Rockefeller University
Digital Commons @ RU

Student Theses and Dissertations

2010

Structural and Functional Characterization of B. Anthracis Udp-Glcnac 2-Epimerase and Salmonella Secreted Effector I

Shyam Bhaskaran

Follow this and additional works at: http://digitalcommons.rockefeller.edu/student_theses_and_dissertations

 Part of the Life Sciences Commons

Recommended Citation

Bhaskaran, Shyam, "Structural and Functional Characterization of B. Anthracis Udp-Glcnac 2-Epimerase and Salmonella Secreted Effector I" (2010). *Student Theses and Dissertations.* Paper 87.

This Thesis is brought to you for free and open access by Digital Commons @ RU. It has been accepted for inclusion in Student Theses and Dissertations by an authorized administrator of Digital Commons @ RU. For more information, please contact mcsweej@mail.rockefeller.edu.

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF *B. ANTHRACIS* UDP-GLCNAC 2-EPIMERASE AND *SALMONELLA* SECRETED EFFECTOR I

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Shyam S. Bhaskaran

June 2010

© Copyright by Shyam S. Bhaskaran, 2010

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF *B. ANTHRACIS* UDP-GLCNAC 2-EPIMERASE AND *SALMONELLA* SECRETED EFFECTOR I

Shyam S. Bhaskaran, Ph.D.

The Rockefeller University 2010

Structural biology has been an exceptionally powerful tool in the study of bacterial pathogenesis, not only in elucidating the molecular basis of disease as the first step towards the design of therapeutics but also in many cases leading to insights into host-pathogen biology that would not have been possible otherwise (Remaut & Waksman 2004). In this work such an approach has been taken towards the study of the *Bacillus anthracis* cell-wall enzyme UDP-GlcNAc 2-Epimerase and the *Salmonella* secreted virulence effector SseI.

UDP-GlcNAc 2-Epimerase is an essential *B. anthracis* enzyme required for formation of the peptidoglycan cell wall precursor UDP-ManAc. We have solved the epimerase's crystal structure to 1.70 Å, explaining the mechanism of its allosteric regulation, as well as showing the first example of direct contact between an allosteric and catalytic substrate molecule. Arg210 was identified as a critical residue for both binding sites. Furthermore, the structure was used to design an inhibitor that would bind to the conserved bacterial allosteric site and inhibit epimerase activity. After multiple rounds of

screening against Gram-positive strains and in *in vitro* assays a lead compound, Na.4030, was identified for further testing in mouse models of *B. anthracis* infection.

SseI is a *Salmonella* virulence effector secreted by its second Type III secretion system that is induced once the bacteria has invaded host cells. It has been implicated in having a direct role in establishing systemic infections. We have solved the structure of its catalytic C-terminal domain to 1.70 Å, using structural homology to identify it as a member of the Cysteine Protease Superfamily, having either acyl hydrolase or transferase activity. Furthermore, Cys178, His216, and Asp231 were identified as its catalytic triad. We were not able to show these activities in *in vitro* assays using small molecule and peptide substrates, but based on co-expression experiments with small Rho GTPases we were able to identify Cdc42 by mass spectrometry as a possible substrate undergoing deamidation at Gln61. The biological relevance of this result during infection is being investigated.

In loving memory of my father

ACKNOWLEDGEMENTS

I would like to sincerely thank my advisor, C. Erec Stebbins, for his mentorship, advice, support, and faith in me during my time in his laboratory. My gratitude also to all the other members of his group (Laboratory of Structural Microbiology) through the years who have been invaluable both in sharing their experiences with me and for their encouragement. I would also like to thank the members of my committee, Vincent Fischetti, Seth Darst, and David Thanassi, for their time and valuable advice on my research.

Furthermore, thank you to all my friends in San Francisco and New York who have been there with me through the ups and downs of my graduate career, I couldn't have done it without you. Special mention goes out to David Elgart and Lindsey Baker, who went above and beyond the call of duty.

Special thanks also to Merry Sherman, CEO and President of Mountain View Pharmaceuticals Inc., who guided my first steps into the world of the scientific unknown.

Finally, to my mother and sister who have loved and believed in me throughout my life, thank you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PART A:

UDP-GlcNAc 2-Epimerase from *Bacillus anthracis*

CHAPTER 1 - INTRODUCTION

1.1 *Bacillus anthracis*

B. anthracis is a non-motile, gram-positive, spore-forming, rod-like bacteria found in the soil and in reservoirs of wild animals that is capable of causing disease in humans and most mammals (Salyers & Whitt 2002). Its spores are extremely resistant to harsh conditions such as heat, UV and ionizing radiation, pressure, and chemical agents (Driks 2009). The distribution of anthrax worldwide can be traced to zones of soil containing high calcium and pHs above 6.1. Under alkaline conditions the spore surface is negatively charged and binds calcium and other divalent cations, enhancing spore survival (Hugh-Jones & Blackburn 2009). Interestingly, the genome sequencing of several members of the *Bacillus* family (single circular chromosome) has shown that anthrax and its non-pathogenic relatives are virtually identical, except for the presence of two virulence plasmids (pOX1, 203 ORFs; pOX2, 110 ORFs) that seem to confer its pathogenic phenotype (Keim et al 2009). The origin of these plasmids is unclear as well as whether they could be transferred to related strains such as *B. cereus* or *B. thuringiensis*.

Anthrax has been subject to study from the very beginning of modern microbiology. Discovered by Robert Koch in 1875, it was the first bacterium used by him to demonstrate his postulates of bacterial pathogenesis in 1876, and Louis Pasteur used it to

create one of the first animal vaccines in 1881 (Schwartz 2009). Historically, 'wool-sorters' disease (*B. anthracis* infection) was commonly found in farmers and workers who handled animal products such as meat, hide, and wool, and was also a major cause of animal death worldwide till the 1900s. However, after the widespread immunization of domestic herds in the 1890s with attenuated strains, both animal and human infections have become quite rare (Beyer & Turnbull 2009). In the US only 1 or 2 human cases are seen annually (usually the less severe cutaneous form), but outbreaks do occur periodically in other parts of the world, such as ~6000 cases reported in Zimbabwe in 1979 (Cieslak & Eitzen 1999).

Recently however there has been renewed interest in the study of anthrax because of its potential as a bioweapon (Franz 2009). The ability of its spores to resist harsh conditions and chemical processing, a spore size of 2-6 microns that is ideal for entry in the alveoli, and its high fatality rate if antibiotics are not taken within a short time-frame after exposure make it appealing as such. Moreover, human-human spread does not usually occur so the 'target area' is contained and controllable. The US, UK, and USSR created weaponized versions of anthrax in the 1940s and it has also been used by both international and domestic terrorists (Pohanka & Skladal 2009). In 1993 the Aum Shinrikyo cult in Japan released an aerosolized version from the roof of their headquarters. Fortunately, unknown to them it was the avirulent Sterne strain. In 2001, 5 letters containing highly processed weaponized anthrax spores were sent via post to newspapers, television stations and two Senators, contaminating 64 people, hospitalizing 17, and killing 5 others. The Senate mailroom took 3 months to fully decontaminate.

1.2 Pathogenesis

Anthrax is transmitted in the spore form to a host from grazing in infected pastures, by skin contact with contaminated animal products, by ingesting infected meat, or by inhalation. Once in the body spores are taken up by phagocytic cells at the infection site, triggering germination, the signals for which are small host molecules such as L-alanine (Salyers & Whitt 2002). The bacteria grow and divide within these cells, eventually bursting out and causing localized necrosis. They can also move to other organs in the body in the spore (carried intracellularly within macrophages) or vegetative (extracellular, fully germinated) form through the lymphatic system, causing systemic infection and septic shock. Symptoms usually appear after a 2-5 day incubation period. Spore formation is triggered by exposure to oxygen when bacteria are shed during the course of the infection or if the host dies (Mock & Fouet 2001). The cadaver and the soil surrounding it then become a reservoir of spores for further infections. The spores can last for up to 200 years (Hudson et al 2008) and infected carcasses must be burnt and contaminated areas treated with harsh chemicals such as chlorine dioxide and formaldehyde.

The majority of people exposed to anthrax spores develop cutaneous anthrax. Spores enter through cuts in the skin and germinate, leading to a large area of localized swelling (edema) which eventually turns black (called an eschar) due to necrosis. Recovery is around 99% if treated with antibiotics such as ciprofloxacin or clears on its own within several weeks. More rarely, ingested spores lead to gastrointestinal anthrax, which is

characterized by vomiting blood, lesions and inflammation of the intestinal tract, and severe diarrhea. Even when treated, fatality is between 25 and 60% if bacteria have spread to other organs. The most serious form of the disease is inhalation anthrax. Spores are picked up by and germinate within macrophages in the alveoli and very quickly move throughout the body. Patients are often asymptomatic until the release of large bacterial loads into the bloodstream, and death occurs quickly within a few days of septacemia. If treatment is given within the first 24 hours, the death rate is reduced from 97 to 45%, but recovery is rare outside of this window (www.cdc.gov).

The original Pasteur ($pXO1^-$) and Sterne ($pXO2^-$) vaccines for animals were based on live-attenuated strains that were either atoxic or unencapuslated, and easily cleared by the immune system. But sometimes these do still cause disease and newer vaccines for human use are based on protein preparations from avirulent strains that only contain an immunogenic toxin component called protective antigen (PA) (Cybulski et al 2009). However, these vaccines only provide short-term immunity and need to be given every 3 months for 18 months, followed by yearly booster shots (Biothrax™, Emergent Biosolutions).

1.3 Virulence Factors

Once its spores have germinated within the initial phagocytic cells and/or disseminated through the body, the extracellular vegetative form of the bacterium is characterized by its subversion of the host immune response (Moayeri & Leppla 2009, Tournier et al 2009,

Turk 2007). *B. anthracis* has two main virulence determinants that are carried on separate plasmids, but both are under the control of the same global transcriptional regulator AtxA, that responds to host-specific factors such as elevated temperature and CO_2 levels (Koehler 2009). pOX1 codes for a three-part toxin responsible for septic shock and edema, and pOX2 codes for enzymes required for the biosynthesis of an anti-phagocytic capsule that shields the vegetative form from immune detection. Thus anthrax is able to cause a combination of infection (due to bacterial multiplication) and toxemia (due to toxin release). These effects can studied individually in mice models depending on the strains of mice and *B. anthracis* used (Goossens 2009).

1.3.1 Toxins

The three parts of the toxin are lethal factor (LF, 90 kDa), edema factor (EF, 89 kDa), and protective antigen (PA, 83 kDa) (Collier & Young 2003). They are secreted by the bacterium into the environment but are non-toxic individually. First, PA must bind to the transmembrane cell surface receptors TEM8 (tumor endothelial marker 8) and CVC2 (capillary morphogenesis protein 2) both of which are ubiquitous and normally bind extracellular matrix components (van der Goot & Young 2009). Then PA is cleaved by furin proteases on the membrane surface with the larger C-terminal part oligomerizing into a heptameric prepore, which can then bind 1-3 molecules of either LF (the binary complex is called LFTx) or EF (called EFTx). This is a signal for endocytosis of LFTx or EFTx. The acidic pH of the late endosomes causes a conformational change in PA converting it into a transmembrane translocon that can transport LF and EF into the

cytosol (Collier 2009). Because of the requirement of PA for the action of both toxin forms and its high immunogenecity, it has been effective as a vaccine component.

LF is a highly specific Zn-protease that once in the cytosol specifically cleaves MKKs (mitogen-activated protein kinase kinases) at a conserved site (D-motif) that reduces their affinity for their cognate MAPKs (mitogen-activated protein kinases) and disrupts their phosphorylation cascade during signaling (Tonello & Montecucco 2009). LF attacks 6 of the 7 mammalian MKKs and shuts down the ERK, JNK, and p38 pathways. These are all essential for functioning of both the innate and adaptive immune system. LF has been shown to affect the release of pro-inflammatory cytokines from macrophages, neutrophil chemotaxis, antigen-presentation by dendritic cells, and the maturation/proliferation of B and T-cells. Some of these immune cells eventually die by necrosis or apoptosis. Later in the infection, circulating LeTx is thought to increase the cellular permeability/integrity of non-immune cells, eventually leading to their death as well.

