspended in 6M urea, 0.5M NaCl, and 20mM Tris-HCl, pH 8.0, and centrifuged again at 16,000rpm for 40 minutes. The supernatant was applied to NiNTA resin (Qiagen) that had been equilibrated in the resuspension buffer. The column was washed with 6M urea, 0.5M NaCl, 30mM imidazole, and 20mM Tris-HCl, pH 8.0. Protein was eluted with 6M urea, 0.5M NaCl, 500mM imidazole, and 20mM Tris-HCl, pH 8.0. Work was carried out at room temperature.

SipC$^{1-120}$ was refolded via dialysis refolding at 4°C with protein at 0.2 mg/mL as detailed in section 2.1.1. 6xHis-rhinovirus 3C protease was added during the final stage of dialysis to remove the histidine tag, and the 3C protease and cleaved histidine tag were separated from SipC$^{1-120}$ via a passage over a NiNTA column, and SipC$^{1-120}$ was then further purified over a 120mL Superdex 75 column (GE Healthcare) and 25mL Superdex 200 column (GE Healthcare) as described above.

SipC$^{200-409}$ and all other SipC C-terminal domain constructs were cloned into a modified pCDF-Duet-1 vector (Novagen) between the *SalI* and *NotI* restriction sites. Protein was expressed in *E.* coli BL21(DE3) competent cells with 1mM IPTG at 21°C for 16 hours. Cells were lysed via high-pressure homogenization into lysis buffer (50mM HEPES, pH 7.5, 300mM NaCl, and 5mM imidazole). Lysate was centrifuged for 40 minutes at 4°C and 16,000rpm. Supernatant was passed over a NiNTA column (Qiagen) pre-equilibrated in lysis buffer at 4°C. The column was washed with 50mM HEPES, pH 7.5, 300mM NaCl, 30mM imidazole, and protein was eluted with 50mM HEPES pH 7.5, 300mM NaCl, 250mM imidazole. Eluted protein was dialyzed with 6xHIs-rhinovirus 3C protease for 16 hours at 4°C against 25mM HEPES, pH 7.5, 200mM NaCl, and 5mM DTT. The protein was then run over a NiNTA column (Qiagen) to remove the histidine tag and 3C protease. Final purification was

done on a 120mL Superdex 75 column (GE Healthcare) into 25mM HEPES, pH 7.5, 200mM NaCl, and 5mM DTT.

M. musculus Exo70$^{1-653}$ and Exo70$^{85-653}$ (mExo70) (Source Bioscience) were inserted between the *SalI* and *NotI* restriction sites of a modified pCDF-Duet-1 vector (Novagen) containing a hisitidine-NusA tag and rhinovirus 3C protease cleavage site. Purification of mExo70 was performed according to the purification scheme described for SipC C-terminal domain constructs, except that buffers were made with 25mM Tris-HCl, pH 8.0 and the final purification step was done on a 120mL Superdex 200 column (GE Healthcare) in 25mM Tris-HCl, pH 8.0, 200mM NaCl, and 5mM DTT. Both mExo$^{1-653}$ and mExo$^{85-653}$ eluted as monomers under these conditions.

To co-express mExo70$^{85-653}$ with SipC$^{200-409}$, SipC$^{200-409}$ was cloned into the second cloning site of a modified NusA-containing pCDF-Duet-1 vector (Novagen) (described above) between the *NdeI* and *XhoI* restriction sites. Purification of the complex was carried out in the same manner as for mExo70 alone, using buffers containing 25mM Tris, pH 8.0. To further probe for complex stability, the elution from NiNTA resin (Qiagen) was applied to a 120mL Superdex 200 column (GE Healthcare) in 25mM Tris pH, 8.0, 200mM NaCl, and 5mM DTT.

S. typhimurium SpvB was purified as described by Margarit, *et al.* [142]. Briefly, SpvB was histidine-tagged and purified using standard NiNTA procedures and protein was eluted in buffer containing 50mM Tris-HCl, pH 8.0, 300mM NaCl, 250mM imidazole. The final stage of purification was performed on a 120mL Superdex 75 column (GE Healthcare) into 10mM Tris-HCl, pH 8.0, 100mM NaCl, and 2mM DTT.

AI.2.2 Limited Proteolysis

Histidine-tagged $SipC^{200-402}$ was subjected to limited proteolysis using subtilisin protease (Sigma Aldrich). 11.5µg of $SipC^{200-409}$ was incubated with 3.45ng to 11.5ng of subtilisin and 5mM $CaCl_2$ for 30 minutes at 4°C. In a similar reaction, 8.05µg of $SipC^{200-330}$ was treated with 48.3ng to 0.483µg of trypsin (Sigma Aldrich) and 5mM $CaCl_2$ for 30 minutes at room temperature. The reactions were terminated by the addition of 10mM PMSF and SDS running buffer. Samples were boiled and run on 15% SDS-PAGE and transferred to PDVF membrane (Millipore) for Edman sequencing by the Rockefeller University Proteomics Resource Center (New York, NY).

AI.2.3 Reductive Methylation of Lysine Residues

The lysine residues of $SipC^{200-409}$, $SipC^{200-295}$, $SipC^{200-290}$, and $SipC^{195-278}$ were reductively methylated using a protocol adapted from Rayment [143]. The

SipC constructs were used in a concentration range of 0.1-0.5mg/mL in a 2mL volume and were in a buffer containing 25mM HEPES, pH 7.5, 200mM NaCl, and 5mM DTT. 40mM dimethyl-amine-borane (DMAB, Sigma Aldrich), prepared in water, and 80mM methanol-free formaldehyde (Polysciences, Inc.), prepared in water, were added to SipC in a falcon tube. The tube was wrapped in aluminum foil and allowed to incubate at 4°C for 2 hours with gentle mixing. An additional 40mM DMAB and 80mM formaldehyde were added to the SipC methylation reaction and the reaction was again incubated, gently mixing, at 4°C for 2 hours. 5mM DMAB was added to the sample, and the reaction was slowly shaken for 16 hours at 4°C. 250mM ammonium sulfate was then added to the sample to end the methylation reaction. As a final purification step, and to confirm successful methylation, SipC samples were run over a 120mL Superdex 75 column (GE Healthcare) into 25mM HEPES, pH 7.5, 200mM NaCl, and 5mM DTT.

AI.2.4 Crystallization

Crystallization screens with $SipC^{200-295}$ were performed in a 96-well sitting drop format using approximately 5mg/mL protein in a 1:1 ratio with reservoir buffer to a total volume of 2µL. Spherulite formation was observed at room temperature in a wide range of different conditions, though mostly at a relatively neutral pH. Spherulite hits were globular with a smooth surface.

The SipC$^{195-278}$ crystal screen set-up was analogous to that used for SipC$^{200-295}$, however, the 2µL sitting drops contained 16mg/mL protein in a 1:1 ratio with reservoir buffer. Spherulite formation was seen at room temperature and at 4°C predominantly in conditions containing low molecular weight PEGs (PEG 2000-5000). Reductively methylated SipC$^{195-278}$ was also tested for crystallization under the same screening conditions used for native SipC$^{195-278}$, and spherulites with thicker needle-like extensions were obtained in 0.1M sodium acetate pH 4.6, 2M ammonium sulfate.

AI.2.5 SipC and Ribosylated Actin

Rabbit skeletal muscle actin (Cytoskeleton, Inc.) was ribosylated as detailed by Margarit, *et al.* [142]. The ribosylation reaction contained 1mg of actin, 100ug purified SpvB, and 200µM NAD in a buffer containing 5mM Tris-HCl, pH 8.0, 0.2mM ATP, 0.2mM $CaCl_2$, and 2mM DTT to a final volume of 1mL. The reaction was incubated at 4°C for 16 hours with gentle mixing. 100µg SpvB in a polymerization buffer with a final composition of 50mM KCl, 2mM $MgCl_2$, 0.2mM ATP was added to the reaction and the mixture was incubated at room temperature for 2 hours. The reaction was then centrifuged at 100,000xg for 1 hour at 4°C and the supernatant was further purified on a 120mL Superdex 75 column (GE Healthcare) into 5mM Tris-HCl, pH 8.0, 0.2mM $CaCl_2$, 0.2mM ATP, 100mM NaCl, and 2mM DTT. Ribosylated actin eluted as a monomer at 60.18mL.

Purified SipC$^{200-330}$ (in 25mM HEPES, pH 7.5, 250mM NaCl, 5mM DTT) was mixed with ribosylated actin in a 1:1 molar ratio, SipC to actin. The binding reaction was carried out in the presence of 150mM NaCl at 4°C for 16 hours with gentle mixing. The reaction was then run over a 120mL Superdex 75 column (GE Healthcare) in 25mM Tris-HCl, pH 8.0, 0.2mM $CaCl_2$, 0.2mM ATP, 150mM NaCl, and 2mM DTT.

AI.2.6 Actin Polymerization Assay

Our actin polymerization assay protocol was based on that detailed by Cytoskeleton, Inc. (Denver, CO) and Machesky, *et al.* [144]. Pyrene-actin (Cytoskeleton, Inc.) was diluted to 0.45mg/mL using G-buffer (5mM Tris-HCl, pH 8.0, 0.2mM $CaCl_2$), 0.2mM ATP, and 1mM DTT, and allowed to sit 1 hour on ice. The pyrene-actin was then centrifuged at 14,000rpm at 4°C for 30 minutes, and the supernatant, containing monomeric pyrene-actin, was removed for use in subsequent experiments. 10µM of unlabeled actin (Cytoskeleton, Inc.) reconstituted in G-buffer was incubated at room temperature for 60 minutes to induce filament formation. Arp2/3 was diluted to 0.3mg/mL in G-buffer prior to the start of the reaction.

The assay was performed in a black-bottom 96-well plate (Sigma Aldrich) at room temperature. Each reaction was performed in a 300µL reaction volume. 200µL of P-buffer (7.5mM Tris-HCl, pH 8.0, 75mM KCl, 3mM $MgCl_2$, and 1.5mM

ATP) was aliquoted into each well. 0.3μM of pre-incubated unlabeled actin in G-buffer was added to the reaction, followed by SipC$^{200-409}$ (0.5μM, 1μM, 2.5μM) in 25mM Tris, pH 8.0, 250mM NaCl, and 5mM DTT, and finally, 0.8μM monomeric pyrene-actin in G-buffer was added to the reaction. The reaction was shaken initially to mix and shaken again before each time point was taken. Time points were taken every 30 seconds for 40 minutes. Excitation was measured at 365nm and emission was measured at 407nm. Arp2/3 (Cytoskeleton, Inc.) together with the VCA domain of WASp (Cytoskeleton, Inc.) served as a positive control.

AI.3 Results and Discussion

AI.3.1 SipC N-terminal Domain

The N-terminal domain of SipC, spanning residues 1 to 120, was found to be either insoluble or unstable in most of the purification schemes examined. Purification attempts focused on the following constructs: $SipC^{1-120}$, $SipC^{5-120}$, and a full construct with an additional random seven residues, LERPHRD, inserted after residue 120, denoted $SipC^{1-120(7+)}$. The $SipC^{5-120}$ construct was identified by M. Lilic as a potential minimal construct through limited proteolysis with subtilisin protease (Sigma Aldrich). GST-tagged $SipC^{1-120}$ and $SipC^{5-120}$ were insoluble (Fig. 19A); however, $SipC^{1-120(7+)}$ was partially soluble and purified material could be concentrated to 0.67mg/mL before precipitation would occur. This construct was also sensitive to salt concentration, and would precipitate under low NaCl (<100mM NaCl) conditions.

Histidine-tagged $SipC^{1-120}$ was also partially soluble and the protein was amenable to purification under denaturing conditions and subsequent dialysis-based refolding (Fig. 19A). Using these methods, $SipC^{1-120}$ was purified with a total yield of 0.7mg of protein per liter of cell culture grown (Fig. 19B). Crystal screens were set-up using refolded $SipC^{1-120}$ at the highest concentration attainable, 1.83mg/mL. These screens did not lead to successful crystallization.