EF is a highly active Ca^{2+} and calmodulin-dependent adenyl cyclase (CaM-AC). Once inside the cell, increased levels of cAMP (it is 100 times more active than endogenous CaM-AC) impair immune cell function with similar resulting effects as LF (Tang & Guo 2009). This is primarily through three signaling families: cAMP-dependent protein kinases, cyclic nucleotide gated channels, and EPAC (a guanine nucleotide exchange factor for Rap1 & 2). The adenyl cyclase toxins CyaA from *Bordetella pertussis* and ExoY from *Pseudomonas aeruginosa* work in a similar manner (Engel & Balachandran 2009, Vojtova et al 2006). Even though EF does not normally lead to apoptosis alone, its

transcription is coordinately regulated with LF and both toxins are thought to boost the other's effect on immune cells. For instance, cAMP can stimulate the expression of TEM8 and CVC2, enhancing toxin entry into immune cells (Maldonado-Arocho et al 2006).

1.3.2. Capsule

Like all Gram-positives *B. anthracis* has a thick peptidoglycan cell wall, made of glycan chains of alternating N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) cross-linked by tetrapeptides of L-Ala,-D-Glu,-meso-DAP,-D-Ala. Attached to the peptidoglycan by both covalent and non-covalent forces are other polysaccharides and proteins. However, vegetative *B. anthracis* has two other surfaces in addition to the peptidoglycan wall (Fouet 2009). The S-layer overlays the peptidoglycan and is composed of multiple crystalline protein layers of either Sap or EA1. Its exact role is unclear as S-layer mutants are still virulent but it seems to work with the capsule in enhancing resistance to host defenses.

The outermost layer is a capsule made from a monopeptide polymer (up to 1.5 MDa) of γ-D-glutamic acid, unlike most other bacterial capsules which are glycan-based rather than peptide-based. pOX2 codes for the enzymes required for γ-D-glutamic acid biosynthesis as well as other enzymes required for capsule formation. The capsule is essential for virulence, being weakly immunogenic, as well as having anti-phagocytic

activity (Scorpio et al 2007). This allows bacteria to evade the immune system when extracellular and contributes to their rapid multiplication once they enter the bloodstream.

1.4 UDP-GlcNAc 2-Epimerase

The non-hydrolyzing bacterial UDP-GlcNAc 2-epimerase (EC 5.1.3.14, 371 aa, 42.6 kDa, pI 5.85) catalyzes the reversible conversion of UDP-N-acetylglucosamine (UDP-GlcNAc) into UDP-N-acetylmannosamine (UDP-ManNAc) (Kawamura et al 1979, Kawamura et al 1978). Another enzyme, UDP-ManNAc dehydrogenase, then converts UDP-ManNAc into UDP-N-acetylmannosaminuronic acid (UDP-ManNAcA). The epimerization and dehydrogenation reaction products are sources of activated ManNAc and ManNAcA, intermediates in the biosynthesis of bacterial cell surface polysaccharides and the enterobacterial common antigen (ECA), a surface-associated glycolipid common to enteric bacteria (Kuhn et al 1988). The importance of UDP-GlcNAc 2-epimerase in the biosynthesis of the cell wall is highlighted by the presence of two functionally redundant copies of this enzyme in species such as *Staphyloccocus aureus* and *B. anthracis* (Kiser et al 1999, Read et al 2003).

Studies support a mechanism whereby the epimerization reaction proceeds via an initial *anti*-elimination of UDP to generate an intermediate 2-acetoamidoglucal, followed by the subsequent *syn*-addition of UDP to yield the product UDP-ManNAc (Tanner 2002) (Figure 1). The bacterial enzyme is related to the bi-functional mammalian UDP-GlcNAc 2-epimerase/ManNAc kinase, a hydrolyzing enzyme that converts UDP-GlcNAc

into separate UDP and ManNAc molecules and phophorylates the latter into ManNAc 6-phosphate (Hinderlich et al 1997, Stasche et al 1997). The mammalian enzyme catalyzes the rate-limiting step in sialic acid biosynthesis and is a key regulator of cell surface sialylation in humans (Keppler et al 1999).

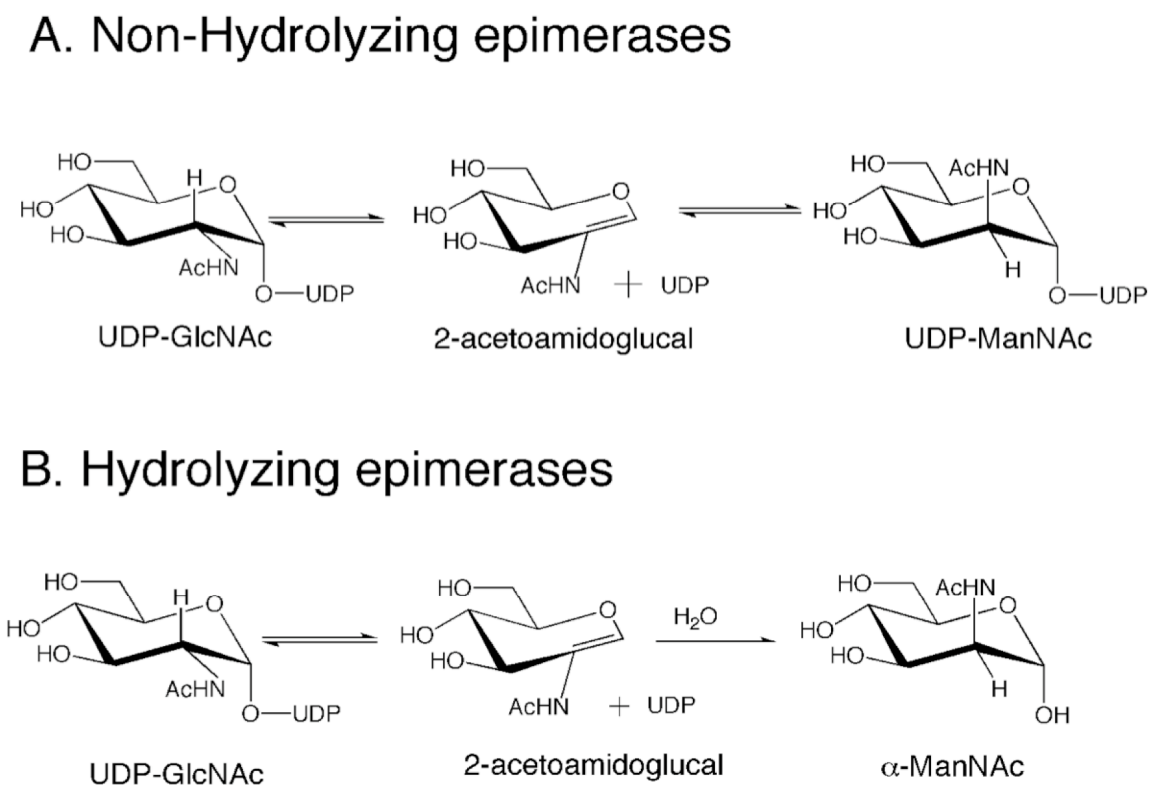


Figure 1: UDP-GlcNAc 2-epimerase catalyzed reaction. Reactions catalyzed by the (A) bacterial and (B) mammalian enzymes.

A previous structure of UDP-GlcNAc 2-epimerase from *E. coli* revealed structural homology to glycosyl transferases such as T4 phage β-glucosyltransferase, and glycogen phosphorylase (Campbell et al 2000). Although the enzyme was crystallized in the presence of UDP-GlcNAc, only UDP, which is an intermediate in the reaction, was

observed in the structure bound at a site which has since been established to be the catalytic site based on mutagenesis studies (Samuel & Tanner 2004).

A unique feature of the bacterial UDP-GlcNAc 2-epimerases is their allosteric regulation by the substrate, UDP-GlcNAc, which acts as an activator. In the absence of this activator, no UDP-ManNAc is epimerized in the reverse reaction (Kawamura et al 1979), but when trace amounts of UDP-GlcNAc are added, the reaction proceeds to equilibrium. This suggests that UDP-GlcNAc is required for the enzyme to acquire a conformation in which it is catalytically competent. The structure of the *E. coli* epimerase revealed that each subunit of the dimer of the enzyme adopted a slightly different conformation due to a 10° interdomain rotation. This was proposed to be part of the allosteric regulatory mechanism – binding of the substrate to one molecule would then alter the conformation of the other molecule also making it catalytically active. However, it remained unclear how UDP-GlcNAc could trigger these changes, as it was absent from the structure.

Our initial structural and biochemical investigation of UDP-GlcNAc 2-epimerase was part of a collaboration with the Laboratory of Bacterial Pathogenesis & Immunology at Rockefeller University. R. Schuch and V. Fischetti were able to identify the epimerase as an essential enzyme for *B. anthracis* survival based on a screen using the cell-binding domain (CBD) of the *B. anthracis*-specific lysin PlyG. Lysins are enzymes released by lytic bateriophages that rapidly degrade the peptidoglycan cell wall of Gram-positive bacteria, facilitating phage release and dissemination after their replication within the host cell (Fischetti 2008, Schuch & Fischetti 2009). The lysin N-terminal catalytic

domain usually has glycosidase activity, with a few variants that are well conserved, but the CBD varies depending on the cell wall component (receptor) that it binds.

Since *B. anthracis* strains resistant to PlyG are rare, its CBD must be targeting an essential component of the cell wall that the bacterium cannot easily change. The Fischetti laboratory was able to identify the binding receptor for PlyG as a *B. anthracis*-specific cell surface carbohydrate made of galactose, N-acetylglucosamine, and N-acetylmannosamine. A bioinformatic survey of the *B. anthracis* genome for enzymes involved in the biosynthesis of these sugars then identified the epimerase as a possible target for further characterization.

1.4 Aims

The conserved nature of non-hydrolyzing UDP-GlcNAc 2-epimerase in bacteria but not mammals makes it a good target for antibiotic development both against *B. anthracis* and Gram-positives in general. We took a structural approach to studying this enzyme, not only to determine its allosteric binding site and explain the mechanism of allosteric regulation, but also to use the crystal structure as the basis for the design of small molecules that would inhibit its activity *in vitro* and against live strains.

CHAPTER 2 - MATERIALS AND METHODS

2.1 Protein expression and purification

The genes coding for both UDP-GlcNAc 2-epimerase (*BA5509*) and UDP-ManNAc dehydrogenase (*BA5512*) were amplified from genomic DNA of *B. anthracis* and cloned into a modified version of a pCDF-Duet-1 vector (Novagen) containing two tandem hexahistidine tags separated by a linker and followed by a downstream rhinovirus 3C protease cleavage site. Bacterial cultures were grown in LB medium at 37ºC to stationary phase and induced with 0.1 mM IPTG at 15ºC overnight. The cells were harvested by centrifugation, re-suspended in a buffer containing 50 mM Tris-HCl pH 8.0, 10 mM Imidazole, 100 mM NaCl, and lysed by passage through a homogenizer (Avestin). Following centrifugation, the lysate was loaded onto a gravity flow column packed with Ni-NTA resin (QIAGEN) which was then washed with 10 column volumes of lysis buffer and 10 column volumes of wash buffer (lysis buffer with 30 mM Imidazole). The column was then eluted with 3 volumes of Elution buffer (lysis buffer + 500 mM Imidazole). Following overnight dialysis and cleavage with the rhinovirus 3C protease, further purification was performed on source Q15 and Superdex 200 columns (GE Healthcare). The purified proteins were concentrated to 20 mg/ml and stored at -80ºC. Point mutants of the epimerase gene were generated using a Quick Change site directed mutagenesis kit (Stratagene).

2.2 Crystallization and structure determination

(conducted by L. Velloso, Laboratory of Structural Microbiology, Rockefeller University)

Crystals of the *B. anthracis* UDP-GlcNAc 2-epimerase could only be obtained upon co-crystallization with the substrate. The protein was crystallized in hanging drop vapor diffusion experiments at 4ºC. Drops consisted of an equal volume of protein (5 mg/ml in 50 mM Tris-HCl pH 8.0, 5 mM UDP-GlcNAc (Sigma), 200 mM NaCl, 2 mM DTT) and well solution (22-26% PEG 4000-6000, 100 mM Tris-HCl pH 8.2-8.8, and 0.2 M Li_2SO_4). Crystals appeared after a week and grew to full size over several weeks. They were cryoprotected by gradual transfer (5% steps) into mother liquor supplemented with 30% glycerol, before being frozen in liquid nitrogen.