A

Construct	Solubility	Yield
GST-SipC(1-120)	No	
GST-SipC(1-120^{7+})	Partially	unstable
GST-SipC(5-120)	No	
His-SipC(1-120)	Partially	0.7 mg/L
His-SipC(5-120)	No	

B

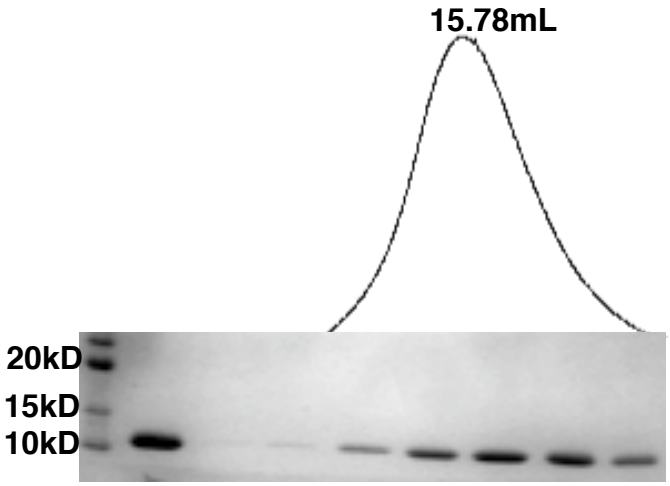


Figure 19. Purification of the N-terminal domain of SipC. (A) The constructs tested for solubility and stability are listed along with their respective solubility and yield. (B) The final purification step for refolded SipC$^{1-120}$ was done on a 25mL Superdex 200 column (GE Healthcare). SipC$^{1-120}$ runs as a trimer, eluting at 15.78mL. Lane 1 (far left) contains the molecular weight marker, lane 2 is the column load material, and the remaining lanes are the fractions corresponding to the elution peak.

AI.3.2 SipC C-terminal Domain

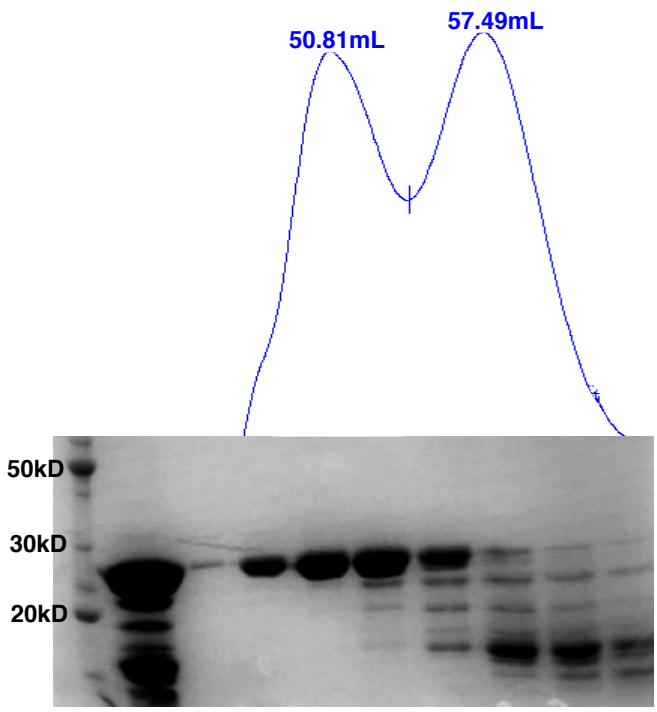
AI.3.2.1 Domain Determination and Purification

Our initial SipC C-terminal domain construct spanned residues 200 to 409, as this was the domain defined in the literature. This construct was purified with an N-terminal histidine tag as described, and was further purified by size exclusion chromatography using a Superdex 75 column (GE Healthcare) with a 120mL column volume into 20mM HEPES, pH 7.5, 200mM NaCl, and 5mM DTT (Fig. 20). Purification using buffers of varying pH were also examined. SipC$^{200\text{-}409}$ was most stable in a 6-8 pH range, and purification in Tris-HCl, pH 8.0, HEPES, pH 7.5, bis-tris-propane, pH 6.8, and MES, pH 5.8 showed similar protein yields and long-term stability. SipC$^{200\text{-}409}$ showed significant precipitation when purified in acetate buffer at pH 4.6.

The histidine tag used to purify the SipC$^{200\text{-}409}$ domain was not cleavable by 3C protease, most likely due to the construct folding in such a way as to obscure the cleavage site. Therefore, SipC$^{200\text{-}409}$, shown in Figure 20A, contains the N-terminal histidine tag and runs approximately 7kD larger than it would without the tag. The His-SipC$^{200\text{-}409}$ construct elutes at 50.18mL, running as a tetramer, on a 120mL Superdex 75 column (GE Healthcare). A characteristic banding pattern, consisting of three predominant lower molecular weight bands, elutes at 57.49mL and was observed in the later stages of purification for the full-

length C-terminal domain (Fig. 20A). This cleavage may be due to activity of proteases inherent to *E. coli* that were still present in the purified material.

A



B

Residue	MW (kD)	pI	Yield (mg/L)
200-409	23	9.3	0.6
200-374	18.8	8.3	
200-330	14.3	9.2	
200-320	13.1	9.2	
200-308	11.5	9.0	
200-300	10.7	10.0	
200-295	10	9.3	
200-290	9.6	9.3	
195-400	23	9.2	0.75
195-295	10.6	9.2	
195-278	8.8	9.3	1.33

Figure 20. The purification of the C-terminal domain of SipC. (A) Histidine-tagged SipC$^{200-409}$ was purified on a 120mL Superdex 75 column into 25mM HEPES, pH 7.5, 200mM NaCl, and 5mM DTT. The full-length construct elutes at 50.81mL, and the characteristic cleavage products elute at 57.49mL. Lane 1 (far left) contains the molecular weight marker, lane 2 is the column load material, and the remaining lanes are the fractions corresponding to the elution peak. (B) The C-terminal domain constructs that were produced and screened for crystallization. Yields are estimates based on comparison to protein of known concentration by SDS-PAGE because the C-terminal domain of SipC does not contain enough aromatic residues to contribute to an accurate UV absorbance, or arginines for meaningful Bradford assay results.

Limited proteolysis using subtilisin on native SipC$^{200-409}$ and denatured and refolded SipC$^{200-409}$ proved unsuccessful in determining a minimal construct (Fig. 21A), so the cleavage products present during purification were sent to the Columbia University Protein Core Facility (New York, NY) for Edman sequencing. N-terminal sequencing indicated that cleavage was occurring C-terminally, defining constructs starting at residue 200 and ending at residue 360, 330, or 320 (Fig. 21B). C-terminal truncations were estimated based on the molecular weight of the cleavage products. This information, combined with secondary structure predictions, led to the construction and purification of SipC$^{200-374}$ and Sip$^{200-320}$. Constructs with various N-terminal starts, at residue 180, 190 and 195, were screened as well; however, only those beginning with residue 195 were screened for crystallization. In addition to the constructs listed in Figure 20B, denatured and refolded SipC$^{200-409}$ and SipC$^{195-400}$, as well as reductively methylated SipC$^{200-409}$, SipC$^{200-295}$, SipC$^{200-290}$, and SipC$^{195-278}$ were tested for crystallization ability.

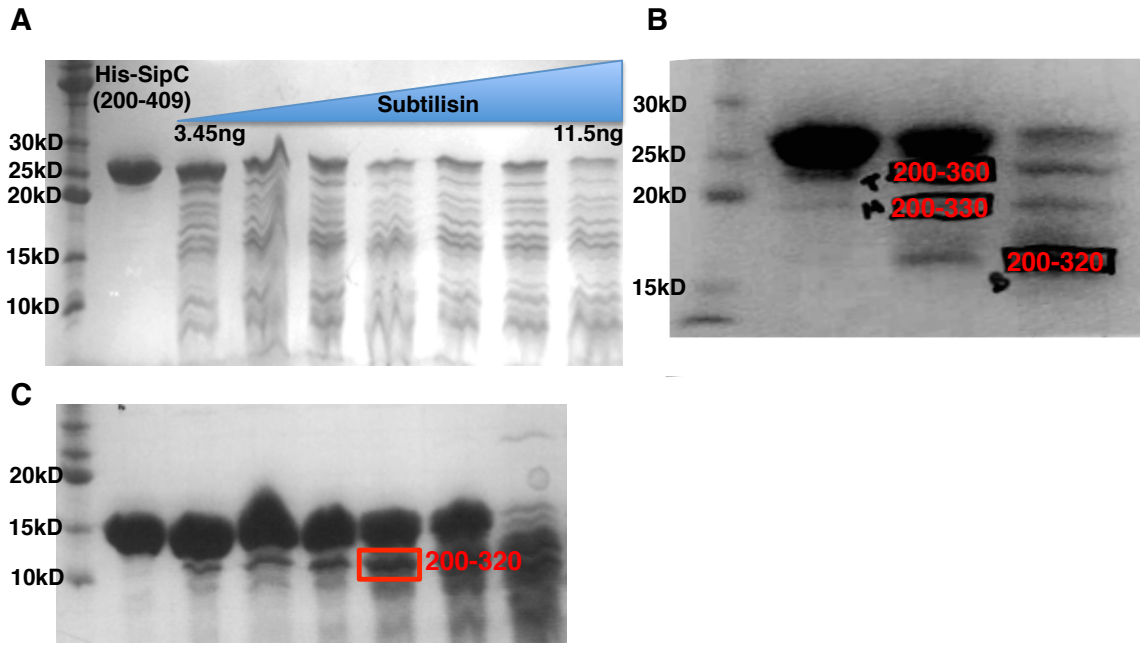


Figure 21. Limited proteolysis of the SipC C-terminal domain. (A) The subtilisin digestion of His-SipC$^{200\text{-}409}$ was performed at 4°C for 30 minutes using 3.45ng to 11.5ng of subtilisin and 11.5µg of SipC. (B) His-SipC$^{200\text{-}409}$ cleavage products appeared during purification, and C-terminal cleavage was determined via Edman sequencing. Corresponding truncations are shown in red. (C) SipC$^{200\text{-}330}$ was treated with trypsin. The band sent for Edman sequencing is indicated in red.

Upon removal of the C-terminal 79 residues for the SipC$^{200\text{-}330}$ construct, 3C protease was able to access the cleavage site and the histidine tag was removed. This suggests that the C-terminus of SipC$^{200\text{-}409}$ was folded towards the N-terminus of the construct and restricted 3C protease's access to the cleavage site. A trypsin digest of SipC$^{200\text{-}330}$ revealed an additional C-terminal cleavage product corresponding to a construct spanning residues 200 to 320,

which was estimated based on molecular weight of the cleavage product and the location of trypsin cleavage sites (C-terminal to arginine and lysine residues) (Fig. 21C). Crystallization trials were set-up with constructs in as high a concentration as could be obtained, usually between 2-8mg/mL. Higher concentrations could be attained with the shorter SipC C-terminal domain constructs, and SipC$^{195-278}$ could be concentrated to 16mg/mL.

We hypothesized that the instability of SipC could stem from exposed membrane-associated regions still potentially present in our constructs. In an effort to mask these areas, we investigated whether the addition of 1M guanine HCl or 1% octyl glucoside (w/v) could enable a higher concentration of SipC$^{200-330}$ to be achieved. We found that 1% octyl glucoside (w/v) allowed for the concentration of SipC$^{200-330}$ to roughly 1.5 times greater than that of the material concentrated in the presence of 1M guanine HCl and approximately 2 times greater than material without either additive, suggesting that the addition of a mild detergent may aid in purification of the SipC C-terminal domain.

AI.3.2.2 Crystallization

SipC$^{200-295}$ was the largest construct tested that produced crystalline growth in crystallization trials (Fig. 22). The purification of SipC$^{200-295}$ was performed as described, and protein eluted as a tetramer on a 120mL Superdex 75 column (GE Healthcare). SipC$^{200-295}$ shows little to no UV absorbance due to

the absence of aromatic residues in the construct. The typical crystal morphology seen with SipC$^{200-295}$ consisted of various spherulites and needle-like clusters that grew in a wide variety of low- to mid-range molecular weight PEG conditions (Fig. 22B). Further screening with PEG-containing conditions; the addition of various additives; and reductive methlyation of SipC$^{200-295}$ did not improve the crystal morphology.

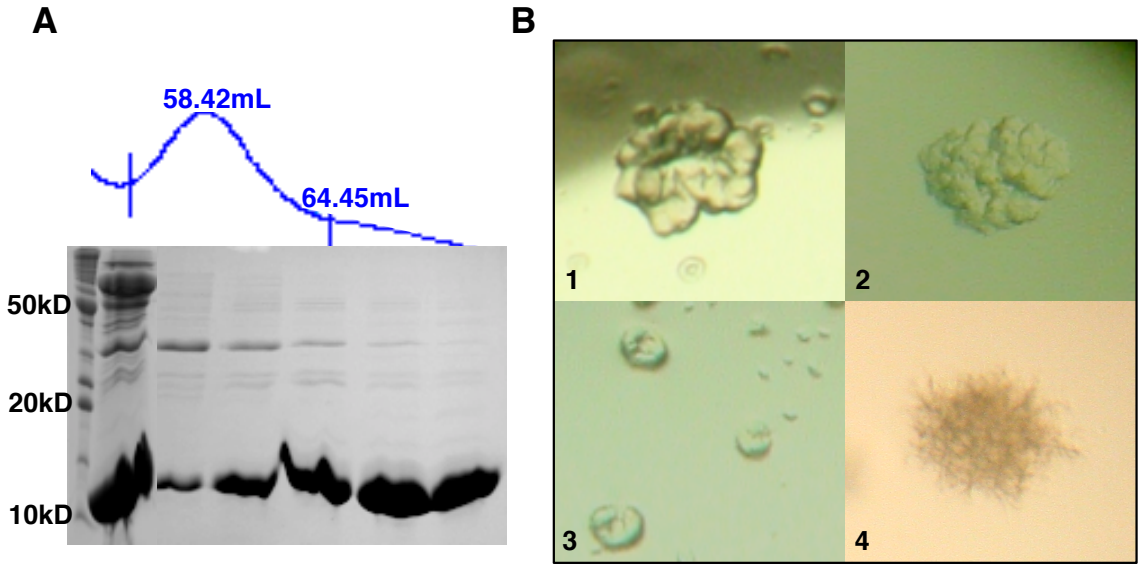


Figure 22. The purification and crystallization of SipC$^{200-295}$. (A) SipC$^{200-295}$ elutes at approximately 65mL on a 120mL Superdex 75 column (GE Healthcare). Lane 1 (far left) contains the molecular weight marker, lane 2 is the column load material, and the remaining lanes are the fractions corresponding to the elution peak. (B) Examples of the crystal morphologies seen with SipC$^{200-295}$. Crystallization conditions are as follows: (1) 0.8M sodium phosphate/1.2M potassium phosphate, 0.1M acetate, pH 4.5; (2) 0.2M ammonium sulfate, 0.1M bis-tris, pH 5.5, 25% PEG 3350; (3) 1M sodium citrate, 0.1M tris, pH 7.0, 0.2M NaCl; (4) 0.2M sodium cacodylate, pH 6.5, 30% PEG 8000.

The most promising crystal morphology was obtained with SipC$^{195-278}$ (Fig. 23A), which produced birefringent needle spherulites (Fig. 23B). Reductive methylation of the lysine residues of SipC$^{195-278}$ led to the formation of thicker, individual needles (Fig. 23C); however, optimization of crystal conditions did not lead to crystals suitable for X-ray diffraction experiments. Seeding using the methylated SipC$^{195-278}$ needles was performed to encourage the growth of larger, ordered crystals, but also proved unsuccessful.

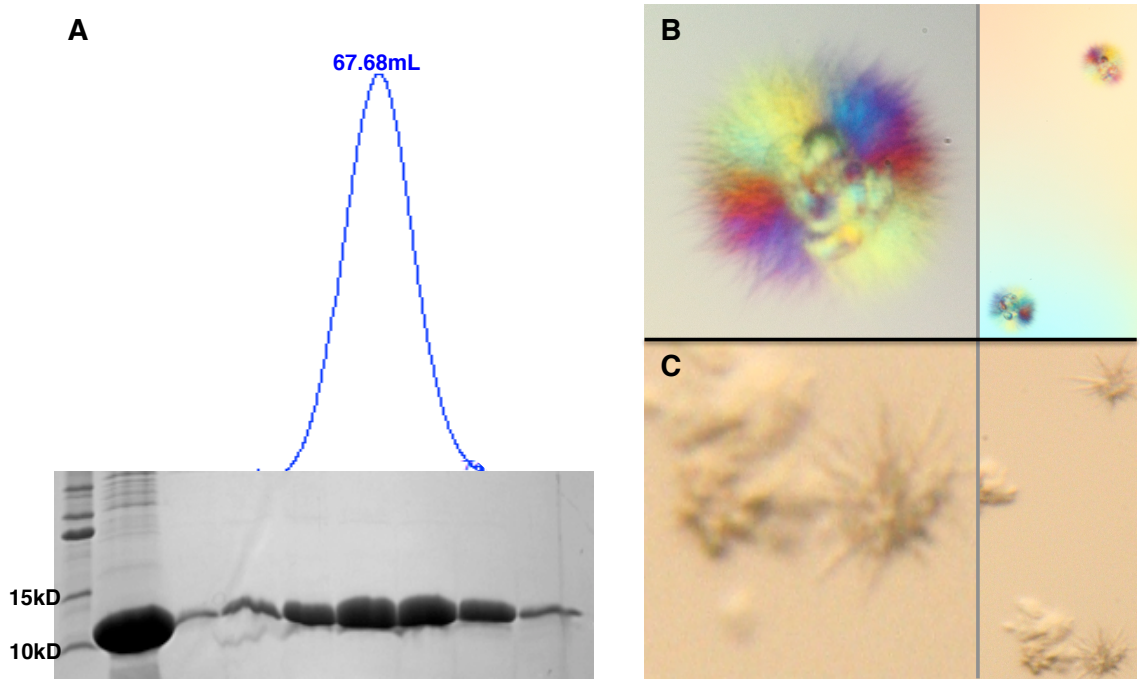


Figure 23. The purification and crystallization of SipC$^{195-278}$. (A) SipC$^{195-278}$ was purified on a 120mL Superdex 75 column (GE Healthcare) and eluted as a trimer at 67.68mL. Lane 1 (far left) contains the molecular weight marker, lane 2 is the column load material, and the remaining lanes are the fractions corresponding to the elution peak. (B) SipC$^{195-278}$ produced a birefringent needle spherulite in 0.2M ammonium citrate, pH 7.0, 20% PEG 3350. (C) Reductively methylated SipC$^{195-278}$ produced thicker, individual needles. The growth condition shown contains 0.1M sodium acetate, pH 4.6, 2M ammonium sulfate.

AI.3.2.3 Binding Partners

As crystallization of the SipC C-terminal domain by itself did not lead to diffraction quality crystals, further work was aimed at purifying and crystallizing SipC in complex with potential C-terminal domain binding partners: ribosylated actin and mouse Exo70. Crystallization with a binding partner often leads to stabilization of the protein of interest and can aid in successful crystallization. The C-terminal domain of SipC is known to polymerize monomeric actin to form actin filaments [30], so we examined the ability of $SipC^{200-330}$, our largest stable SipC C-terminal domain construct, to form a complex with ribosylated actin. Ribosylated actin was produced by using SpvB, an ADP-ribosyltransferase effector protein from *S.* typhimurium [142], to modify monomeric actin with ADP-ribose, thereby preventing actin polymerization and trapping it in its monomeric state. ADP-ribosylation does not cause notable conformational change in actin [142]; therefore, we hypothesized that ribosylated actin could serve as a stable interacting partner with $SipC^{200-320}$.

In order to form a complex, $SipC^{200-330}$ and ribosylated actin were incubated together as described and run over a 120mL Superdex 75 column (GE Healthcare). A stable complex was not formed, as ribosylated actin and $SipC^{200-330}$ do not co-elute on the size exclusion column (Fig. 24). It is possible that complex formation is not favored under the given conditions, or that ribosylated actin and/or the truncated C-terminal domain are not ideal interacting partners.

We also do not know how stable an interaction the SipC C-terminal domain has with monomeric actin. The interaction may only be transient, and as a result, would not be useful in crystallization studies.

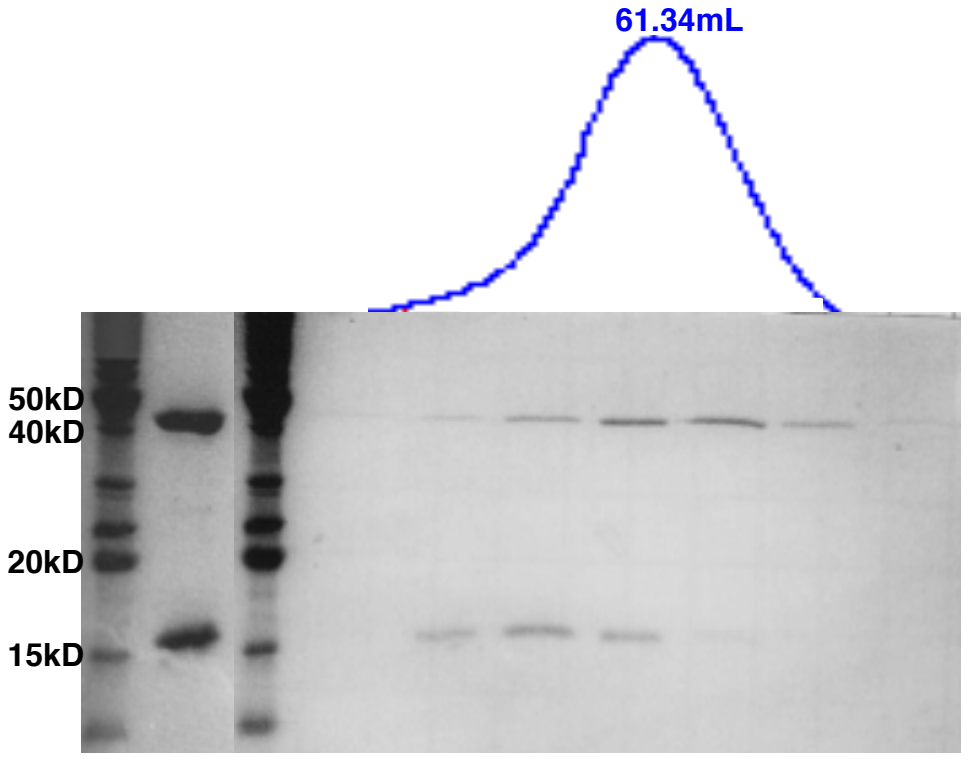


Figure 24. SipC$^{200-330}$ and ribosylated actin do not form a stable complex. SipC$^{200-330}$ was incubated in a 1:1 molar ratio with ribosylated actin for 15 hours at 4°C, and then run over a 120mL Superdex 75 column (GE Healthcare). Ribosylated actin is approximately 45kD and elutes as a monomer, separately from 17kD SipC$^{200-330}$. SipC$^{200-330}$ is not detected as a separate peak on the UV spectrum because it does not contain the aromatic residues necessary to give a substantial UV absorbance; however, a visible shift between SipC$^{200-330}$ and ribosylated actin elution can be detected by SDS-PAGE. . Lane 1 (far left) contains the molecular weight marker, lane 2 is the column load material, Lane 3 is another molecular weight marker, and the remaining lanes are the fractions corresponding to the elution peak.

Nichols, *et al.* reported that SipC$^{200-409}$ interacts with the N-terminal domain of human Exo70 [138]. There are six isoforms of human Exo70, making in-house cloning of the gene problematic. Additionally, the full-length clone was not available commercially, so we elected to work with mouse Exo70, a 653 residue homolog that shares a 96% overall identity with human Exo70 but has an 82 residue deletion in the N-terminal domain. *M. musculus* Exo70 (mExo70) was soluble and stable when expressed with a large N-terminal tag, so experiments were performed with either a 63kD 6xHistidine-NusA tag or a 43kD maltose-binding protein (MBP) tag on mExo70.

Various methods for complex formation were investigated. Purified SipC constructs, 195-400, 200-330, and 200-295, were mixed with purified mExo70 constructs, 1-653 and 85-653, in a 1:1 molar ratio and incubated for 16 hours at 4°C. Stable complex formation was probed by size exclusion chromatography using a Superdex 200, 120mL column (GE Healthcare). SipC and mExo70 did not form a complex, and eluted in individual peaks for all construct combinations analyzed. Co-expression of histidine-tagged SipC$^{200-409}$ and MBP-tagged mExo70 was also unsuccessful in forming a stable complex. By co-expressing SipC$^{200-409}$ with N-terminally 6x-histidine-NusA tagged mExo70$^{85-653}$, we were able to pull-down SipC$^{200-409}$ from cell lysate using NiNTA resin (Qiagen) (Fig. 25). The complex was not stable in the long-term, as SipC and mExo70 eluted

separately when run over a Superdex 200, 120mL column (GE Healthcare). Therefore, crystallization with this complex was not attempted.