Data was collected at beamline X29 at the National Synchroton Light Source and processed with HKL2000 (Otwinowski 1997). The dataset used for refinement was then reprocessed using MOSFLM (Leslie 1992) and SCALA (Evans 1997). Crystals belonged to space group $P2_1$ (a=46.4, b=188.2 c=61.7, β=112.2) with a Mathews coefficient of 3.0 $Å^3$/Da (60% solvent), consistent with the observed presence of one dimer in the asymmetric unit. Molecular replacement was performed with the program PHASER (McCoy et al 2005), using the structure of the UDP-GlcNAc 2-epimerase from *Bacillus subtlis* as a search model (PDB ID 1O6C, solved as part of a structural genomics program by Structural GenomiX (SGX)) (Badger et al 2005). Notably, the latter structure was solved in the absence of any ligand and has many disordered regions.

The resulting model from molecular replacement was put though several cycles of automated model building and refinement in Arp/Warp (Perrakis et al 1999), which resulted in a nearly complete model of the *Bacillus anthracis* enzyme. At this point, clear density was seen in the active sites of both dimer subunits for both UDP and UDP-GlcNAc molecules. These ligands were fit using the LigandBuild option of ArpWarp. Additional model building was performed in Coot (Emsley & Cowtan 2004) and the structure was further refined with Refmac5 (Murshudov et al 1997), using NCS restraints. Simulated-annealing omit maps were calculated in CNS (Brunger et al 1998). Figures were generated using PyMol (www.pymol.org).

2.3 Enzyme kinetics

The kinetic parameters of *B. anthracis* UDP-GlcNAc 2-epimerase for the epimerization of UDP-GlcNAc were obtained using a modified version of a previously described coupled assay that employs UDP-ManNAc dehydrogenase (Morgan et al 1997). Assay mixtures contained 15 mM NAD^+, 100 mM Tris-HCl pH 8.8, 10 mM DTT, 0.125 mM (0.75 mg) of Dehydrogenase, and varying concentrations of UDP-GlcNAc (SIGMA) in 100 μl volume. The UDP-GlcNAc stock concentration was determined by measuring the absorbance at 262 nm, using a ε_{262}= 9890 M^{-1}. The reactions were initiated by addition of the epimerase enzyme (72 nM for wild-type and 0.3 to 33 μM for mutants) and the rates were determined by continuous monitoring of the absorbance at 340 nm (NADH formation) at 37°C. Kinetic parameters were determined from initial reaction velocities fit to the Hill equation using Prism 4 (GraphPad Software Inc):

$$v = V_{max}\,[S]^n/([S]^n + K_m{}^n) \qquad \text{and} \qquad k_{cat}=V_{max}/[E].$$

2.4 Measuring cellular epimerase activity

B. anthracis cell lysates from lines containing knockout mutants of the two versions of the epimerase genes were obtained from A. Pelzek (Laboratory of Bacterial Pathogenesis & Immunology). They were normalized for protein concentration and then tested in duplicate in the coupled assay. 34 μL of lysate at ~3 mg/mL total protein was added per well in a 100 μL final volume. 10 mM substrate was used in each case to achieve saturating rates, and wells without substrate were used to correct for non-specific dehydrogenase activity from the lysates. Known amounts of epimerase were used as standards to calculate the total amount of epimerase present in each strain.

2.5 Inhibitor testing on bacterial strains

(conducted by A. Pelzek, Laboratory of Bacterial Pathogenesis & Immunology, Rockefeller University)

All compounds were solubilized to 5 mM in DMSO as stocks before testing. Any undissolved clumps were broken up by sonication. Gram positives are able to handle up to 3% DMSO so appropriate dilutions were made to ensure that the same amount of DMSO was added in each case. Inhibitors were initially tested at 150 μM and then at again at 30 μM since many compounds precipitated when added to the cell culture medium at the higher concentration. Overnight cultures were inoculated 1:100 into fresh

BHI media and grown for ~3 hrs at 30°. The cultures were then normalized to an OD_{600} of 0.2 and added to 100 μL of an inhibitor (5 μL in DMSO) and BHI (95 μL) mix in a 96-well plate. Growth at 30° in the presence of inhibitor was measured at OD_{600} every two minutes with 40 s of shaking between reads for 16 hrs (Spectramax Plus, Molecular Devices). Growth in DMSO without inhibitor was used as a control. The strains tested were *Bacillus anthracis* ΔSterne, *Staphylococcus aureus* MRSA, and *Streptococcus* Group B (Type II).

2.6 Inhibitor testing *in vitro*

The assay was modified before testing for inhibitory activity. Many of the compounds have a strong absorbance at 340 nm, which saturated the signal due to NADH formation and made quantification difficult. Halving the volume to 50 μL reduced the absolute absorbance of each well and enabled accurate rate measurements. Also, the first round of compounds tested were provided as 5 mM stocks in DMSO. As a result, at the tested inhibitor concentration of 0.5 mM, 10 % DMSO was present in the final reaction mix. 0.6 μg epimerase and 0.5mM substrate was present in each well. Reactions were started by adding epimerase to wells containing substrate and inhibitor that had been pre-incubated together for 20 mins at 37°C. The second round of compounds were tested at inhibitor concentrations of 0.1 mM, 0.25 mM, and 0.5 mM, with 0.6 μg epimerase and 0.5 mM substrate. The final compound, Na.4030, was tested at multiple concentrations between 0.05 mM and 0.75 mM (2 mM substrate) in the original version of the assay, without added DMSO. All compounds in all three rounds were tested in triplicate.

CHAPTER 3 - RESULTS AND DISCUSSION

3.1 Overall crystal structure

The structure of the UDP-GlcNAc 2-epimerase from *B. anthracis* was solved to 1.70 Å by Molecular Replacement from crystals grown in the presence of UDP-GlcNAc (Table 1). The final model contained 750 protein residues (375 from each chain, including 4 vector-encoded residues), 2 UDPs, 2 UDP-GlcNAcs, and 577 water molecules. The Ramachandran plot shows 91.6% of residues in the most favoured regions, and 8.1% and 0.3% in the generously and additionally allowed regions, respectively. It is very similar to the previously determined structures of the homologous *E. coli* UDP-GlcNAc 2-epimerase in complex with UDP (PDB ID 1F6D) (Campbell et al 2000) or with UDP-GalNAc (1VGV) (Badger et al 2005) (Figure 2). There are two molecules in each asymmetric unit with each molecule being composed of two α/β/α domains with a deep substrate cleft between them. In contrast to the *E. coli* structure, both chains of the *B. anthracis* epimerase dimer are in the same conformation. Secondary structure matching (SSM) superposition of equivalent Cα atoms of the dimer subunits yields an r.m.s.d of 0.06 Å and 1.88 Å for the *B. anthracis* and *E. coli* (1F6D) enzymes respectively. The *E coli* structure showed strong density for a UDP molecule bound to one chain in each dimer (called the 'closed' form) but very weak density in the same region of the other chain ('open' form). Comparison between the closed form of the *E. coli* enzyme (1F6D) and the *B. anthracis* enzyme yields an r.m.s.d of 1.63 Å for 341 Cα atoms. Hereafter,

unless otherwise stated, the closed form of the *E. coli* structure is used for all structural comparisons.

Table 1: Crystallographic statistics. Values in parentheses are for the highest resolution shell.

Table 1: Crystallographic statistics[a]	
Data collection	
Wavelength (Å)	1.08090
Resolution range	60-1.7
Number of unique reflections	106456
Average redundancy	3.5 (3.4)
Completeness	99.4 (99.0)
R_{merge} (%)	8.7 (34.6)
I/σ	14.6 (3.5)
Wilson B-factor *(A^2)	16.1
Refinement	
Number of reflections	
Work set	101096
Test set	5360
R_{factor} (%)	22.1
R_{free} (%)	26.2
r.m.s.d. for bond lengths (Å)	0.018
r.m.s.d. for bond angles (deg)	1.76
Average B-factor ($Å^2$)	
Main chains	14.8
Side chains	16.7
Waters	27.4
UDP	12.4
UDP-GlcNAc	10.6

3.2 Active site

Clear electron density for both UDP and UDP-GlcNAc molecules was observed in the active site of the *B. anthracis* enzyme (Figure 2B). UDP and 2-acetoamidoglucal are thermodynamically favored intermediates of the epimerase catalyzed reaction and are released into solution upon prolonged incubation of the enzyme with UDP-GlcNAc (Samuel & Tanner 2004). All of the interactions between UDP and the epimerase are conserved between the *B. anthracis* and the *E. coli* enzymes and a similar presence of water molecules and lack of electron density for 2-acetoamidoglucal in the active site was observed. Although the residues responsible for proton abstraction/addition have not been unambiguously identified, mutations in the *E. coli* enzyme of ionizable residues found in the region where the glucosamine moiety of the substrate would be positioned, identified Asp95 and Glu131 as candidates for the proton abstraction and Glu117 as being involved in the second reaction step. These residues are conserved in *B. anthracis* (as Asp100, Glu136, Glu122) and are in similar positions in the structure, making them strong candidates for catalytic residues.

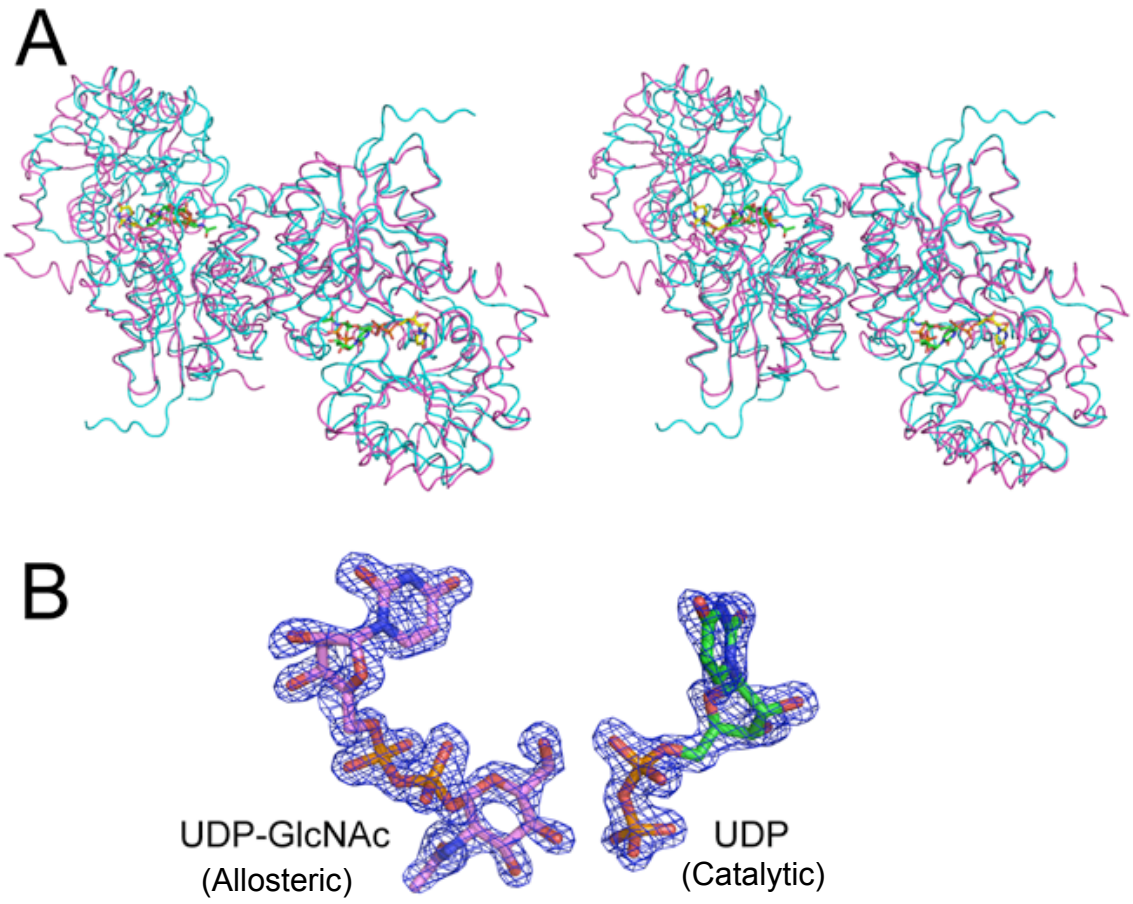


Figure 2: UDP-GlcNAc 2-epimerase structure. A. Stereo view of the superposition of the structures of UDP-GlcNAc 2-epimerase from *B. anthracis* (light blue) and *E. coli* (magenta). UDP and UDP-GlcNAc from the *B. anthracis* structure are shown as sticks. B. Fo-Fc simulated annealing omit electron density map contoured at 3σ for UDP-GlcNAc and UDP calculated using the final model minus these molecules.

3.3 Allosteric site

This is the first time that UDP-GlcNAc has been observed in the active site of UDP-GlcNAc 2-epimerase. The molecule lies in an extended pocket lined by several hydrophilic side chains and forms hydrogen bonds to the side chains of residues Gln43, Gln46, Gln70, His44, His242, Arg210 and Glu136 (Figure 3). It also makes hydrogen bonds to the main chain of the enzyme and to water molecules. UDP-GlcNAc makes two hydrogen bonds to the α and β phosphates of the adjacent UDP molecule, helping to hold it in place for the second step of the epimerization reaction.

Sequence alignments of UDP-GlcNAc 2-epimerases from several bacterial species show that all of the residues that interact with UDP-GlcNAc via their side chains are highly conserved in the non-hydrolyzing enzymes but variant on the hydrolyzing ones (Figure 4). Therefore, this UDP-GlcNAc binding site is conserved and probably unique to the non-hydrolyzing bacterial UDP-GlcNAc 2-epimerases.

Although the crystals of the *E. coli* enzyme (1F6D) were also grown in the presence of UDP-GlcNAc and belong to the same space group with related unit cell dimensions, they were obtained at pH 5.25 where the enzyme is only marginally active (Kawamura et al 1979). The *B. anthracis* epimerase crystals were grown in a pH range of 8.2-8.8, which is the optimal pH range for the enzyme. Thus, it may be that UDP-GlcNAc binds optimally to the enzyme only at a higher pH, which might help explain how we trapped the enzyme in its UDP-GlcNAc bound form.

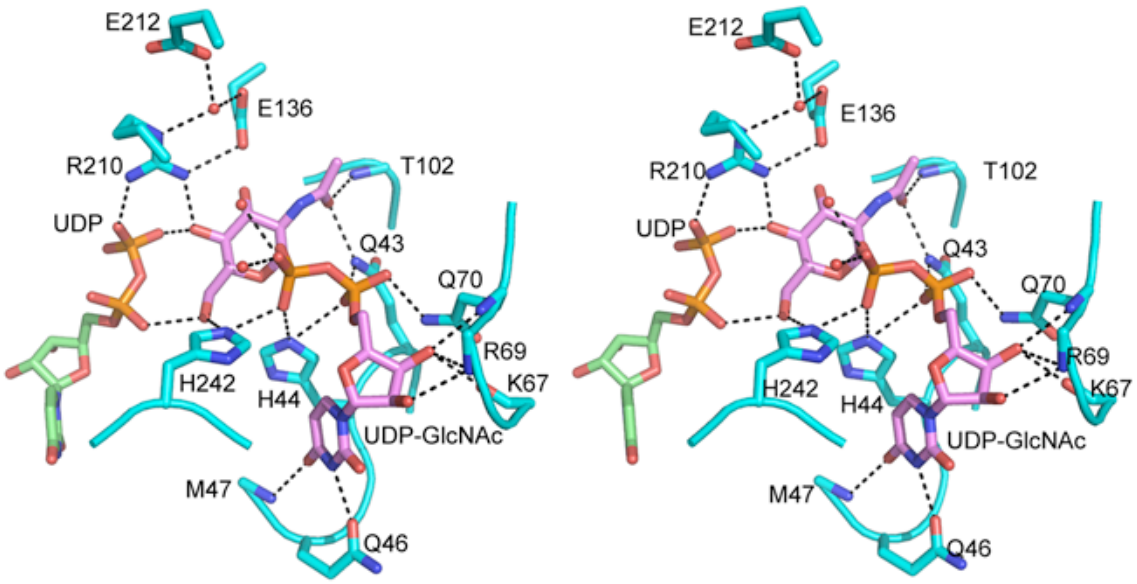


Figure 3: Interactions between the enzyme and UDP-GlcNAc. Stereo view of the hydrogen bond interactions between UDP-GlcNAc (pink), UDP (green) and *B*. anthracis epimerase residues (light blue). Water molecules are shown as red spheres and hydrogen bonds as dotted lines.

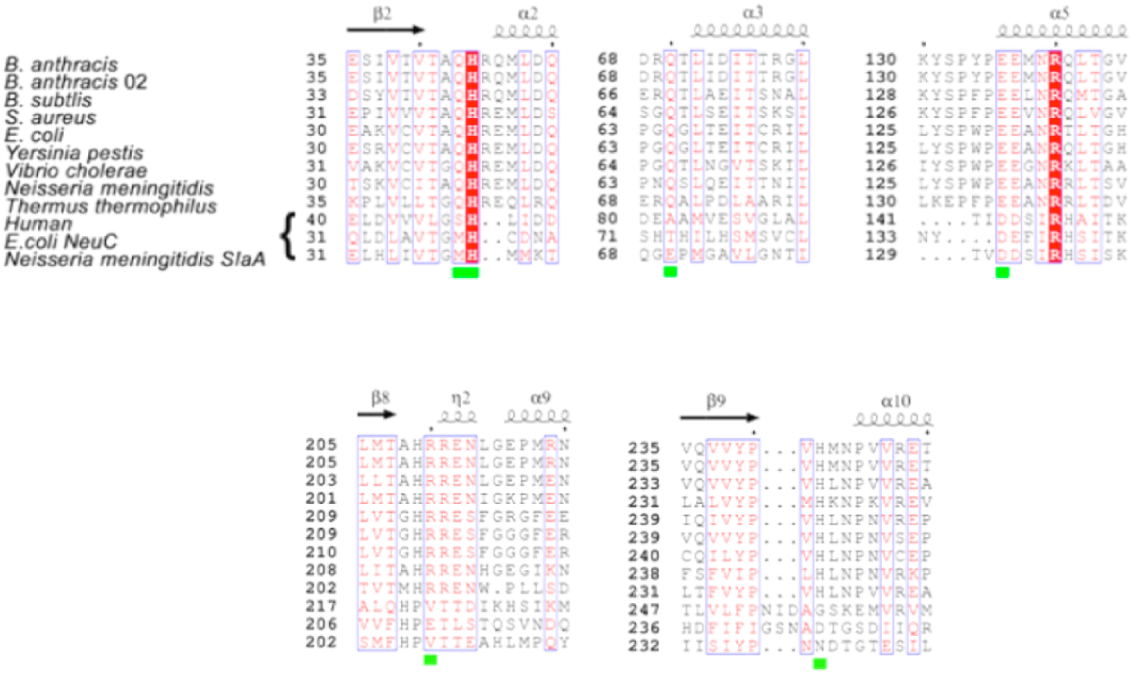


Figure 4: Sequence alignments showing the conservation of the UDP-GlcNAc interacting residues in different non-hydrolyzing UDP-GlcNAc 2-epimerases. The bottom 3 sequences in the alignments (human, SiaA and NeuC) represent hydrolyzing UDP-GlcNAc epimerases. Positions of conserved allosteric site residues are indicated by green squares.

3.4 Conformational changes induced by substrate binding

Comparison between the UDP-GlcNAc 2-epimerase structure and that of the *E. coli* enzyme in complex with UDP reveals that binding of the substrate UDP-GlcNAc triggers major but localized conformational changes which take place mostly in three loops (His209-Gly215, Ile65-Leu72, and Val241-Pro245) (Figure 5) which contain the residues interacting with the pyrophosphate and glucosamine moieties of UDP-GlcNAc.

The flipping of the loop His209-Gly215 (His213-Gly219 in *E. coli*) is the most striking conformational change and leads to shifts of 6.5 Å and 12 Å in the Cα atom and guanidium group positions of residue Arg210 (Arg214). This contrasts with a shift of only 0.5 Å for the Cα atom of His209 (His213), which makes hydrogen bonds to the β-phosphate of UDP in both the *B. anthracis* and the *E. coli* enzyme structures. This loop flip brings the side chain of Arg210 into position to make hydrogen bonds to the β-phosphate of UDP, a hydroxyl group of UDP-GlcNAc, the carboxyl group of Glu136 (Glu131) and a water molecule (Figure 3). Thus, Arg210 plays an important role in anchoring both UDP and UDP-GlcNAc, and since its guanidium is likely to be fully protonated at the optimum pH of the enzyme (between 7 and 9) it probably stabilizes the leaving UDP in the *anti*-elimination reaction.

The loop Ile65-Leu72 (Ile60-Leu67) was partly disordered in the *E. coli* epimerase enzyme structures where residues 66-69 (61-64) were not seen. In the *B. anthracis*

structure, it makes four hydrogen bonds from Arg69 and Gln70 to the ribose and α-phosphate of UDP-GlcNAc (Figure 3).

The conformational changes that take place on the Val241-Pro245 (Val245-Pro249) loop are reminiscent of those in the His209-Gly215 (His213-Gly219) loop, shifting the Cα and Nε2 atoms of residue His242 (His246) by about 3 Å and 9 Å respectively between both structures. These movements bring His242 into hydrogen bonding position to both the β-phosphate and a hydroxyl of the UDP-GlcNAc molecule.

The binding of UDP-GlcNAc to the allosteric site of the enzyme not only triggers closure of the active site, but also obstructs its access to the solvent, thus preventing the products of the *anti*-elimination reaction from diffusing away before they can react again in the *syn*-addition. These enzymes have been shown to release the intermediates UDP and 2-acetoamidoglucal into solution at a rate of about 1/400 that of the UDP-GlcAc epimerization (Morgan et al 1997, Tanner 2002) and since these are thermodynamically more stable than either the substrate or the product of the reaction, the enzyme must kinetically trap them to avoid their release into solution. Consequently, following every catalytic cycle, the active site must open for product release and substrate uptake.

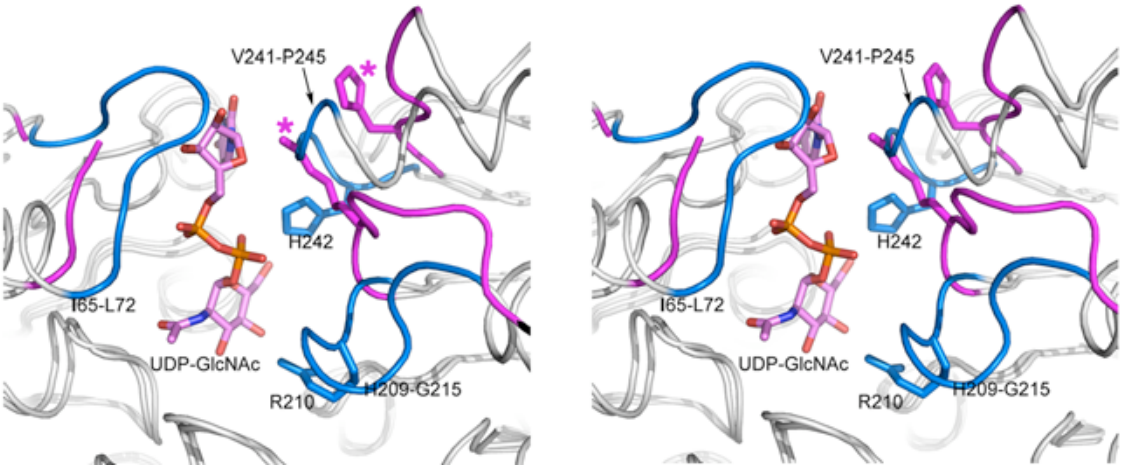


Figure 5: Conformational changes triggered by UDP-GlcNAc. The *B. anthracis* UDP-GlcNAc 2-epimerase is shown superposed with the closed form of the *E. coli* enzyme (PDB ID 1F6D, chain B). The largest conformational changes are shown in blue and pink for the *B. anthracis* and *E. coli* and enzymes, respectively. The side chains of Arg210 and His242 are shown for both enzymes (asterisks mark *E. coli*), and UDP-GlcNAc in pink.

3.5 Kinetic parameters of allosteric binding site mutants

Since bacterial UDP-GlcNAc 2-epimerases are tightly allosterically regulated by the substrate and the UDP-GlcNAc interacting residues are strictly conserved it seems that the allosteric site is key to the conformational changes needed to convert the enzyme to an active state. To investigate this, we generated point mutants with alanine substitutions of UDP-GlcNAc-interacting residues. All mutants, except His44Ala were soluble and behaved as wild-type enzyme during the purification. As the His44Ala mutation completely abolished the solubility of the protein, a His to Gln change was made which overcame the solubility problem.