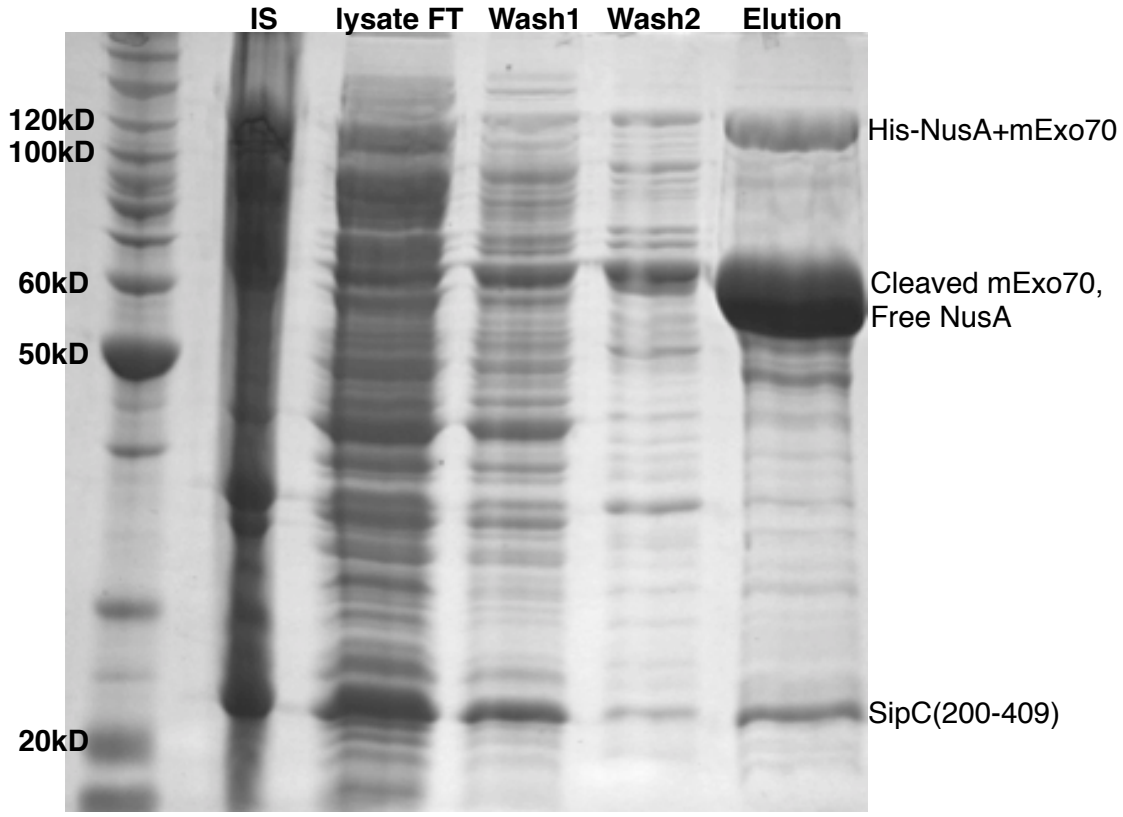


Figure 25. Pull-down of SipC$^{200-409}$ by histidine-NusA-tagged mExo70$^{85-653}$. Histidine-NusA tagged mExo70$^{85-653}$ and SipC$^{200-409}$ were co-expressed and protein was loaded onto a NiNTA column (Qiagen). The insoluble fraction is indicated with IS. The column elution contains His-NusA-mExo70$^{85-653}$ running at roughly 110kD and untagged SipC$^{200-409}$ running at about 23kD.

AI.3.2.4 Actin Polymerization Assay

In order to test our C-terminal domain SipC constructs for their ability to stimulate actin polymerization, we developed an assay based largely on an actin polymerization assay kit from Cytoskeleton, Inc. (Denver, CO) as well as the protocol detailed by Machesky, *et al.* [144]. The assay is performed with pyrene-labeled actin (Cytoskeleton, Inc.), which emits a greater fluorescent signal when it is polymerized into F-actin [145]. Pyrene labeling does not alter the critical concentration of actin or interfere with normal filament formation [133]. We used Arp2/3 (Cytoskeleton, Inc.) in combination with the VCA domain of human Wiskott-Aldrich syndrome protein (WASp) (Cytoskeleton, Inc.) for a positive control of actin polymerization (Fig. 26). The Arp 2/3 complex promotes actin filament nucleation in eukaryotic cells by binding to existing actin filaments and creating branch points for further filament formation [146]. Arp2/3 is able to stimulate rapid actin polymerization when bound to the C-terminal VCA domain conserved among WASp family members [147, 148]. When Arp2/3+VCA was added to pyrene-actin, a dramatic increase in pyrene fluorescence was observed (indicative of actin polymerization), leveling off approximately 10 minutes into the reaction (Fig. 26).

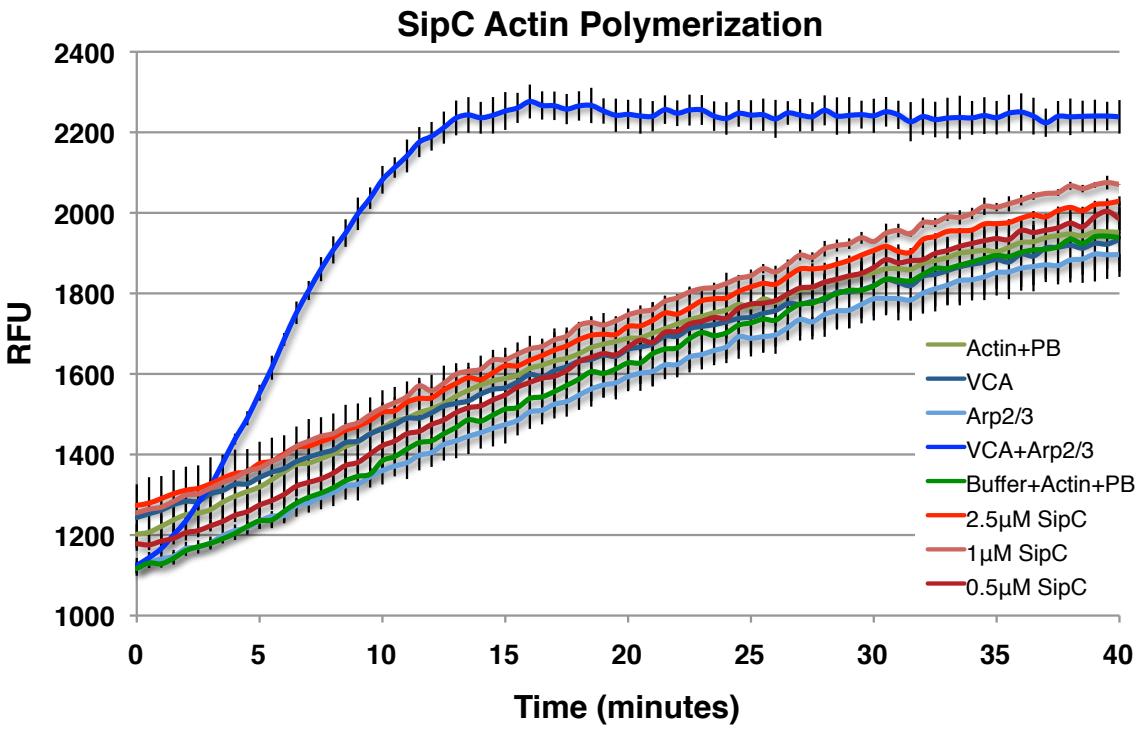


Figure 26. SipC$^{200-409}$ actin polymerization assay. SipC$^{200-409}$ did not show actin polymerization activity above the background; however, the positive control, Arp2/3 + VCA, successfully polymerizes pyrene-actin under the given conditions. Data shown is the average of two assay runs for each component. Standard error of the mean indicated with a black bar for each time point. Reaction conditions detailed in section AI.2.6.

Our SipC$^{200\text{-}409}$ construct did not show actin polymerization activity above that detected for actin alone (Fig. 26). We initially performed this assay under the conditions described by Hayward, *et al.* [30], but when this proved unsuccessful, we sought to identify more ideal reaction conditions. Hayward, *et al*. [30] and McGhie, *et al*. [134] used a 1:10 SipC to actin ratio for their polymerization assays, and the same buffer and assay conditions as are detailed in section AI.2.6. We investigated a range of SipC to actin ratios, screening our SipC$^{200\text{-}409}$ and SipC$^{195\text{-}400}$ constructs for activity with SipC to actin ratios ranging from 1:100 to 5:1. We also examined different buffer and salt conditions, and tried various pyrene-actin concentrations, including combinations of pyrene-actin and unlabeled actin. No activity was detected under any of these reaction conditions. Urea denatured and refolded SipC$^{200\text{-}409}$ was also tested for activity, and also gave a negative result. Therefore, we were not able to show actin nucleation activity with our C-terminal constructs of SipC.

AI.4 Conclusions

SipC plays an integral part in *Salmonella* host entry, through its role as a translocase and through manipulation of the host cytoskeleton via actin nucleation and filament bundling. We sought to define the structure of SipC in order to gain an understanding of how this protein is able to perform its many functions. We investigated the N- and C-terminal effector domains of SipC as crystallization targets separately, and found that the N-terminal domain, $SipC^{1-120}$, could not be purified in its native state, requiring denaturation and refolding in order to remain soluble *in vitro*.

The C-terminal domain proved more amenable to *in vitro* work, and extensive crystallization screening was performed with different constructs of this SipC domain. $SipC^{200-295}$ and $SipC^{195-278}$ led to spherulite and needle-like crystal formation in crystallization trials; however, efforts to optimize these initial hits did not lead to diffraction quality crystals. A fluorescence-based assay was implemented to test for actin nucleation ability of SipC, although C-terminal domain constructs did not show activity in this assay. These results are in contrast to the work published by Hayward, *et al.*, which demonstrated actin nucleation activity of the SipC C-terminal domain using a similar fluorescence-based approach with refolded $SipC^{200-409}$ [30]. Our assays were performed with native $SipC^{200-409}$ as well as denatured/refolded $SipC^{200-409}$ under identical conditions as used by Hayward, *et al.* [30]. We were not able to replicate their

experiment under their described conditions, nor were we able to find more optimal conditions in which to detect actin nucleation activity. One explanation for this discrepancy is that we used our own protocol for refolding SipC$^{200-409}$; however, our refolded material was successfully refolded, as indicated by its elution profile by size exclusion chromatography, so we believe our protocol was adequate for proper protein folding. It is still feasible that there were differences in refolding that might have led to our construct's inactivity, so further experimentation will be necessary in order to rule out this possibility.

Further examination of purification conditions for SipC domains could prove useful for successful crystallization. Despite testing a variety of different conditions, our SipC constructs always behaved as multimers. There may be a yet untested set of conditions that could yield monomeric material, potentially aiding in activity assaying as well as crystallization. Cytokeratin 8 may serve as a better binding partner for SipC$^{200-409}$ and this interaction has yet to be confirmed and probed for crystallization.

APPENDIX II. AVIRULENCE PROTEIN A

AII.1 Introduction

Salmonella Avirulence Protein A (AvrA) is a SPI-1 encoded effector protein that is conditionally expressed in enteritis-related *Salmonella* strains, such as *S.* typhimurium [149, 150]. AvrA is a 34kD protein with a high sequence similarity, and 56% sequence identity, to *Yersinia pseudotuberculosis* YopJ [151]. AvrA does not appear to be a phenocopy of YopJ, as AvrA and YopJ do not perform the same functions in host cells and are not interchangeable [152]. AvrA is predicted to contain a cysteine protease catalytic domain, comprised of His123-Glu142-Cys186, which is characteristic of Clan CE cysteine proteases (Fig. 27) [153]. Members of Clan CE are similar to those of Clan CA in that they both employ the same catalytic mechanism driven by an active cysteine thiol, but they differ in the arrangement of their triad residues. AvrA is grouped into the YopJ-like C55 family within Clan CE, which contains members with sumoyl peptidase or acetyltransferase activity [113, 154].

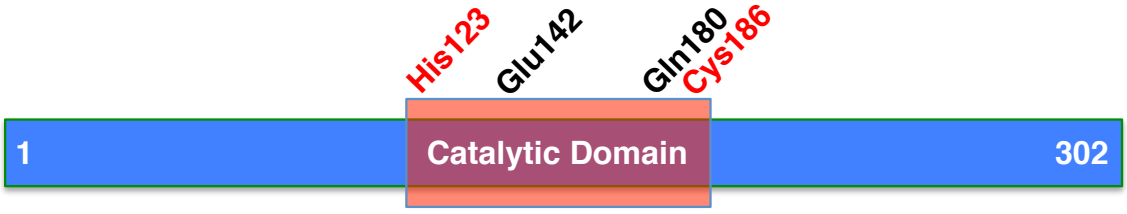


Figure 27. A schematic representation of AvrA and its catalytic domain. AvrA is 302 residues in length and is hypothesized to contain a catalytic triad of His123, Glu142, and Cys186. Gln180 aids in the formation of the oxyanion hole necessary for catalysis.

AvrA's involvement in cell infection remains poorly understood, as both AvrA's catalytic function and targeted host signaling pathways are unclear. AvrA does appear to have anti-apoptotic and anti-inflammatory effects on the host, and potential roles for AvrA include: stabilization of host cell permeability and tight junctions [149]; inhibition of the NF-κB signaling pathway [155]; counterbalance of the SPI-I TTSS effector activated JNK pathway and suppression of cell death [46, 156]. In order to carry out these functions in the host, AvrA has been attributed with having deubiquitinase and acetyltransferase activity.