In order to measure the kinetic parameters of the UDP-GlcNAc 2-epimerase and mutants in its allosteric binding site, a modified microplate version of a previously published cuvette assay was created. The assay had initially been used to measure the activity of the *E. coli* version of the enzyme. It works in two steps. First, UDP-GlcNAc is converted to UDP-ManNAc by the epimerase and then to UDP-N-acetylmannosaminuronic acid (UDP-ManNAcA) by *B. anthracis* UDP-ManNAc dehydrogenase. The latter reaction uses NAD^+ and the appearance of NADH can be monitored at 340 nm.

Since the assay contains two coupled enzymes, each of its components were tested to check that they were not limiting. It was found that significantly higher amounts of NAD^+ (15 mM vs 4 mM) were needed than for the *E. coli* enzyme to ensure complete coupling of the two reactions and accurate measurement of *B. anthracis* epimerase activity. This might be due to a weaker binding affinity for NAD^+ in the *B. anthracis* version of the dehydrogenase enzyme, which has not been previously characterized. In addition, 10 mM DTT was used instead of 2 mM. The dehydrogenase has a Cys in its active site that is critical for function and it has been shown previously for the *B. cereus* enzyme that DTT is needed to reduce inactivating disulphide bond formation between enzyme molecules (Kawamura et al 1978). It should be noted that the dehydrogenase is present at ~1000 times the amount of epimerase, even in the *E. coli* version of the assay. The assay was also tested for interferences from common buffer components and found to be robust and not affected by 50 mM Na^+, 50 mM K^+, 50 mM Cl^-, 5 mM $CaCl^{2+}$, or 5 mM EDTA.

Kinetic data for the wild-type and mutant enzymes are summarized in Figure 6 and Table 2. The K_M for the wild-type enzyme (2.2 mM) is between 2 to 3 fold higher than those reported for the *E. coli* (0.6 mM) and *Bacillus cereus* enzymes (1.1 mM), even though the k_{cat} is similar to *E. coli* (7.9 s^{-1} vs 7.1 s^{-1}), whereas the Hill coefficient is slightly lower than both (1.4 vs 1.8) (Kawamura et al 1979, Kawamura et al 1978, Morgan et al 1997, Samuel & Tanner 2004). A possible explanation is that *B. anthracis* contains two redundant and identical copies of the UDP-GlcNAc 2-epimerase gene in contrast to the single copy present in *E. coli*. Thus, it might still achieve comparable UDP-GlcNAc turnover rates to *E. coli* by using higher amounts of UDP-GlcNAc 2-epimerase.

The Arg210Ala mutant shows a 10-fold increase in K_M and a 400-fold decrease in kcat, suggesting this residue has an important role in stabilizing the incoming UDP-GlcNAc that enters the active site, and the resulting UDP produced in the *anti*-elimination reaction. Since Arg210 also makes hydrogen bonds to a UDP-GlcNAc hydroxyl that interacts with UDP, it plays a role in stabilizing two β-phosphate oxygen atoms of the latter. The lower Hill coefficient of the Arg210Ala mutant suggests that this residue is also involved in the allosteric regulation of the enzyme. Notably, Arg210 also interacts with Glu136 (Glu131 in *E. coli*) via two hydrogen bonds (one direct and one water-mediated), and at the optimal pH range of the epimerase reaction (7 to 9), these side chains are likely to be ionized with opposite charges. The mutation Glu131Gln in the *E. coli* epimerase led to a 10,000-fold decrease in k_{cat} (Samuel & Tanner 2004), and the disruption of this interaction with Glu136 in the Arg210Ala *B. anthracis* mutant would explain its large decrease in k_{cat}. Interestingly, Arg210 is conserved only in the non-

hydrolyzing bacterial UDP-GlcNAc 2-epimerases, whereas Glu136 (Glu131) is conserved in character in all UDP-GlcNAc 2-epimerases suggesting that the crucial role of Arg210 and the conformational changes associated with it are unique to the bacterial non-hydrolyzing UDP-GlcNAc 2-epimerases.

The His242Ala mutant also shows a moderate increase in K_M (5-fold) and a pronounced decrease in kcat (36-fold). This residue makes hydrogen bonds to the β-phosphate and to a hydroxyl group of UDP-GlcNAc that interacts with the neighboring UDP molecule. Thus, it likely plays a role in both the binding of UDP-GlcNAc and the stabilization of the UDP intermediate.

The mutants Gln70Ala, Gln43Ala and His44Gln have in common moderate decreases (2 to 7 fold) in kcat and increases in K_M (3 to 8-fold). All these residues make hydrogen bonds to the axial oxygens of the α-phosphate of the UDP-GlcNAc. His44 also makes hydrogen bonds to one of the axial oxygens of the β-phosphate, and Gln43 to the carbonyl of the acetyl moiety of UDP-GlcNAc. Thus, it appears that interactions between the epimerase and the α-phosphate of UDP-GlcNAc help position the latter for optimal interaction with the UDP in the catalytic site. The weakened α-phosphate interactions due to the loss of a few hydrogen bonds in these mutants probably lead to a lower affinity of the allosteric site for UDP-GlcNAc, and also as a consequence for the bound UDP.

The relatively larger changes in the K_M and k_{cat} of the R210A mutant compared to those of the other mutants confirm that this residue is central to the conformational changes that

lead to the UDP-GlcNAc-induced activation of the *B. anthacis* UDP-GlcNAc 2-epimerase and to the stabilization of the interaction between UDP-GlcNAc molecules in the active and allosteric site.

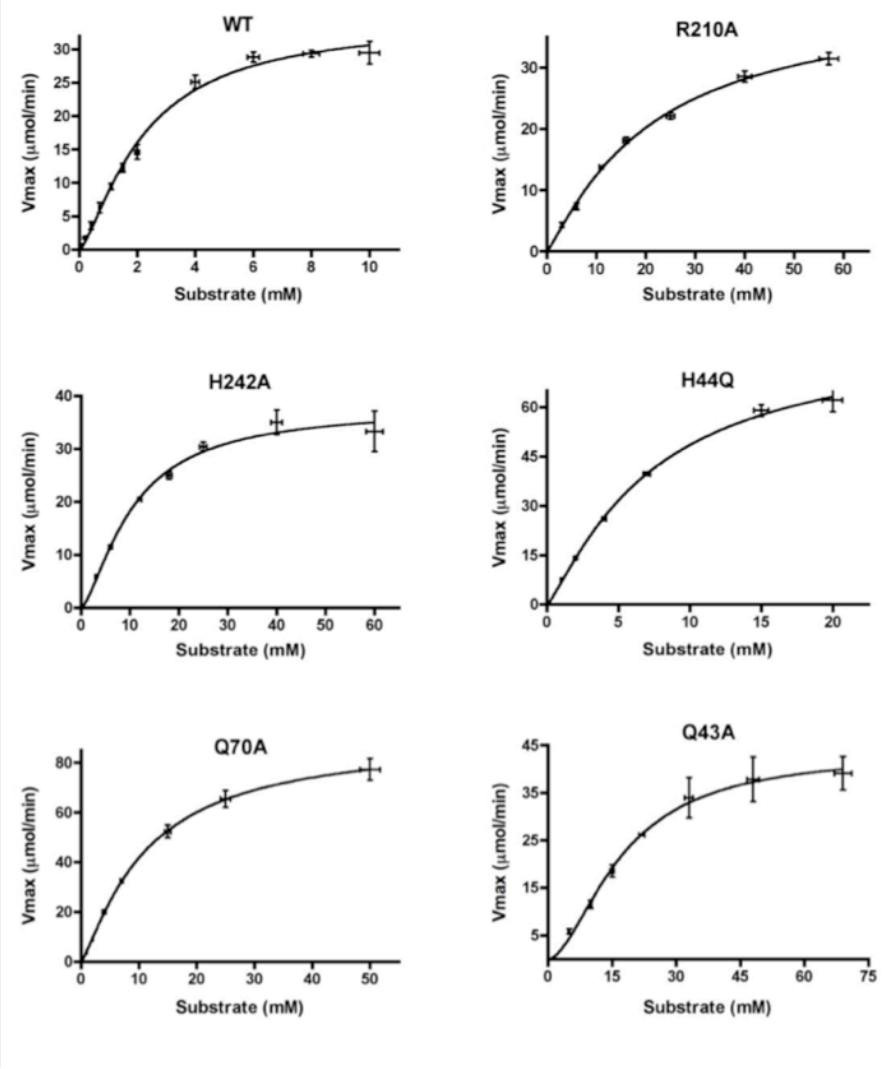


Figure 6: Substrate saturation plots for wild-type and mutants of *B. anthracis* UDP-GlcNAc 2-epimerase. The velocities (± SD) were obtained using kinetic data from three replicates for each enzyme. Enzyme concentrations used were as follows: WT (0.072 μM), R210A (16.8 μM), H242A (2.9 μM), H44Q (0.34 μM), Q70A (0.72 μM), Q43A (0.60 μM).

Table 2: Kinetic data (± SEM, n=3) for the epimerization of UDP-GlcNAc by wild-type and mutant *B. anthracis* UDP-GlcNAc 2-epimerase.

Enzyme	kcat (s^{-1})	K_M (mM)	kcat/K_M ($s^{-1}.mM^{-1}$)	n
WT	7.9 ± 0.3	2.2 ± 0.2	3.6	1.4 ± 0.1
R210A	0.021 ± 0.001	20.9 ± 2.2	0.001	1.15 ± 0.07
H242A	0.22 ± 0.10	10.4 ± 1.0	0.021	1.47 ± 0.15
H44Q	4.0 ± 0.2	7.3 ± 0.8	0.55	1.21 ± 0.08
Q70A	2.06 ± 0.07	11.1 ± 0.8	0.19	1.24 ± 0.06
Q43A	1.20 ± 0.07	17.2 ± 1.6	0.070	1.77 ± 0.23

3.6 Measuring epimerase activity in mutant *B. anthracis* strains

As a further test of the versatility of the coupled assay, it was also used to measure epimerase activity in the soluble fraction of whole cell lysates from WT and epimerase knockout *B. anthracis* strains (provided by A. Pelzek) (Table 3). Controls showed that the lysis buffer components did not interfere with the assay, and epimerase activity was readily detectable. There was significant non-specific dehydrogenase activity in the lysates (~30% of signal) and wells with no added substrate were used to correct for this background. It was also found that epimerase activity was not inhibited by other lysate components since known amounts of epimerase spiked into lysate showed the same activity as enzyme present in buffer only.

The necessity of epimerase activity for bacterial survival was evident in that all knockouts showed much slower growth than the WT strains and by the non-viability of a double knockout (A. Pelzek, personal communication). The epimerase also seems to account for 0.1 to 0.2 % of all soluble cellular protein. It is interesting that when one copy of the epimerase gene is deleted, cellular levels of epimerase from the remaining gene are higher than total WT expression from both genes. Perhaps the two genes are

normally tightly controlled and not expressing at their maximal levels, but disrupting one leads to maximal transcription of the other as a survival response.

Table 3: Lysate activity of epimerase gene knockouts. Each strain contained an IPTG-inducible plasmid with the corresponding gene that had been knocked out of the genome. +/- indicates strains grown in uninduced and induced conditions.

Strain Name	Description	μg epimerase/ mg soluble protein
45	WT	1.1
856-	BA5509 KO, inducible	1.5
856+		3.2
862-	BA5509 KO, inducible	1.8
862+		3.4
877-	BA5509 KO, non-inducible	1.7
877+		1.5
1033	BA5433 KO	1.7
1035	BA5433 KO	2.4
1205-	BA5433 KO, BA5509 inducible	1.7
1205+		2.6

3.7 Screening allosteric site inhibitors

Since the allosteric substrate site is conserved among the non-hydrolyzing bacterial UDP-GlcNAc 2-epimerases, but not the mammalian hydrolyzing version, it makes it a good candidate for structure-based design of compounds that would work against *B. anthracis* but not harm the host. This was done in collaboration with the pharmaceutical company SuperGen and the Laboratory of Bacterial Pathogenesis & Immunology at Rockefeller University.

Using the allosteric site rather than the catalytic site, a group at SuperGen was able to screen *in silico* for small molecules that would bind strongly to this regulatory site. Based on these hits and further modeling they then synthesized a hundred compounds as possible inhibitors. These were tested against the Gram-positives *B. anthracis* ΔSterne, *Staphylococcus aureus* MRSA, and *Streptococcus* Group B (Type II) for growth inhibition (A. Pelzek, Laboratory of Bacterial Pathogenesis & Immunology). Compounds (55-93) were also tested against the Gram-negatives *Pseudomonas aeruginosa* and *Vibrio vulnificus*. Several compounds showed activity against multiple Gram-positive strains but none showed growth inhibition against the Gram-negatives, except for 55, which worked against *V. vulnificus* (Table 4).