Deubiquitinases are cysteine proteases that cleave ubiquitin-protein bonds, preventing degradation of the targeted proteins. Potential substrates of AvrA catalyzed deubiquitination are NF-κB, IκBα, β-catenin, and Wnt2 [46, 157, 158]. Acetyltransferases, also commonly cysteine proteases, are responsible for the transfer of actyl groups to specific residues on their target protein. AvrA O-acetylates a key threonine residue located in the activation loop of MAPK kinase 4, inhibiting JNK and subsequently leading to the attenuation of the host inflammatory response [159]. Through acetylation of p53, AvrA activates the p53 pathway to stimulate cell cycle arrest and block apoptosis [160]. Inositol hexakisphosphate, a eukaryotic host cell factor, has been shown to stimulate AvrA's acetyltransferase activity *in vitro* and is believed to serve as an allosteric activator of AvrA [161]. We sought to determine the X-ray structure of AvrA in order to gain a more precise understanding of its role in *Salmonella* infection.

AII.2 Materials and Methods

AII.2.1 Protein Expression and Purification

All AvrA constructs were amplified from *Salmonella Typhimurium* strain SL1344, and inserted into a modified pCDF-Duet-1 vector (Novagen) between the *SalI* and *NotI* restriction sites. The catalytically inactive mutant, $AvrA^{C186A}$, was generated by point mutagenesis. Protein was expressed in *E. coli* BL21(DE3) cells with 1mM IPTG for 16 hours at 21°C. Harvested cells were pelleted, resuspended in 50mM HEPES, pH 7.5, 300mM NaCl, and 5mM imidazole, and lysed via high-pressure homogenization. Cleared lysate was run over Ni-NTA resin (Qiagen) and protein was eluted in 50mM HEPES, pH 7.5, 300mM NaCl, and 500mM Imidazole. Protein was dialyzed against 25mM HEPES, pH 7.5, 250mM NaCl, and 5mM DTT, and the N-terminal histidine tag was cleaved off with 6xHis-rhinovirus 3C protease. The material was passed over Ni-NTA to remove the histidine tag and 3C protease. The final purification step was size exclusion chromatography using a Superdex 75 column with a 120mL resin volume (GE Healthcare). All AvrA constructs were purified into 25mM HEPES, pH 7.5, 250mM NaCl, and 5mM DTT.

ARFGEF2 constructs were cloned out of human cDNA extracted from breast cancer cells (generously provided by L. Fish). PCR products were inserted between *SalI* and *NotI* restriction sites in a modified pGEX4 vector (GE

Healthcare) that was engineered to contain a 6xHistidine tag followed by a rhinovirus 3C protease cleavage site all 3' to the GST tag. For co-expression of AvrA and ARFGEF2, ARFGEF2 was cloned between the *SalI* and *NotI* restriction sites of a modified pCDF-Duet-1 vector (Novagen) and AvrA was inserted between the *NdeI* and *XhoI* restriction sites. Co-expression was also performed with AvrA in the first cloning site between *SalI* and *NotI* and ARFGEF2 in the second site between *NdeI* and *XhoI*. Purification of ARFGEF2 alone and in combination with AvrA was performed as described for AvrA by itself.

AII.2.2 Limited Proteolysis

Limited proteolysis on 31.5µg of $AvrA^{63-302}$ was performed using subtilisin protease (Sigma Aldrich), ranging from 3.25ng to 1.89µg, with 5mM $CaCl_2$. The reaction was allowed to proceed at room temperature for 20 minutes, and was stopped with the addition of 10mM PMSF and SDS loading buffer. Reaction products were separated via 15% SDS-PAGE and were transferred to PVDF membrane (Millipore) for visualization by SYPRO® Ruby protein stain (Sigma Aldrich). Major cleavage product bands were excised from the membrane and sent to the Columbia University Protein Core Facility (New York, NY) for N-terminal Edman degradation sequencing.

AII.2.3 Yeast Two-Hybrid Sample Preparation

AvrA$^{15\text{-}302}$ and catalytically inactive AvrA$^{15\text{-}302,\ C186A}$ were cloned into the pGBT9 vector (Clontech Laboratories, Inc.) between the *EcoRI* and *SalI* restriction sites. The vectors containing the AvrA bait constructs were sent to the Protein Interaction Screening department of the Genomics and Proteomics Core Facility at the German Cancer Research Center (Heidelberg, Germany) for yeast two-hybrid screening. A pre-screen with varying 3-aminotriazole concentrations was performed using both wild-type and the catalytic mutant of AvrA$^{15\text{-}302}$. It was determined that AvrA$^{15\text{-}302,\ C186A}$ could be screened at a higher stringency with positive hits clearly above the background, so the catalytic mutant was used in four independent screens using a human universal normalized cDNA library with 0mM 3-aminotriazole.

AII.3 Results and Discussion

AII.3.1 Domain Determination and Purification

In order to obtain a structure of AvrA, we tested various constructs for their crystallization propensity (Table 5). Full-length AvrA was poorly soluble, but it was it was found that expression of AvrA$^{15-302}$ led to a significantly higher protein yield. Limited proteolysis of AvrA$^{15-302}$ using subtilisin protease identified new constructs with further N-terminal truncations, AvrA$^{27-302}$ and AvrA$^{63-302}$. Of the constructs screened, AvrA$^{63-302}$ provided the greatest yield of 14.2mg of protein per liter of culture grown. Treatment of AvrA$^{63-302}$ with subtilisin protease revealed C-terminal truncations at residues 275, 258, and 244, estimated based on molecular weight of the cleavage products (Fig. 28). Additional C-terminal truncations, at 220 and 214, resulted in insoluble constructs.

Table 5. AvrA constructs produced for crystallization.

Residue	MW (kD)	pI	Soluble	Yield (mg/L)
1-302	33.7	5.4	Y	0.3
15-302	32	5.4	Y	4.7
27-302	31	5.3	Y	7.9
63-302	27	6.2	Y	14.2
63-244	20.5	5.3	Y	10.8
63-220	17.7	4.8	N	N/A
63-214	17.2	4.9	N	N/A

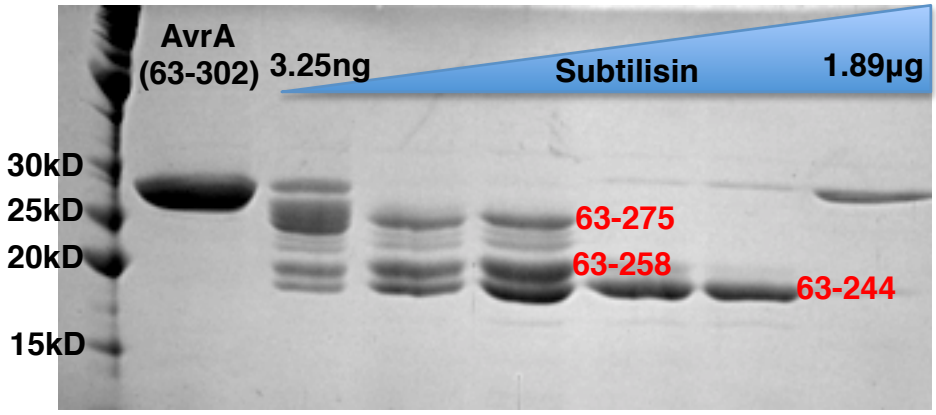


Figure 28. Limited proteolysis of AvrA$^{63-302}$. AvrA$^{63-302}$ was treated with subtilisin protease, ranging from 3.25ng to 1.89µg, for 20 minutes at room temperature. There were three main cleavage products, all of which were the result of C-terminal cleavage. The products correspond to construct of 63-275, 63-258, and 63-244.

AvrA$^{63-244}$ was soluble with an N-terminal histidine tag, and yielded approximately 10.8mg of protein per liter of culture grown. The final step in the purification of this construct is shown in Figure 29. On the Superdex 75, 120mL column under the given conditions, AvrA$^{63-244}$ runs as both a dimer and monomer, eluting at 61.16mL and 70.58mL, respectively. Crystal screens were performed using the monomeric material, but did not lead to crystals. None of the AvrA constructs successfully crystallized; therefore, we attempted to find a binding partner that would aid in crystallization.

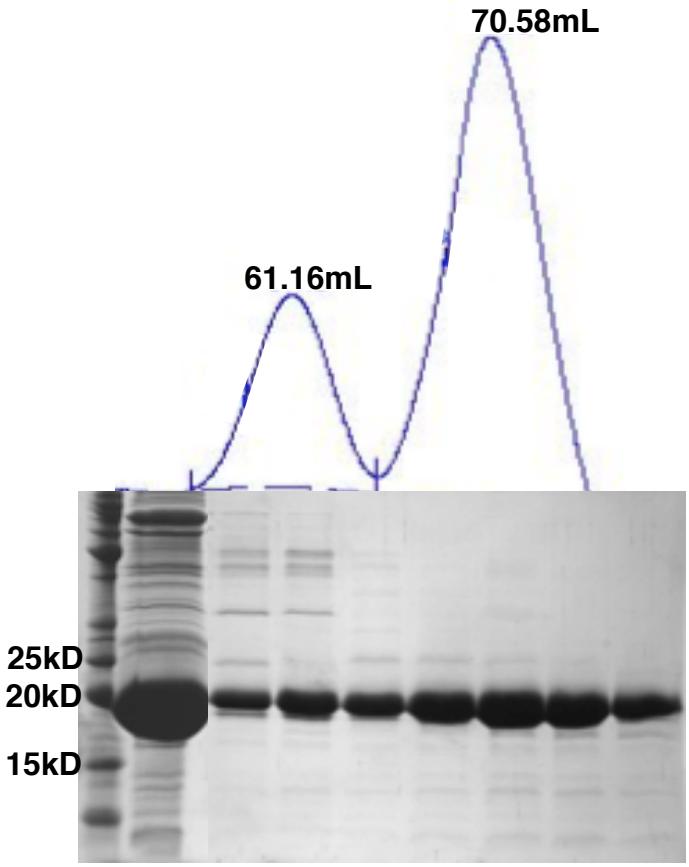


Figure 29. The purification of AvrA$^{63-244}$. AvrA$^{63-244}$ was purified on a Superdex 75, 120mL column (GE Healthcare) into 25mM HEPES, pH 7.5, 250mM NaCl, and 5mM DTT. There are two elution peaks, corresponding to elution as a dimer (61.16mL) and as a monomer (70.58mL). Lane 1 (far left) contains the molecular weight marker, lane 2 is the column load material, and the remaining lanes are the fractions corresponding to the elution peak.

AII.3.2 ARFGEF2: A Potential Binding Partner

To identify a potential binding partner for AvrA, we contracted with the Genomics and Proteomics Core Facility at the German Cancer Research Center

(Heidelberg, Germany) to perform a yeast two-hybrid screen using catalytically inactive AvrA$^{15\text{-}302,\ C186A}$ as bait. A yeast two-hybrid screen is a method for probing the interaction between two proteins. The protein of interest is used as the bait, and is expressed in yeast cells containing a prey plasmid generated from a cDNA library that is derived from the organism or tissue of choice.

Table 6. Yeast two-hybrid results. AvrA$^{15\text{-}302,\ C186A}$ was used as bait in four independent yeast two-hybrid screens against a human universal normalized cDNA library.

Prey	No. of times prey has been isolated	No. of times fragment starts in 5' UTR	No. of times fragment starts in 3' UTR	No. of times fragment starts in coding sequence	Avg. No. of bases of 5' UTR in prey sequence	Prey promiscuity
ARFGEF2	3	0	0	3	0	1
ASAP1-IT1 (non-protein)	2	0	0	2	0	2
ANK3	2	2	0	0	618	5

Four independent screens were executed against a human universal normalized cDNA library at the lowest stringency level possible using 0mM 3-aminotriazole. Three prey proteins were identified as potential interacting partners with the catalytically inactivate AvrA$^{15\text{-}302}$: Brefeldin A-inhibited Guanine Nucleotide Exchange Protein 2 (ARFGEF2); ASAP1 intronic transcript 1 (ASAP1-IT1); ankyrin 3, node of Ranvier (ANK3). The results are summarized in Table 6. ANK3 was isolated two times in the screening process; however the isolated fragments were all located in the 5' untranslated region (UTR). ASAP1-IT1 was

likely an artifact as it is a non-protein coding sequence. ARFGEF2 was the most promising result of this experiment. ARFGEF2 was isolated in three out of the four screens, and in all three instances, the isolated fragment started in the coding sequence of the protein. ARFGEF2 also had a prey promiscuity of 1, which indicates that this bait/prey pair is highly specific.