Table 4: Round 1 compounds active against bacterial strains

Species	Inhibitory compounds
B. anthracis	55 56 74 75 81 82 85
Staph MRSA	55 56 74 75 77 81 85
GBS Type II	55 75 81 85
Vibrio C1	55

In order to show that these compounds were inhibiting growth through the epimerase and not some other mechanism, they were tested in the coupled microplate assay. First the assay was modified to contain 10 % DMSO (to keep the compounds soluble), and the kinetic parameters of the WT epimerase retested in the new buffer. Both the K_M and k_{cat} were found to be significantly different in 10 % DMSO (Table 5). It is likely that the polar substrate is destabilized in the now hydrophobic assay buffer, enhancing its binding to the hydrophilic active site of the enzyme (lower K_M) and also slowing its release back

into solution after epimerization (lower k_{cat}). The Hill co-efficient shows that allosteric regulation is still present and a similar catalytic efficiency means that the mechanism of catalysis is likely to be the same. The lower K_M also meant that the inhibitors were tested in the presence of 0.5 mM rather than 2 mM substrate, which is the enzyme's K_M in aqueous conditions. Substrate concentrations around the K_M are usually used when testing for inhibition since the activity in this region is linear with substrate and hence most sensitive to inhibitory effects. At 2 mM substrate, the enzyme's activity in DMSO would be saturated and any inhibition masked.

Table 5: Kinetic parameters in DMSO

Parameter	0% DMSO	10% DMSO
K_m	2.19	0.43
n	1.4	1.8
k_{cat}	7.9	1.9
k_{cat}/K_m	3.6	4.3

Compounds 55, 56, 74, 75, 77, 81, 82, and 85 were then tested in the modified assay. Compounds 1 and 2 were used as negative controls (Table 6). Compounds 55 and 74 were found to be weak inhibitors, 75 a strong inhibitor, and 77 a moderate activator of epimerase activity. The latter is interesting in that it suggests that some of the compounds are actually binding the allosteric site as they were designed to do since binding to the catalytic site would explain an inhibitory effect but not an activating one.

Table 6: Inhibition of epimerase activity. Negative values indicate activation.

Compound #	% Inhibition
55	19.5 ± 4.5
74	21.1 ± 7.8
75	63.9 ±5.0
77	-36.6 ± 4.2

These 4 compounds were then retested with the order of addition to start the reaction being reversed (substrate added to a enzyme/inhibitor pre-incubation rather than enzyme to substrate/inhibitor). This was not found to change the magnitude of their effects significantly suggesting that it was competitive binding at the allosteric site and not a covalent bond being formed between the enzyme and inhibitor that was leading to inhibition. This was also shown by increasing the substrate concentration to 3 mM, which reduced the magnitude of the inhibition as would be expected with competitive inhibition. Since there is a second enzyme present in the assay it is possible that these compounds are affecting the dehydrogenase rather than the epimerase. To test this the amount of dehydrogenase in each well was doubled but everything else kept the same. This did not reduce the magnitude of inhibition, suggesting that the epimerase, and not the dehydrogenase, was being targeted by these compounds.

Based on the structures of compounds 55, 74, 75, and 77, a new batch of nine compounds were synthesized by SuperGen. These all inhibited growth of the three Gram-positive strains used and were then tested in the *in vitro* assay at concentrations of 0.1 mM, 0.25 mM, and 0.5 mM, from which their IC_{50} values were estimated (Table 7). Based on the substrate concentration of 0.5 mM, their K_i values were then calculated (assuming competitive inhibition) using the equation:

$$K_i = IC_{50}/(1 + [S]/K_M) \quad \text{(Prism 4, Graphpad Software Inc.)}$$

All of the compounds showed significant inhibition, with K_i values ranging from 70 to 280 μM.

Table 7: Estimated IC_{50} and K_i values for second round inhibitors

Compound #	IC_{50}, mM	Ki, mM
4003	0.600	0.28
4011	0.295	0.14
4023	0.250	0.12
4025	0.200	0.09
4026	0.195	0.09
4029	0.265	0.12
4030	0.150	0.07
4031	0.295	0.14
4033	0.150	0.07

The related compounds 4030 and 4033 were the best inhibitors tested (K_is of 70 μM), but both were very hydrophobic and insoluble in aqueous solution. For use under physiological conditions, a modified version of 4030 was synthesized by SuperGen to contain a negative charge to enhance its solubility (provided as a sodium salt, Na.4030). This new compound was retested in the original version of the assay (without added DMSO) and found to have an estimated K_i of 82 μM in aqueous buffer (Figure 7). It is now being tested in a mouse model of *B. anthracis* infection.

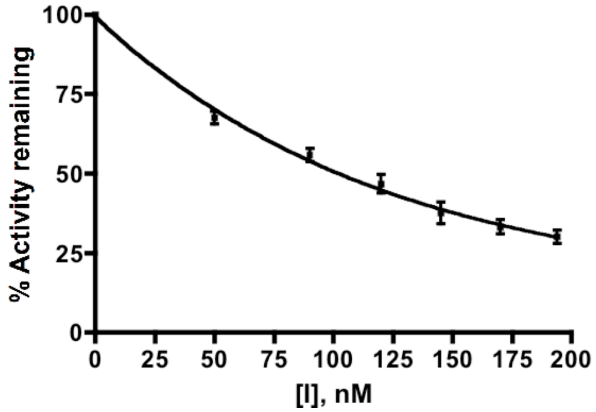


Figure 7: Inhibition by Na.4030. The compound was tested in triplicate at multiple concentrations and in the presence of 2mM substrate.

3.8 Na.4030 binding to epimerase

Although compound Na.4030 shows significant inhibition in the *in vitro* assays conducted, the assay itself contains ManNAc dehydrogenase as well as UDP-GlcNAc 2-epimerase, and it is possible that the compound is inhibiting the dehydrogenase and not the epimerase, leading to a decrease in NADH production. While previous assay controls suggested that this was not occurring, we wanted to definitively rule out this possibility. The inhibitor might also be causing generalized protein aggregation, which would lead to a decrease in overall signal as well. Several studies were conducted to show direct binding of the inhibitor to the epimerase.

First we tried to obtain crystals of the inhibitor-enzyme complex, which would prove direct binding as well as confirming that the compound bound to the allosteric site. The modified compound Na.4030 is more soluble in aqueous solution than the original compound 4030, but it still could only be dissolved to a maximum of 0.5mM. In order to get the highest concentration of inhibitor possible, it was dissolved directly into the epimerase sample rather than mixing a stock solution with the enzyme prior to setting up crystallization screens. Increasing amounts of inhibitor were slowly added to the epimerase sample (10 mg/mL, 20 mM Tris pH 8.0, 150 mM NaCl, 2 mM DTT) with shaking until it was saturated i.e. when there were visible clumps of compound remaining. Any precipitation/undissolved compound was removed by centrifugation. This complex was then extensively screened using commercially available trays at both RT and 4°C. However no reproducible hits were found.

In order to increase the inhibitor/enzyme ratio the screens were redone with samples prepared as previously but with the addition of 10 % DMSO to the protein solution before dissolving the inhibitor in it, as well as in the crystallization screening conditions (the inhibitor crystallized out by itself if DMSO was not present in the well reservoirs). Again no reproducible hits were found. It could be that inhibitor binding to just the allosteric site is not enough to stabilize the epimerase in a crystallizable conformation and that a substrate molecule is also required in the active site. The requirement of substrate for the formation of crystals of both the *B. anthracis* and *E. coli* enzymes provides some evidence for this. However, the high substrate affinity for the allosteric site makes obtaining inhibitor/epimerase complexes problematic if also in the presence of substrate.

The next method tried was to test if an inhibitor/epimerase complex would stay together on size exclusion chromatography. Na.4030 has a distinctive absorbance peak at 445nm and if the complex stays together during gel filtration it should be detectable co-eluting with the protein peak. 20 mg of epimerase (10 mg/mL, 20 mM Tris pH 8.0, 150 mM NaCl, 2 mM) in the presence on 0.5 mM Na.4030 was loaded onto an SD200 column and the protein-containing fractions tested for the presence of inhibitor. However, the inhibitor is highly hydrophobic and bound the resin at the top of the column extremely tightly (only eluted in 50% DMSO). That Na.4030 was not detected in the protein-containing fractions might have been due to it being stripped away by the column, and doesn't prove a lack of interaction. It is also possible that the binding is transient and the inhibitor continually binds/dissociates from the enzyme, in which case they might not co-

elute on gel filtration, depending on the competition between the binding/dissociation rate of the complex and the entry rate of protein into the matrix of the resin beads.

Finally we tried to see binding of Na.4030 to epimerase by isothermal calorimetry (ITC) (Roselin et al 2010). Samples were sent for testing to the High Throughput Screening Resource Center at Rockefeller University. Briefly, inhibitor present at 0.5 mM in 20 mM Tris 8.0, 150 mM NaCl, 2 mM DTT was added in 2μL increments to 0.4 mM epimerase in the same buffer. The energy required to maintain the sample cell temperature constant was measured. Again however, the inhibitor was extremely hydrophobic and bound to the tubing and injection needle of the machine, making the data extremely noisy and difficult to ascertain if binding was occurring.

Although direct binding of compound Na.4030 to the allosteric site of UDP-GlcNAc 2-epimerase has not been shown, the controls in the *in vitro* assays and the appearance of activating behavior in some of the initially tested compounds from which Na.4030 was derived provide strong evidence that its inhibitory activity in cells is through the epimerase rather than the dehydrogenase. More soluble versions of Na.4030 are now being designed (SuperGen) and might be used to eventually prove direct binding of Na.4030 to the enzyme.

CHAPTER 4 - CONCLUSIONS

In this work we were able to determine the first crystal structure of UDP-GlcNAc 2-epimerase from *B. anthracis*. The structure showed substrate molecules bound to separate active and allosteric sites in each molecule of the dimeric protein, and provided the first example of an enzyme where a substrate molecule bound at the allosteric site directly interacts with a substrate molecule bound at the catalytic site. Moreover, we were able to disprove a previous model for allosteric regulation proposed based on the structure of the *E. coli* version of the enzyme. Rather than substrate binding to one molecule of the physiological enzyme dimer conformationally activating the other subunit, we have shown that each molecule in the dimer contains separate allosteric and catalytic sites and can be activated independently of one another. Furthermore, since the catalytic molecule is trapped within the catalytic site and not fully exposed to solvent, the enzyme must convert between an open and closed form after every cycle of epimerization to release the product and allow entry of a new substrate molecule.

We have also developed a robust coupled microplate assay to measure epimerase activity quantitatively, even in whole cell lysates. Mutations in the allosteric binding site showed the importance of these residues for proper functioning of the enzyme, especially Arg 210, which makes contacts with both the allosteric and catalytic site substrate molecules. The allosteric site was then used to *in silico* screen and design possible inhibitors of the epimerase which would be active against the bacterial non-hydrolyzing version of the

enzyme but not the mammalian hydrolyzing form. After three rounds of testing against a number of Gram-positive strains and in the coupled *in vitro* assay, compound Na.4030 (K_i of 82 μM) was identified as a possible lead for further development as a therapeutic drug. It is now being tested in a mouse model of *B. anthracis*.

Although direct binding between Na.4030 and the epimerase was not shown, there is strong circumstantial evidence that the compound's observed bactericidal activity is through the epimerase and not some other enzyme or non-specific effect. More soluble and less hydrophobic versions of the compound are now being synthesized which might be able to definitively show biophysically that such an interaction is occurring.

PART B:

Salmonella Secreted Effector I

CHAPTER 5 - INTRODUCTION

5.1 *Salmonella enterica*

S. enterica is a large group of gram-negative, motile, rod-like bacteria that are found in reservoirs of reptiles, livestock animals, and humans. It is capable of causing a wide range of diseases in mammalian hosts, including an asymptomatic carrier state, a self-limiting diarrheal infection, septacemia, and fatal systemic infections (Lan et al 2009). It was first described in 1880 by Karl Eberth as the possible cause of typhoid fever, which was subsequently confirmed in 1884 when Georg Gaffky was able to culture it directly from patients. The genus is named for Frank Salmon who created the first animal health & meat safety department within the US Department of Agriculture, The Bureau of Animal Industry. His collaborator Theobald Smith isolated *Salmonella cholerae suis* as the causative agent of hog cholera in 1885 and named it in his honor (www.wikipedia.org).