The isolated fragment of ARFGEF2 corresponds to the C-terminal domain of the protein, spanning residues 1638 to 1786. ARFGEF2 is a large, 200kD guanine nucleotide exchange factor that is required for vesicle trafficking from the trans-Golgi network [162]. ARFGEF2 stimulates the exchange of GDP for GTP to activate ADP-ribosylation factors (ARFs), which then assemble coat proteins around intracellular vesicles, mediating vesicle cycling between the trans-Golgi network, endosome, and plasma membrane [163, 164]. The nucleotide exchange activity of ARFGEF2 is attributed to the presence of a Sec7 domain, correlating to residues 644 to 832 (Fig. 30A). The Sec7 domain was originally identified in yeast and is highly conserved in proteins possessing GEF activity [165]. ARFGEF2 has also been found to associate with recycling endosome-like peripheral vesicles and has been implicated in the coordination of actin cytoskeleton mechanics and membrane traffic in cell migration via an interaction with integrin β1 [166]. The C-terminal domain of ARFGEF2 has not been associated with a specific function; however, it is possible that AvrA's interaction

with this domain could affect ARFGEF2's role in vesicular transport and/or in actin cytoskeleton dynamics to aid *Salmonella* in evading the host defenses.

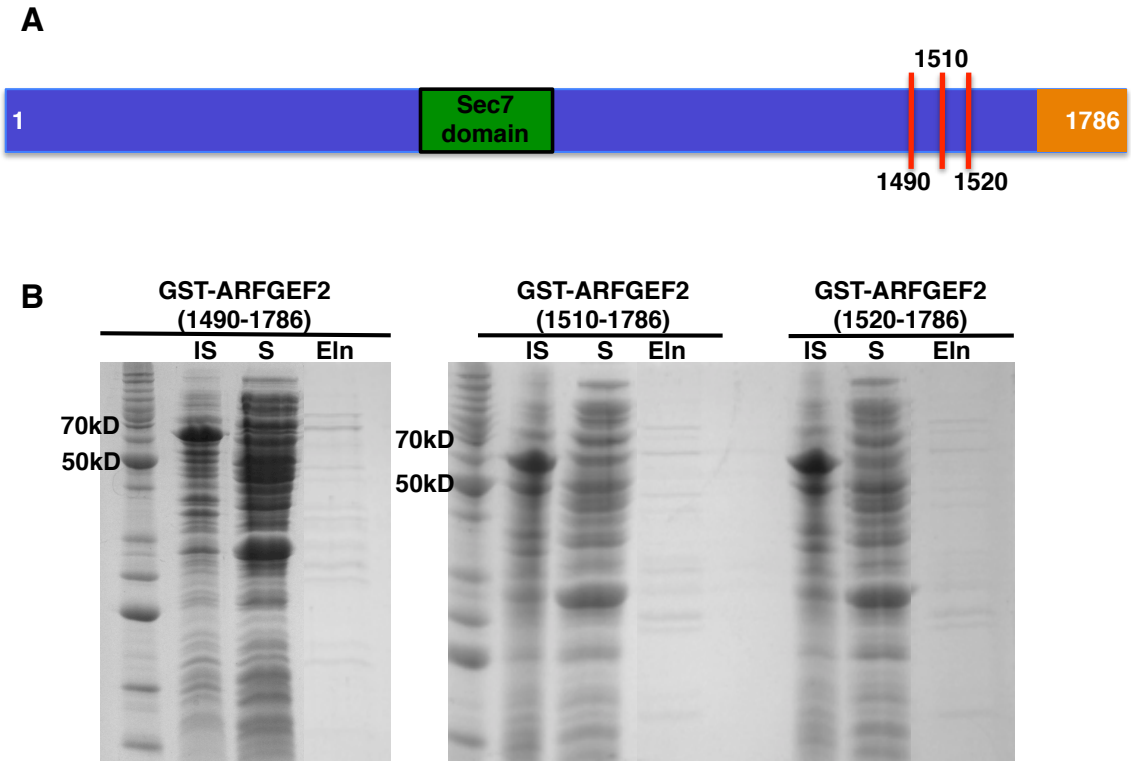


Figure 30. The AvrA interacting domain of ARFGEF2 and its purification. Through a yeast two-hybrid screen, the C-terminal domain of ARFGEF2, residues 1638 to 1786 (orange), was identified as a potential binding partner of AvrA$^{15\text{-}302,\ C186A}$. (A) ARFGEF2 is 1786 residues in length and contains a Sec7 domain. The initial ARFGEF2 constructs designed to probe for AvrA interaction had N-terminal starts at residues 1490, 1510, and 1520, respectively, and all ended at residue 1786. (B) ARFGEF2 constructs were expressed with N-terminal GST-6xHis tags, running at approximately 55-60kD. All constructs were found in the insoluble fraction, which is denoted as IS, the soluble fraction as S, and the Ni-NTA elution as Eln. The first lane on each gel (far left) contains the molecular weight marker.

In order to confirm the interaction between AvrA and the C-terminal domain of ARFGEF2, we designed three ARFGEF2 constructs based on secondary structure predictions generated by Phyre [167]. Constructs started at residue 1490, 1510, and 1520, and all terminated with residue 1786 (Fig. 30A). These constructs were expressed with an N-terminal GST-6xHis tag and purified over a Ni-NTA column (Fig. 30B). All constructs were insoluble.

We also investigated the complex formation between AvrA and ARFGEF2 by co-expressing AvrA$^{15\text{-}302,\ C186A}$ and ARFGEF2 on separate plasmids, pCDF-Duet-1 (Novagen) for AvrA and pGEX4 (GE Healthcare) for ARFGEF2, in the same *E. coli* cells. Expression was seen for AvrA$^{15\text{-}302,\ C186A}$, but not for ARFGEF2. Co-expression from the same vector, pCDF-Duet-1 (Novagen), was also examined, looking at His-tagged ARFGEF2 with untagged AvrA$^{15\text{-}302,\ C186A}$ as well as His-tagged AvrA$^{15\text{-}203,\ C186A}$ with untagged ARFGEF2. His-tagged AvrA$^{15\text{-}302,\ C186A}$ was soluble and could be purified on Ni-NTA resin, while untagged ARFGEF2 remained insoluble.

AII.4 Conclusions

We have defined a minimal domain of AvrA, which spans residues 63 to 244, and contains the predicted catalytic triad of AvrA. Although under the given purification conditions, $AvrA^{63-244}$ was not successful in crystallization trials, more work can be done to refine the purification conditions and identify those in which the protein behaves in a uniform state in solution, either as a dimer or a monomer. Through a yeast two-hybrid screen, we found the C-terminal domain of ARFGEF2, residues 1638 to 1786, to be a potential interacting partner of AvrA. Although co-expression of ARFGEF2 C-terminal domain constructs with AvrA did not yield an AvrA-ARFGEF2 complex, it may prove successful to purify the GST-His-ARFGEF2 constructs under denaturing conditions, refold, and then probe for binding to $AvrA^{15-302,\ C186A}$. The interaction could also be examined through co-immunoprecipitation or an *in vivo* immunofluorescence experiment.

REFERENCES

1. Porwollik S, Boyd EF, Choy C, Cheng P, Florea L, Proctor E, McClelland M (2004) Characterization of Salmonella enterica subspecies I genovars by use of microarrays. *J Bacteriol* **186:** 5883-5898

2. Coburn B, Grassl GA, Finlay BB (2007) Salmonella, the host and disease: a brief review. *Immunol Cell Biol* **85:** 112-118

3. Fierer J, Guiney DG (2001) Diverse virulence traits underlying different clinical outcomes of Salmonella infection. *J Clin Invest* **107:** 775-780

4. Raffatellu M *et al* (2005) Host restriction of Salmonella enterica serotype Typhi is not caused by functional alteration of SipA, SopB, or SopD. *Infect Immun* **73:** 7817-7826

5. Baumler AJ, Tsolis RM, Ficht TA, Adams LG (1998) Evolution of host adaptation in Salmonella enterica. *Infect Immun* **66:** 4579-4587

6. Deng W, Liou SR, Plunkett G, 3rd, Mayhew GF, Rose DJ, Burland V, Kodoyianni V, Schwartz DC, Blattner FR (2003) Comparative genomics of Salmonella enterica serovar Typhi strains Ty2 and CT18. *J Bacteriol* **185:** 2330-2337

7. McClelland M *et al* (2004) Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of Salmonella enterica that cause typhoid. *Nat Genet* **36:** 1268-1274

8. Parkhill J *et al* (2001) Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. *Nature* **413:** 848-852

9. McClelland M *et al* (2001) Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. *Nature* **413:** 852-856

10. Majowicz SE *et al* (2010) The global burden of nontyphoidal Salmonella gastroenteritis. *Clin Infect Dis* **50:** 882-889

11. Crump JA, Mintz ED (2010) Global trends in typhoid and paratyphoid Fever. *Clin Infect Dis* **50:** 241-246

12. Monack DM, Mueller A, Falkow S (2004) Persistent bacterial infections: the interface of the pathogen and the host immune system. *Nat Rev Microbiol* **2:** 747-765

13. Monack DM (2012) Salmonella persistence and transmission strategies. *Curr Opin Microbiol* **15:** 100-107

14. Foley SL, Lynne AM (2008) Food animal-associated Salmonella challenges: pathogenicity and antimicrobial resistance. *J Anim Sci* **86:** E173-187

15. Garai P, Gnanadhas DP, Chakravortty D (2012) Salmonella enterica serovars Typhimurium and Typhi as model organisms: revealing paradigm of host-pathogen interactions. *Virulence* **3:** 377-388

16. Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, Griffin PM, Tauxe RV (1999) Food-related illness and death in the United States. *Emerg Infect Dis* **5:** 607-625

17. Chen HM, Wang Y, Su LH, Chiu CH (2013) Nontyphoid salmonella infection: microbiology, clinical features, and antimicrobial therapy. *Pediatr Neonatol* **54:** 147-152

18. Ramos-Morales F (2012) Impact of Salmonella enterica Type III Secretion System Effectors on the Eukaryotic Host Cell. *ISRN Cell Biology* **2012:** 36

19. Sabbagh SC, Forest CG, Lepage C, Leclerc JM, Daigle F (2010) So similar, yet so different: uncovering distinctive features in the genomes of Salmonella enterica serovars Typhimurium and Typhi. *FEMS Microbiol Lett* **305:** 1-13

20. Hacker J, Blum-Oehler G, Muhldorfer I, Tschape H (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* **23:** 1089-1097

21. Marlovits TC, Kubori T, Sukhan A, Thomas DR, Galan JE, Unger VM (2004) Structural insights into the assembly of the type III secretion needle complex. *Science* **306:** 1040-1042

22. Kosarewicz A, Konigsmaier L, Marlovits TC (2012) The blueprint of the type-3 injectisome. *Philos Trans R Soc Lond B Biol Sci* **367:** 1140-1154

23. Schraidt O, Marlovits TC (2011) Three-dimensional model of Salmonella's needle complex at subnanometer resolution. *Science* **331:** 1192-1195

24. Galan JE, Wolf-Watz H (2006) Protein delivery into eukaryotic cells by type III secretion machines. *Nature* **444:** 567-573

25. Scherer CA, Cooper E, Miller SI (2000) The Salmonella type III secretion translocon protein SspC is inserted into the epithelial cell plasma membrane upon infection. *Mol Microbiol* **37:** 1133-1145

26. Edgren T, Forsberg A, Rosqvist R, Wolf-Watz H (2012) Type III secretion in Yersinia: injectisome or not? *PLoS Pathog* **8:** e1002669

27. Akopyan K, Edgren T, Wang-Edgren H, Rosqvist R, Fahlgren A, Wolf-Watz H, Fallman M (2011) Translocation of surface-localized effectors in type III secretion. *Proc Natl Acad Sci U S A* **108:** 1639-1644

28. Ivanov MI, Noel BL, Rampersaud R, Mena P, Benach JL, Bliska JB (2008) Vaccination of mice with a Yop translocon complex elicits antibodies that are protective against infection with F1- Yersinia pestis. *Infect Immun* **76:** 5181-5190

29. Rosqvist R, Forsberg A, Wolf-Watz H (1991) Intracellular targeting of the Yersinia YopE cytotoxin in mammalian cells induces actin microfilament disruption. *Infect Immun* **59:** 4562-4569

30. Hayward RD, Koronakis V (1999) Direct nucleation and bundling of actin by the SipC protein of invasive Salmonella. *EMBO J* **18:** 4926-4934

31. Menard R, Sansonetti P, Parsot C, Vasselon T (1994) Extracellular association and cytoplasmic partitioning of the IpaB and IpaC invasins of S. flexneri. *Cell* **79:** 515-525

32. Galan JE, Collmer A (1999) Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284:** 1322-1328

33. Galan JE (2001) Salmonella interactions with host cells: type III secretion at work. *Annu Rev Cell Dev Biol* **17:** 53-86