There are over 2000 serovars described to date, distinguished by the reaction of specific antisera to their O (polysaccharide cell wall components), H (flagellar proteins), and Vi (capsule component) antigens. *S. enterica* serovars Typhimurium, Enteritidis, Newport, and Heidelberg are the major cause of gastroenteritis while serovars Typhi and Paratyphi lead to a more serious systemic infection that is restricted to humans (Salyers & Whitt 2002). Every year there are estimated to be 4 million cases of *Salmonella*-based food

poisoning in the US, leading to 15,000 hospitalizations, and 600 deaths, at a cost of $4 billion. Worldwide there are over 1.3 billion cases/year with higher fatality rates especially in developing nations and with HIV-infected patients. Typhoid cases are rare in the US, usually restricted to international travelers, but are more common in the rest of the world, with an estimated 20 million cases per year and over 200,000 deaths (www.who.int, www.cdc.gov, www.cfsan.fda.gov).

Apart from its disease importance, *Salmonella* has been studied extensively as a model of bacterial metabolism, genetics, virulence, and modulation of host cellular responses. It is also being investigated as a possible carrier for oral vaccines. The idea is to introduce genes coding for an antigen of interest into an avirulent strain which would then express the antigen on its surface. Bacteria would multiply in the Peyer's patches after administration, eliciting both a mucosal and humoral response against the antigen (Galen et al 2009). Its Type III secretion system (TTSS) (see Section 1.3) is also being investigated as a way of delivering protein therapeutics directly into the cytosol of cells. Moreover, the appearance of multi-drug resistant strains of *S. typhi* (in the developing world) and *S. enterica* (in industrialized nations, due to the use of antibiotics in animal feed to promote rapid growth) is becoming a cause for concern (Foley & Lynne 2008, Rowe et al 1997).

The sequencing of multiple *Salmonella* genomes shows extensive genetic degradation of the typhoidal strains *S. typhi* (210 pseudogenes) and *S. paratyphi* (173 pseudogenes) compared to *S. typhimurium* (39 pseudogenes) (McClelland et al 2004, McClelland et al

2001), which might explain its host restriction to humans. However, the reason for its invasiveness and lack of diarrheal symptoms remains unclear. One difference is that *S. typhi* produces an additional capsular polysaccharide composed of N-Acetylgalacturonic acid (Vi antigen, not present in *S. typhimurium*) that might allow it to evade the innate immune system in the intestinal epithelium and move quickly to other organs (hence not eliciting a fluid loss response) (Raffatellu et al 2006).

5.2 Pathogenesis

Non-typhoidal salmonella is transmitted by contaminated food, usually egg, meat, or poultry, although milk and even contaminated marijuana have lead to large outbreaks. Symptoms appear within 24 hrs, starting with nausea and vomiting, followed by abdominal pain, diarrhea, and sometimes fever. Usually no antibiotic treatment is required and symptoms subside in a few days, except in the case of the immunocompromised or the elderly, where bacteria can move into the bloodstream to cause septicemia. Patients continue to shed bacteria for up to three months. Vaccines are available for poultry and cattle but are usually only protective against specific strains and endemic farm infections remains a problem (Salyers & Whitt 2002).

Typhoid fever has a much longer incubation period of 1 to 4 weeks. After ingestion from contaminated food or water, bacteria survive within intestinal macrophages and migrate to the spleen and liver through the bloodstream and lymphatic system. After rapid multiplication large numbers of free bacteria are released back into the bloodstream. This

causes a high fever, chills, convulsions, and delirium. The bacteria also move from the liver to the gall bladder and are shed back into the intestine and can cause severe ulceration (Salyers & Whitt 2002). Fatality can be as high as 20% if not treated by antibiotics such as ampicillin, trimethoprim-sulfamethoxazole, and ciprofloxacin. Travellers are advised to be vaccinated using either a live attenuated strain (Ty21a, Vivotif Berna™) or the Vi capsular polysaccharide vaccine (or ViCPS, Typhim Vi™). In some people, *S. typhi* can persist in the gall bladder asymptomatically and be shed for years, as was the case with Mary Mallon ('Typhoid Mary') who worked as a cook in New York City in the 1900s and infected hundreds of people.

At the molecular level, every aspect of the multi-stage *Salmonella* infection is controlled by a large array of virulence effectors delivered directly into host cells by two separate Type III secretion systems, encoded by pathogenecity islands 1 & 2 (Abrahams & Hensel 2006, Haraga et al 2008, Lilic & Stebbins 2004, McGhie et al 2009). TTSS-1 is induced extracellularly and is activated either by contact with host cells or by environmental cues in the intestine. Its effectors are transported across the host cell membrane and act to facilitate membrane ruffling and invasion of both M cells and enterocytes. They also induce the inflammatory and secretory responses – a mixture of neutrophil influx, mucosal damage, and increased Cl^- secretion, which leads to massive fluid loss (Zhang et al 2003). Once internalized the bacteria is enclosed within a phagosomal compartment called the *Salmonell*a-containing vacuole (SCV), which leads to expression of TTSS-2 in response to cues such as low Mg^{2+} and low pH. Its effectors are transported across the SCV membrane and interact with endocytic pathways to maintain the SCV during

Salmonella replication and subsequent release back into the intestinal lumen. It also plays a role in modulation of the immune system and systemic infection.

Salmonella has been shown to contain three other TTSSs (SPI 3-5) that have an effect on virulence, but SPI-1 and SPI-2 seem to be the major determinants of pathogenicity (Rychlik et al 2009).

5.3 Type III secretion system

Type III secretion systems (injectisomes) have been found in over 25 species of Gram-negative bacteria that have either a symbiotic or pathogenic relationship with plant and animal hosts, such as *Salmonella*, *Yersinia*, *Bordetella*, *Shigella*, *E. coli*, *Chlamydia*, *and Pseudomonas syringae* (Cornelis 2006, He et al 2004). They are evolutionarily related to the flagellar apparatus (Saier 2004) and are carried on pathogenicity islands encoding all the proteins needed for complex assembly (>20) and many of the translocated effectors, even though effectors in other parts of the genome can be integrated into the system by transcriptional control. The effectors themselves tightly modulate host cell responses during the infection and many have been shown to be mimics of host proteins in structure and biochemical function (although not necessarily at the sequence level), representing both modifications of genes that have been acquired by horizontal transfer and those created by convergent evolution (Stebbins & Galan 2001).

The TTSS of *S. typhimurium* has been well studied and is used to show general principles of injectisome function (Figure 8) (Galan & Wolf-Watz 2006, Marlovits & Stebbins 2010). The main transport machinery is called the needle complex, through which effectors move past both the bacterial and host cellular membranes. It is composed of a basal structure and a needle-like projection that sticks out for ~50nm past the cell wall. The base is made of inner and outer rings (with ~20-fold symmetry) that are anchored to the bacterial inner and outer bacterial membranes respectively. There is also a cylindrical inner rod inside the base that traverses it and connects to the needle (Marlovits et al 2004).

In *S. typhimurium*, the base is created first by three proteins that are targeted to the bacterial membrane by the *Sec* system: InvG forms the outer ring, PrgH and PrgK the inner ring. This structure then is able to secrete proteins that are needed to assemble the inner rod (PrgJ) and needle (PrgI). InvJ is a regulatory protein that is thought to stabilize the bottom of the inner rod during its formation and completion of inner rod assembly is a signal to switch substrates to translocated effectors. PrgI also polymerizes to form the needle at the same time as PrgJ assembly, and needle length has been shown to be coupled to the rate of formation of the inner rod (Marlovits et al 2006).

Although not confirmed by electron microscopy in *S. typhimurium*, based on homology to the flagellum, a further structure called the C-ring is thought to form on the cytoplasmic side of the inner ring of the basal body by the Spa family of proteins (Morita-Ishihara et al 2006). These recruit further accessory proteins and an ATPase

(InvC, and associated Org family proteins), which are necessary for transport. Although competent for effector secretion into the extracellular media, translocation past the host cell membrane requires the formation of a translocon made of three secreted proteins that insert themselves into the host membrane and bind to the tip of the needle (SipB and SipC form the pore, and SipD contacts the needle) (Deane et al 2010, Mueller et al 2008).

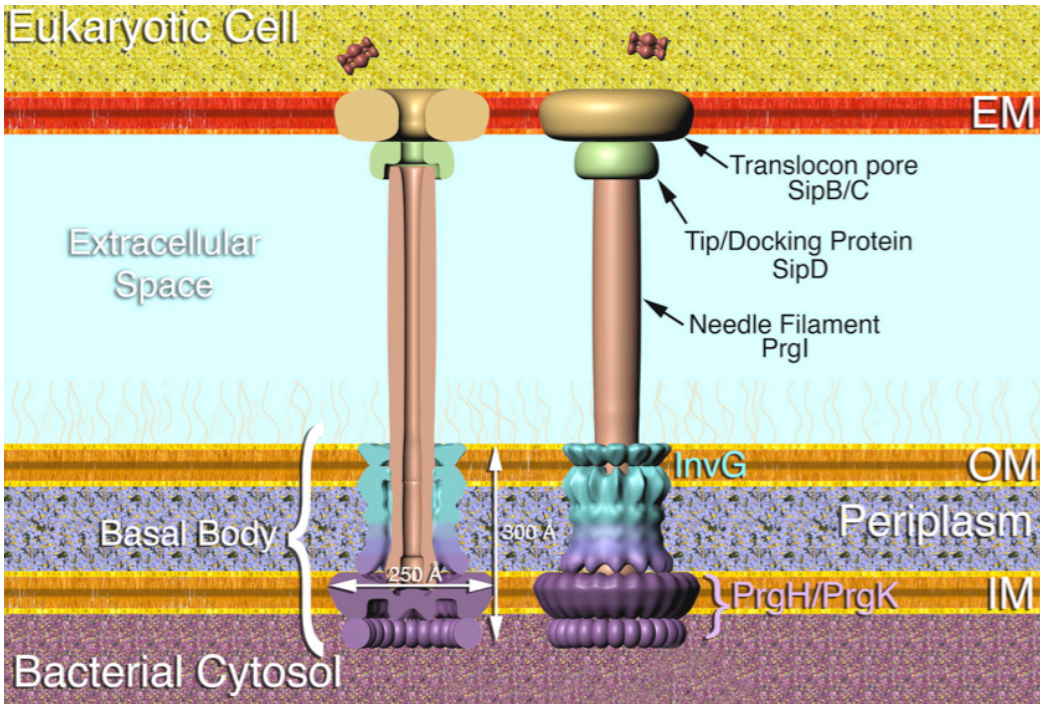


Figure 8: Schematic of the injectisome from *S. typhimurium*. (Figure 1 in Marlovits & Stebbins 2010)

Effector targeting to the TTSS apparatus and consequent secretion is a well-regulated process. The secretion signal is in the first 20-30 amino acids (not cleaved off, unlike the *Sec* system), but in the absence of certain small, acidic, homodimeric accessory proteins called chaperones, secretion is significantly less efficient. If the chaperone binding domain is deleted but the secretion signal kept, transport through the TTSS-related flagellar system instead has been observed, suggesting that the secretion signal is a

'generalized' TTSS/flagellar export signal (Lee & Galan 2004). Chaperones bind to a ~50-100 amino acid domain downstream from the secretion signal and stabilize their cognate effectors in the cytoplasm, regulate when the effectors are 'seen' by the secretion machinery, and maintain the effectors in a partially unfolded form that primes them for transport. The needle's inner diameter is ~28 Å, making it necessary for the effector to fully unfold before passing through it (Marlovits et al 2004). The energy for this unfolding process and the separation of the effector-chaperone complex comes from an ATPase (InvC) present at the base of the needle (Akeda & Galan 2005).

For SPI-1, most chaperones identified have a single effector substrate (Class 1A chaperone) but InvB can bind multiple substrates (Class 1B) (Parsot et al 2003). Structures of chaperone-effector complexes show that the dimeric chaperones usually bind a single effector molecule at two distinct sites stretching out the N-terminal region of the molecule into an unfolded state, and although not obvious by sequence, a conserved structural motif for chaperone binding has been proposed (Lilic et al 2006).

SPI-2 is interesting in that only chaperones for its injectisome components have been identified within it, and none for its encoded effectors. Moreover, a highly homologous ~140 amino acid N-terminal domain is present in all seven of its secreted effectors that are encoded outside the SPI-2 island (SspH1, SspH2, SlrP, SifA, SifB, SseI, SseJ) (Miao & Miller 2000). Called an STE domain, it is sufficient for transport through the TTSS-2. It has been proposed that having only a single domain necessary for localization and export through the TTSS-2 allows the bacteria to evolve at much quicker rates by

'terminal reassortment', a process whereby new effectors can be added to its arsenal by genetic rearrangements in which the addition of one element fully integrates an acquired gene into the TTSS transcriptional and transport systems (Stavrinides et al 2006).