34. Cirillo DM, Valdivia RH, Monack DM, Falkow S (1998) Macrophage-dependent induction of the Salmonella pathogenicity island 2 type III secretion system and its role in intracellular survival. *Mol Microbiol* **30:** 175-188

35. Steele-Mortimer O (2008) The Salmonella-containing vacuole: moving with the times. *Curr Opin Microbiol* **11:** 38-45

36. Knodler LA, Steele-Mortimer O (2003) Taking possession: biogenesis of the Salmonella-containing vacuole. *Traffic* **4:** 587-599

37. Steele-Mortimer O, Brumell JH, Knodler LA, Meresse S, Lopez A, Finlay BB (2002) The invasion-associated type III secretion system of Salmonella enterica serovar Typhimurium is necessary for intracellular proliferation and vacuole biogenesis in epithelial cells. *Cell Microbiol* **4:** 43-54

38. Steele-Mortimer O, Meresse S, Gorvel JP, Toh BH, Finlay BB (1999) Biogenesis of Salmonella typhimurium-containing vacuoles in epithelial cells involves interactions with the early endocytic pathway. *Cell Microbiol* **1:** 33-49

39. Smith AC, Cirulis JT, Casanova JE, Scidmore MA, Brumell JH (2005) Interaction of the Salmonella-containing vacuole with the endocytic recycling system. *J Biol Chem* **280:** 24634-24641

40. Drecktrah D, Knodler LA, Ireland R, Steele-Mortimer O (2006) The mechanism of Salmonella entry determines the vacuolar environment and intracellular gene expression. *Traffic* **7:** 39-51

41. Rathman M, Sjaastad MD, Falkow S (1996) Acidification of phagosomes containing Salmonella typhimurium in murine macrophages. *Infect Immun* **64:** 2765-2773

42. Salcedo SP, Holden DW (2003) SseG, a virulence protein that targets Salmonella to the Golgi network. *EMBO J* **22:** 5003-5014

43. Garcia-del Portillo F, Zwick MB, Leung KY, Finlay BB (1993) Salmonella induces the formation of filamentous structures containing lysosomal membrane glycoproteins in epithelial cells. *Proc Natl Acad Sci U S A* **90:** 10544-10548

44. Garcia-del Portillo F, Zwick MB, Leung KY, Finlay BB (1993) Intracellular replication of Salmonella within epithelial cells is associated with filamentous structures containing lysosomal membrane glycoproteins. *Infect Agents Dis* **2:** 227-231

45. Fink SL, Cookson BT (2007) Pyroptosis and host cell death responses during Salmonella infection. *Cell Microbiol* **9:** 2562-2570

46. Wu H, Jones RM, Neish AS (2012) The Salmonella effector AvrA mediates bacterial intracellular survival during infection in vivo. *Cell Microbiol* **14:** 28-39

47. Kim JM, Eckmann L, Savidge TC, Lowe DC, Witthoft T, Kagnoff MF (1998) Apoptosis of human intestinal epithelial cells after bacterial invasion. *J Clin Invest* **102:** 1815-1823

48. Brennan MA, Cookson BT (2000) Salmonella induces macrophage death by caspase-1-dependent necrosis. *Mol Microbiol* **38:** 31-40

49. Monack DM, Navarre WW, Falkow S (2001) Salmonella-induced macrophage death: the role of caspase-1 in death and inflammation. *Microbes Infect* **3:** 1201-1212

50. van der Velden AW, Lindgren SW, Worley MJ, Heffron F (2000) Salmonella pathogenicity island 1-independent induction of apoptosis in infected macrophages by Salmonella enterica serotype typhimurium. *Infect Immun* **68:** 5702-5709

51. Everest P, Wain J, Roberts M, Rook G, Dougan G (2001) The molecular mechanisms of severe typhoid fever. *Trends Microbiol* **9:** 316-320

52. Raffatellu M, Chessa D, Wilson RP, Dusold R, Rubino S, Baumler AJ (2005) The Vi capsular antigen of Salmonella enterica serotype Typhi reduces Toll-like receptor-dependent interleukin-8 expression in the intestinal mucosa. *Infect Immun* **73:** 3367-3374

53. Weinstein DL, O'Neill BL, Hone DM, Metcalf ES (1998) Differential early interactions between Salmonella enterica serovar Typhi and two other pathogenic Salmonella serovars with intestinal epithelial cells. *Infect Immun* **66:** 2310-2318

54. Merrell DS, Falkow S (2004) Frontal and stealth attack strategies in microbial pathogenesis. *Nature* **430:** 250-256

55. Haghjoo E, Galan JE (2004) Salmonella typhi encodes a functional cytolethal distending toxin that is delivered into host cells by a bacterial-internalization pathway. *Proc Natl Acad Sci U S A* **101:** 4614-4619

56. Spano S, Liu X, Galan JE (2011) Proteolytic targeting of Rab29 by an effector protein distinguishes the intracellular compartments of human-adapted and broad-host Salmonella. *Proc Natl Acad Sci U S A* **108:** 18418-18423

57. Jinadasa RN, Bloom SE, Weiss RS, Duhamel GE (2011) Cytolethal distending toxin: a conserved bacterial genotoxin that blocks cell cycle progression, leading to apoptosis of a broad range of mammalian cell lineages. *Microbiology* **157:** 1851-1875

58. Spano S, Ugalde JE, Galan JE (2008) Delivery of a Salmonella Typhi exotoxin from a host intracellular compartment. *Cell Host Microbe* **3:** 30-38

59. Song J, Gao X, Galan JE (2013) Structure and function of the Salmonella Typhi chimaeric A(2)B(5) typhoid toxin. *Nature* **499:** 350-354

60. Spano S, Galan JE (2012) A Rab32-dependent pathway contributes to Salmonella typhi host restriction. *Science* **338:** 960-963

61. Ho TD, Figueroa-Bossi N, Wang M, Uzzau S, Bossi L, Slauch JM (2002) Identification of GtgE, a novel virulence factor encoded on the Gifsy-2 bacteriophage of Salmonella enterica serovar Typhimurium. *J Bacteriol* **184:** 5234-5239

62. Soding J (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21:** 951-960

63. Ebine K, Ueda T (2009) Unique mechanism of plant endocytic/vacuolar transport pathways. *J Plant Res* **122:** 21-30

64. Stenmark H, Olkkonen VM (2001) The Rab GTPase family. *Genome Biol* **2:** REVIEWS3007

65. Oesterlin LK, Goody RS, Itzen A (2012) Posttranslational modifications of Rab proteins cause effective displacement of GDP dissociation inhibitor. *Proc Natl Acad Sci U S A* **109:** 5621-5626

66. Schwartz SL, Cao C, Pylypenko O, Rak A, Wandinger-Ness A (2007) Rab GTPases at a glance. *J Cell Sci* **120:** 3905-3910

67. Goody RS, Rak A, Alexandrov K (2005) The structural and mechanistic basis for recycling of Rab proteins between membrane compartments. *Cell Mol Life Sci* **62:** 1657-1670

68. Geyer M, Wittinghofer A (1997) GEFs, GAPs, GDIs and effectors: taking a closer (3D) look at the regulation of Ras-related GTP-binding proteins. *Curr Opin Struct Biol* **7:** 786-792

69. Gideon P, John J, Frech M, Lautwein A, Clark R, Scheffler JE, Wittinghofer A (1992) Mutational and kinetic analyses of the GTPase-activating protein (GAP)-p21 interaction: the C-terminal domain of GAP is not sufficient for full activity. *Mol Cell Biol* **12:** 2050-2056

70. Pfeffer SR (2005) Structural clues to Rab GTPase functional diversity. *J Biol Chem* **280:** 15485-15488

71. Simon I, Zerial M, Goody RS (1996) Kinetics of interaction of Rab5 and Rab7 with nucleotides and magnesium ions. *J Biol Chem* **271:** 20470-20478

72. Dumas JJ, Zhu Z, Connolly JL, Lambright DG (1999) Structural basis of activation and GTP hydrolysis in Rab proteins. *Structure* **7:** 413-423

73. Stroupe C, Brunger AT (2000) Crystal structures of a Rab protein in its inactive and active conformations. *J Mol Biol* **304:** 585-598

74. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302:** 205-217

75. Wasmeier C, Romao M, Plowright L, Bennett DC, Raposo G, Seabra MC (2006) Rab38 and Rab32 control post-Golgi trafficking of melanogenic enzymes. *J Cell Biol* **175:** 271-281

76. Bultema JJ, Ambrosio AL, Burek CL, Di Pietro SM (2012) BLOC-2, AP-3, and AP-1 proteins function in concert with Rab38 and Rab32 proteins to mediate protein trafficking to lysosome-related organelles. *J Biol Chem* **287:** 19550-19563

77. Dell'Angelica EC, Mullins C, Caplan S, Bonifacino JS (2000) Lysosome-related organelles. *FASEB J* **14:** 1265-1278

78. Raposo G, Marks MS (2007) Melanosomes--dark organelles enlighten endosomal membrane transport. *Nat Rev Mol Cell Biol* **8:** 786-797

79. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A* **276:** 307-326

80. Adams PD *et al* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66:** 213-221

81. Terwilliger TC, Adams PD, Read RJ, Mccoy AJ, Moriarty NW, Grosse-Kunstleve RW, Afonine PV, Zwart PH, Hung LW (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallographica Section D-Biological Crystallography* **65:** 582-601

82. Mccoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *Journal of Applied Crystallography* **40:** 658-674

83. Grosse-Kunstleve RW, Adams PD (2003) Substructure search procedures for macromolecular structures. *Acta Crystallographica Section D-Biological Crystallography* **59:** 1966-1973

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

85. Winn MD *et al* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67:** 235-242

86. Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, Murshudov GN (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallographica Section D-Biological Crystallography* **60:** 2184-2195

87. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography* **66:** 486-501

88. Winn MD, Murshudov GN, Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Macromolecular Crystallography, Pt D* **374:** 300-321

89. Painter J, Merritt EA (2006) TLSMD web server for the generation of multi-group TLS models. *Journal of Applied Crystallography* **39:** 109-111

90. McNicholas S, Potterton E, Wilson KS, Noble ME (2011) Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr* **67:** 386-394

91. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9:** 671-675

92. Boggon TJ, Shapiro L (2000) Screening for phasing atoms in protein crystallography. *Structure* **8:** R143-149

93. Hendrickson WA, Ogata, C.M. (1997) Phase determination from multiwavelength anomalous diffraction measurements. *Methods Enzymol* **276:** 494-523

94. Laskowski RA (2009) PDBsum new things. *Nucleic Acids Res* **37:** D355-359

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

96. Holm L, Kääriäinen, S., Rosenström, P., Schenkel, A. (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics* **24:** 2780-2781

97. Rawlings ND, Barrett AJ (1994) Families of cysteine peptidases. *Methods Enzymol* **244:** 461-486

98. Rawlings ND, Barrett AJ, Bateman A (2012) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **40:** D343-350

99. Dirix G, Monsieurs P, Dombrecht B, Daniels R, Marchal K, Vanderleyden J, Michiels J (2004) Peptide signal molecules and bacteriocins in Gram-negative bacteria: a genome-wide in silico screening for peptides containing a double-glycine leader sequence and their cognate transporters. *Peptides* **25:** 1425-1440

100. Umezawa H (1982) Low-Molecular-Weight Enzyme-Inhibitors of Microbial Origin. *Annual Review of Microbiology* **36:** 75-99

101. Krengel U, Schlichting I, Scherer A, Schumann R, Frech M, John J, Kabsch W, Pai EF, Wittinghofer A (1990) Three-dimensional structures of H-ras p21 mutants: molecular basis for their inability to function as signal switch molecules. *Cell* **62:** 539-548

102. Scheffzek K, Ahmadian MR, Kabsch W, Wiesmuller L, Lautwein A, Schmitz F, Wittinghofer A (1997) The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* **277:** 333-338

103. Scita G, Tenca P, Frittoli E, Tocchetti A, Innocenti M, Giardina G, Di Fiore PP (2000) Signaling from Ras to Rac and beyond: not just a matter of GEFs. *EMBO J* **19:** 2393-2398

104. Ridley A (2000) Rho GTPases. Integrating integrin signaling. *J Cell Biol* **150:** F107-109

105. Feig LA (1999) Tools of the trade: use of dominant-inhibitory mutants of Ras-family GTPases. *Nat Cell Biol* **1:** E25-27

106. Hsu Y, Jubelin G, Taieb F, Nougayrede JP, Oswald E, Stebbins CE (2008) Structure of the cyclomodulin Cif from pathogenic Escherichia coli. *J Mol Biol* **384:** 465-477

107. Cui J *et al* (2010) Glutamine deamidation and dysfunction of ubiquitin/NEDD8 induced by a bacterial effector family. *Science* **329:** 1215-1218

108. Merlet J, Burger J, Gomes JE, Pintard L (2009) Regulation of cullin-RING E3 ubiquitin-ligases by neddylation and dimerization. *Cell Mol Life Sci* **66:** 1924-1938

109. Rabut G, Peter M (2008) Function and regulation of protein neddylation. 'Protein modifications: beyond the usual suspects' review series. *EMBO Rep* **9:** 969-976

110. Saha A, Deshaies RJ (2008) Multimodal activation of the ubiquitin ligase SCF by Nedd8 conjugation. *Mol Cell* **32:** 21-31

111. Crow A, Hughes RK, Taieb F, Oswald E, Banfield MJ (2012) The molecular basis of ubiquitin-like protein NEDD8 deamidation by the bacterial effector protein Cif. *Proc Natl Acad Sci U S A* **109:** E1830-1838

112. Taieb F, Nougayrede JP, Oswald E (2011) Cycle inhibiting factors (cifs): cyclomodulins that usurp the ubiquitin-dependent degradation pathway of host cells. *Toxins (Basel)* **3:** 356-368

113. Mukherjee S, Hao YH, Orth K (2007) A newly discovered post-translational modification--the acetylation of serine and threonine residues. *Trends Biochem Sci* **32:** 210-216

114. Ding J, McGrath WJ, Sweet RM, Mangel WF (1996) Crystal structure of the human adenovirus proteinase with its 11 amino acid cofactor. *EMBO J* **15:** 1778-1783

115. Paquette N *et al* (2012) Serine/threonine acetylation of TGFbeta-activated kinase (TAK1) by Yersinia pestis YopJ inhibits innate immune signaling. *Proc Natl Acad Sci U S A* **109:** 12710-12715

116. Orth K (2002) Function of the Yersinia effector YopJ. *Curr Opin Microbiol* **5:** 38-43

117. Viboud GI, Bliska JB (2005) Yersinia outer proteins: role in modulation of host cell signaling responses and pathogenesis. *Annu Rev Microbiol* **59:** 69-89

118. Mesquita FS, Thomas M, Sachse M, Santos AJ, Figueira R, Holden DW (2012) The Salmonella deubiquitinase SseL inhibits selective autophagy of cytosolic aggregates. *PLoS Pathog* **8:** e1002743

119. Rytkonen A, Poh J, Garmendia J, Boyle C, Thompson A, Liu M, Freemont P, Hinton JC, Holden DW (2007) SseL, a Salmonella deubiquitinase required for macrophage killing and virulence. *Proc Natl Acad Sci U S A* **104:** 3502-3507

120. Figueira R, Holden DW (2012) Functions of the Salmonella pathogenicity island 2 (SPI-2) type III secretion system effectors. *Microbiology* **158:** 1147-1161

121. Le Negrate G *et al* (2008) Salmonella secreted factor L deubiquitinase of Salmonella typhimurium inhibits NF-kappaB, suppresses IkappaBalpha ubiquitination and modulates innate immune responses. *J Immunol* **180:** 5045-5056

122. Collazo CM, Galan JE (1997) The invasion-associated type III system of Salmonella typhimurium directs the translocation of Sip proteins into the host cell. *Mol Microbiol* **24:** 747-756

123. Darwin KH, Miller VL (2001) Type III secretion chaperone-dependent regulation: activation of virulence genes by SicA and InvF in Salmonella typhimurium. *EMBO J* **20:** 1850-1862

124. Tucker SC, Galan JE (2000) Complex function for SicA, a Salmonella enterica serovar typhimurium type III secretion-associated chaperone. *J Bacteriol* **182:** 2262-2268

125. Hueck CJ, Hantman MJ, Bajaj V, Johnston C, Lee CA, Miller SI (1995) Salmonella typhimurium secreted invasion determinants are homologous to Shigella Ipa proteins. *Mol Microbiol* **18:** 479-490

126. Kaniga K, Tucker S, Trollinger D, Galan JE (1995) Homologs of the Shigella IpaB and IpaC invasins are required for Salmonella typhimurium entry into cultured epithelial cells. *J Bacteriol* **177:** 3965-3971

127. Buttner D, Bonas U (2002) Port of entry--the type III secretion translocon. *Trends Microbiol* **10:** 186-192

128. Osiecki JC, Barker J, Picking WL, Serfis AB, Berring E, Shah S, Harrington A, Picking WD (2001) IpaC from Shigella and SipC from Salmonella possess similar biochemical properties but are functionally distinct. *Mol Microbiol* **42:** 469-481

129. Myeni SK, Wang L, Zhou D (2013) SipB-SipC complex is essential for translocon formation. *PLoS One* **8:** e60499

130. McGhie EJ, Hayward RD, Koronakis V (2004) Control of actin turnover by a salmonella invasion protein. *Mol Cell* **13:** 497-510

131. Furukawa R, Fechheimer M (1997) The structure, function, and assembly of actin filament bundles. *Int Rev Cytol* **175:** 29-90

132. May RC, Machesky LM (2001) Phagocytosis and the actin cytoskeleton. *J Cell Sci* **114:** 1061-1077

133. Tobacman LS, Korn ED (1983) The kinetics of actin nucleation and polymerization. *J Biol Chem* **258:** 3207-3214

134. McGhie EJ, Hayward RD, Koronakis V (2001) Cooperation between actin-binding proteins of invasive Salmonella: SipA potentiates SipC nucleation and bundling of actin. *EMBO J* **20:** 2131-2139

135. Chang J, Myeni SK, Lin TL, Wu CC, Staiger CJ, Zhou D (2007) SipC multimerization promotes actin nucleation and contributes to Salmonella-induced inflammation. *Mol Microbiol* **66:** 1548-1556

136. Chang J, Chen J, Zhou D (2005) Delineation and characterization of the actin nucleation and effector translocation activities of Salmonella SipC. *Mol Microbiol* **55:** 1379-1389

137. Gonias SL, Hembrough TA, Sankovic M (2001) Cytokeratin 8 functions as a major plasminogen receptor in select epithelial and carcinoma cells. *Front Biosci* **6:** D1403-1411

138. Nichols CD, Casanova JE (2010) Salmonella-directed recruitment of new membrane to invasion foci via the host exocyst complex. *Curr Biol* **20:** 1316-1320

139. Moore BA, Robinson HH, Xu Z (2007) The crystal structure of mouse Exo70 reveals unique features of the mammalian exocyst. *J Mol Biol* **371:** 410-421

140. Wu H, Rossi G, Brennwald P (2008) The ghost in the machine: small GTPases as spatial regulators of exocytosis. *Trends Cell Biol* **18:** 397-404

141. Zuo X, Zhang J, Zhang Y, Hsu SC, Zhou D, Guo W (2006) Exo70 interacts with the Arp2/3 complex and regulates cell migration. *Nat Cell Biol* **8:** 1383-1388

142. Margarit SM, Davidson W, Frego L, Stebbins CE (2006) A steric antagonism of actin polymerization by a salmonella virulence protein. *Structure* **14:** 1219-1229

143. Rayment I (1997) Reductive alkylation of lysine residues to alter crystallization properties of proteins. *Methods Enzymol* **276:** 171-179

144. Machesky LM, Mullins RD, Higgs HN, Kaiser DA, Blanchoin L, May RC, Hall ME, Pollard TD (1999) Scar, a WASp-related protein, activates nucleation of actin filaments by the Arp2/3 complex. *Proc Natl Acad Sci U S A* **96:** 3739-3744

145. Kouyama T, Mihashi K (1981) Fluorimetry study of N-(1-pyrenyl)iodoacetamide-labelled F-actin. Local structural change of actin protomer both on polymerization and on binding of heavy meromyosin. *Eur J Biochem* **114:** 33-38

146. dos Remedios CG, Chhabra D, Kekic M, Dedova IV, Tsubakihara M, Berry DA, Nosworthy NJ (2003) Actin binding proteins: regulation of cytoskeletal microfilaments. *Physiol Rev* **83:** 433-473

147. Condeelis J (2001) How is actin polymerization nucleated in vivo? *Trends Cell Biol* **11:** 288-293

148. Higgs HN, Pollard TD (1999) Regulation of actin polymerization by Arp2/3 complex and WASp/Scar proteins. *J Biol Chem* **274:** 32531-32534

149. Liao AP, Petrof EO, Kuppireddi S, Zhao Y, Xia Y, Claud EC, Sun J (2008) Salmonella type III effector AvrA stabilizes cell tight junctions to inhibit inflammation in intestinal epithelial cells. *PLoS One* **3:** e2369

150. Streckel W, Wolff AC, Prager R, Tietze E, Tschape H (2004) Expression profiles of effector proteins SopB, SopD1, SopE1, and AvrA differ with systemic, enteric, and epidemic strains of Salmonella enterica. *Mol Nutr Food Res* **48:** 496-503

151. Rytkonen A, Holden DW (2007) Bacterial interference of ubiquitination and deubiquitination. *Cell Host Microbe* **1:** 13-22

152. Schesser K, Dukuzumuremyi JM, Cilio C, Borg S, Wallis TS, Pettersson S, Galyov EE (2000) The Salmonella YopJ-homologue AvrA does not possess YopJ-like activity. *Microb Pathog* **28:** 59-70

153. Orth K, Xu Z, Mudgett MB, Bao ZQ, Palmer LE, Bliska JB, Mangel WF, Staskawicz B, Dixon JE (2000) Disruption of signaling by Yersinia effector YopJ, a ubiquitin-like protein protease. *Science* **290:** 1594-1597

154. Roden J, Eardley L, Hotson A, Cao Y, Mudgett MB (2004) Characterization of the Xanthomonas AvrXv4 effector, a SUMO protease translocated into plant cells. *Mol Plant Microbe Interact* **17:** 633-643

155. Collier-Hyams LS, Zeng H, Sun J, Tomlinson AD, Bao ZQ, Chen H, Madara JL, Orth K, Neish AS (2002) Cutting edge: Salmonella AvrA effector inhibits the key proinflammatory, anti-apoptotic NF-kappa B pathway. *J Immunol* **169:** 2846-2850

156. Du F, Galan JE (2009) Selective inhibition of type III secretion activated signaling by the Salmonella effector AvrA. *PLoS Pathog* **5:** e1000595

157. Ye Z, Petrof EO, Boone D, Claud EC, Sun J (2007) Salmonella effector AvrA regulation of colonic epithelial cell inflammation by deubiquitination. *Am J Pathol* **171:** 882-892

158. Sun J, Hobert ME, Rao AS, Neish AS, Madara JL (2004) Bacterial activation of beta-catenin signaling in human epithelia. *Am J Physiol Gastrointest Liver Physiol* **287:** G220-227

159. Jones RM, Wu H, Wentworth C, Luo L, Collier-Hyams L, Neish AS (2008) Salmonella AvrA Coordinates Suppression of Host Immune and Apoptotic Defenses via JNK Pathway Blockade. *Cell Host Microbe* **3:** 233-244

160. Wu S, Ye Z, Liu X, Zhao Y, Xia Y, Steiner A, Petrof EO, Claud EC, Sun J (2010) Salmonella typhimurium infection increases p53 acetylation in intestinal epithelial cells. *Am J Physiol Gastrointest Liver Physiol* **298:** G784-794

161. Mittal R, Peak-Chew SY, Sade RS, Vallis Y, McMahon HT (2010) The acetyltransferase activity of the bacterial toxin YopJ of Yersinia is activated by eukaryotic host cell inositol hexakisphosphate. *J Biol Chem* **285:** 19927-19934

162. Sheen VL *et al* (2004) Mutations in ARFGEF2 implicate vesicle trafficking in neural progenitor proliferation and migration in the human cerebral cortex. *Nat Genet* **36:** 69-76

163. Pacheco-Rodriguez G, Moss J, Vaughan M (2002) BIG1 and BIG2: brefeldin A-inhibited guanine nucleotide-exchange proteins for ADP-ribosylation factors. *Methods Enzymol* **345:** 397-404

164. Yamaji R, Adamik R, Takeda K, Togawa A, Pacheco-Rodriguez G, Ferrans VJ, Moss J, Vaughan M (2000) Identification and localization of two brefeldin A-inhibited guanine nucleotide-exchange proteins for ADP-ribosylation factors in a macromolecular complex. *Proc Natl Acad Sci U S A* **97:** 2567-2572

165. Mossessova E, Gulbis JM, Goldberg J (1998) Structure of the guanine nucleotide exchange factor Sec7 domain of human arno and analysis of the interaction with ARF GTPase. *Cell* **92:** 415-423

166. Shen X, Li CC, Aponte AM, Shen RF, Billings EM, Moss J, Vaughan M (2012) Brefeldin A-inhibited ADP-ribosylation factor activator BIG2 regulates cell migration via integrin beta1 cycling and actin remodeling. *Proc Natl Acad Sci U S A* **109:** 14464-14469

167. Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4:** 363-371