5.4 *Salmonella* Secreted Effectors

The effectors translocated by TTSS-1 and TTSS-2 have been extensively studied and much progress made in determining their targets and biochemical functions (Abrahams & Hensel 2006, Haraga et al 2008, McGhie et al 2009). They are summarized in Tables 8 & 9. Briefly, the effectors first act to induce membrane deformation and cytoskeletal rearrangements, triggering bacterial internalization, after which the cytoskeleton is returned to its resting state. They then help the SCV mature and migrate to the perinuclear region. Once the SCV is positioned, bacteria begin to replicate and Sifs (*Salmonella* induced filaments), a mixture of vesicles and tubulin filaments, are induced to form around the SCV, diverting host nutrients needed for bacterial replication. During this time the inflammatory response is also stimulated leading to the diarrheal symptoms. There is also evidence that cell death and eventual bacterial escape is also under effector control.

Table 8: Effectors secreted by *Salmonella* TTSS-1. Adapted from Table 1 (McGhie et al 2009).

Effector	Gene Location	Activity	Host cell target(s)	Role in infection
AvrA	SPI-1	Cysteine protease with deubiquitinase activity, acetyltransferase	MKK4/7 IκBα, β-catenin	Inhibits inflammation, represses apoptosis & epithelial innate immunity, stabilises tight junctions
SipA (SspA)	SPI-1	Actin binding / stabilising	Actin, T-plastin	Increases internalisation efficiency, enhances actin assembly, potentiates SipC activity, triggers PMN transmigration, maintains perinuclear SCV positioning, disrupts tight junctions
SipB (SspB)	SPI-1	SPI-1 TTSS translocon component	Cholesterol	SPI-1 effector delivery, apoptosis of phagocytes
SipC (SspC)	SPI-1	SPI-1 TTSS translocon component, actin nucleation & bundling	Actin	SPI-1 effector delivery, induces membrane ruffling
SipD (SspD)	SPI-1			Regulates SPI-1 effector secretion
SopA	Outside SPI-1	E3 ubiquitin ligase	HsRMA1	Disrupts SCV integrity, induces PMN transmigration
SopB (SigD)	SPI-5	Inositol polyphosphate phosphatase	Inositol phosphates	Promotes membrane fission & macropinosome formation, maintains perinuclear SCV positioning, promotes epithelial cell survival, triggers nitric oxide production in macrophages, promotes fluid secretion, disrupts tight junctions
SopE	Bacteriophage SopEϕ	Guanine exchange factor (GEF) mimic	Rac-1, Cdc42	Induces membrane ruffling & proinflammatory responses, promotes fusion of SCV with early endosomes, disrupts tight junctions
SopE2	In vicinity of bacteriophage remnants	GEF mimic	Cdc42	Induces membrane ruffling & proinflammatory responses, increases macrophage iNos expression, disrupts tight junctions
SptP	SPI-1	GTPase activating protein (GAP) mimic, tyrosine phosphatase	Cdc42, Rac-1, vimentin	Returns host cytoskeleton to resting state following bacterial entry, downregulates proinflammatory responses
SlrP*		Ubiquitin ligase?		Confers host specificity?
SopD*	Outside SPI-1/SPI-2			Promotes membrane fission & macropinosome formation, contributes to *Salmonella* virulence and persistence in mice, induces fluid secretion, promotes invasion of T84 cells
SspH1*	Bacteriophage Gifsy-3	E3 ubiquitin ligase	PKN1	Downregulates proinflammatory responses
SteA* (STM1583)	Outside SPI-2			Required for efficient mouse spleen colonisation
SteB* (STM1629)	Outside SPI-2	Putative picolinate reductase		

* Can also be translocated via TTSS-2

Table 9: Effectors secreted by *Salmonella* TTSS-2. Adapted from Table 2 (McGhie et al 2009).

Effector	Gene Location	Activity	Host cell target(s)	Role in infection
GogB	Bacteriophage Gifsy-1			
PipB	SPI-5			
PipB2	Outside SPI-2		Kinesin-1	Promotes Sif extension, recruits kinesin-1 to SCV
SifA	Outside SPI-2	Rab mimic?	SKIP, Rab7/9	Required for SCV membrane integrity & Sif formation, maintains perinuclear SCV positioning, redirects exocytic vesicles to SCV
SifB	Outside SPI-2			
SopD2	Outside SPI-2			Contributes to Sif formation, required for efficient bacterial replication in macrophages & mice
SpiC (SsaB)	SPI-2		Hook 3, TassC	Interferes with vesicular trafficking, role in SCV-associated actin polymerisation (VAP) and Sif formation, controls order of protein export through SPI-2 TTSS
SseF	SPI-2			Contributes to Sif formation, recuits dynein to SCV, maintains perinuclear SCV positioning, required for formation of microtubule bundles around SCV, redirects exocytic transport vesicles to SCV
SseG	SPI-2			As SseF
SseI (SrfH/ GtgB)	Bacteriophage Gifsy-2		Filamin, TRIP6	Remodels SCV associated F-actin? Increases cellular detachment, inhibits directed cellular migration
SseJ		Deacylase, phospholipase A & glycerol-phospholipid :cholesterol acyltransferase	Cholesterol	Negative regulation of Sifs, antagonises SifA SCV stabilisation
SseK1	Outside SPI-2			
SseK2	Outside SPI-2			
SseK3* (NleB)	ST64B coliform bacteriophage			
SseL	Outside SPI-2	Cysteine protease with deubiquitinase activity	IκBα	Macrophage apoptosis, downregulates inflammatory responses
SspH2	In vicinity of bacteriophage remnants	Inhibits actin polymerisation *in vitro*	Filamin, Profilin	Remodels SCV associated F-actin? E3 ubiquitin ligase
SteC (STM1698)	Outside SPI-2	Serine/Threonine kinase		Required for VAP
SpvB[#]	pSLT (*S.typhimurium*)	ADP ribosyl transferase, inhibits actin polymerisation *in vitro*, depolymerises F-actin upon transfection	Actin	Inhibition of VAP, apoptosis of infected cells, required for full virulence in mice
SpvC[#]	pSLT (*S.typhimurium*)	Phosphothreonine lyase		Required for full virulence in mice

* Found in *S.typhimurium* SL1344, not LT2

[#] Found in non-typhoid *Salmonella* serovars

5.4 *Salmonella* secreted effector I

SseI (322 aa, 36.8k Da, pI 5.12), also called SrfH, is a TTSS-2 effector that has been directly linked to systemic infection. It is present in a chromosomal prophage, Gifsy-2, that is located outside the SPI-2 island (gene STM1051 in *S. typhimurium*). However, it is still under the control of the SPI-2 transcriptional regulator SsrB, allowing its expression and secretion in parallel with the other SPI-2 encoded effectors. During infection, SsrB increases SseI's expression 78-fold in macrophages and 166-fold in the spleen (Worley et al 2000). Its N terminus contains an STE domain that is found in all SPI-2 effectors present outside SPI-2 and is thought to allow localization and transport through TTSS-2 without need for a cognate chaperone (Miao & Miller 2000). The Gifsy-2 phage is inducible (Figueroa-Bossi & Bossi 1999) and an analysis of SseI distribution among *S. enterica* strains showed that it is only present in a few classes, raising the possibility that it is a recent addition to the SPI-2 effector repertoire and is still spreading through the genus (Hansen-Wester et al 2002). SseI's C terminus does not show significant sequence homology to any other protein.

SseI has been shown to be actively translocated across the SCV membrane and localize to regions of active actin polymerization through an interaction of its N terminus with filamin (a dimeric protein that cross-links F-actin) (Miao et al 2003). However its exact role in pathogenesis was unclear until a genome-wide screen in a persistent mouse model (129X1/SvJ, bacteria can remain for up to a year) of *S. typhimurium* infection identified it as being necessary for systemic infection, not emerging from the screen until two weeks

post-infection (Lawley et al 2006). In normal mouse models (BALB/c and C57Bl/6) *S. typhimurium* behaves like *S. typhi* in humans and leads to death from systemic infection within 7-10 days (Tierrez & Garcia-del Portillo 2005).

A yeast two-hybrid screen identified TRIP6 (Thyroid receptor interacting protein 6) as a possible interaction partner (Worley et al 2006). TRIP6 is an adaptor protein that binds components of the Rac1 signaling pathway, and is critical for cell motility and the NF-κB inflammatory pathways (Li et al 2005, Xu et al 2004). Transfection of RAW276.7 macrophage-like cell lines showed co-localization of TRIP6 and SseI but the interaction was not confirmed biochemically. They were also able to show that both RAW276.7 cells and JAWS dendritic cells migrated faster towards a chemoattractant than mutant strains in a Boyden chamber. Furthermore, wild-type (WT) bacteria present within $CD18^{+}$ macrophages were found in the bloodstream in significantly greater numbers than a SseI mutant strain 30 mins after oral infection. These results led the authors to postulate that SseI increased cellular migration, allowing macrophages infected with bacteria to move actively into the bloodstream and to the other organs, rather than passively through the lymphatic system, which can take 12 - 20 hrs (MacPherson et al 1995)

Recently, pull-downs with primary macrophage lysates identified IQGAP1 as another binding partner (McLaughlin et al 2009). IQGAP1 is a large scaffolding protein that binds actin and several small G proteins, playing a role in cellular motility and its directionality (Brown & Sacks 2006). This interaction was also shown biochemically

with purified IQGAP1 and SseI, as well as during infection of primary bone marrow derived macrophages (BMDM). Transfection of BMDMs with SseI showed its co-localization with IQGAP1 and polymerizing actin at the cell periphery. However, the previously reported interaction with TRIP6 could not be confirmed in these cells, perhaps due to different expression patterns than RAW264.7 cells. When they looked at cellular migration, surprisingly they found that SseI inhibited the directed migration of BMDM cells, causing cells to frequently reverse their direction of movement. Furthermore, dendritic cell (DC) migration was also inhibited *in vivo*. When DC cells were infected with GFP-labelled *Salmonella* and injected intraperitoneally, WT strains showed lower DC numbers in the spleen after 6 hrs compared to mutant strains, which also correlated with lower numbers of activated $CD4^+$ T cells. This led them to postulate that inhibiting immune cell migration in areas of infection allows the bacteria to evade clearance, leading to systemic infection. Further investigation revealed that the increased motility phenotype seen previously with RAW276.7 cells (Worley et al 2006) was actually due to increased detachment from the extracellular matrix in these types of cells, rather than increased migration. This would also explain why infected macrophages were seen within 30 mins in the bloodstream, since a decreased cellular adherence to the extracellular matrix would allow them to quickly move through the layers of the gut epithelium and into the vascular system.

5.5 Aims

The drastic effect of SseI on *S. typhimurium* pathogenesis, converting it from a gut-localized infection to a serious systemic infection, makes it an effector of considerable interest. However, its molecular functions and targets remain uncharacterized. We took a structural approach to studying this enzyme, attempting to solve a crystal structure of its C-terminal catalytic domain to find close structural homologs that would give clues as to its biochemical function(s) and substrate(s). Any structure-based hypotheses would then be tested both *in vitro* and in live cells.

CHAPTER 6 - MATERIALS AND METHODS

6.1 Gene constructs

SseI (1-322) was cloned from *S. typhimurium* strain LT2 genomic DNA into a modified pCDF-Duet (Novagen) vector containing two hexahistidine tags followed by a rhinovirus 3C protease cleavage site before the protein. Subsequent constructs (1-125, 1-138, 24-125, 24-134, 137-289) were sub-cloned from this original plasmid.

Human Cdc42 (1-184) and RhoA F25N (1-181) were obtained from G. Prehna (Laboratory of Structural Microbiology, Rockefeller) as N-terminal GST fusions in a pGEX-4T-3 (GE Healthcare) vector with an added 3C cleavage site between the tag and protein. Rac1 (1-184) was also received as a GST fusion, but with a thrombin cleavage site instead. RhoA F25N was used because of its enhanced solubility and stability during purification. It has been shown to behave similarly to WT RhoA (Self & Hall 1995) and is referred to as RhoA henceforth.

Human TRIP6 (1-476) was cloned from HeLa cell cDNA into both the modified pCDF-Duet and pGEX-4T-3 vectors. A smaller construct (276-476) was further subcloned from this plasmid. IQGAP1 (1-1657) was obtained from D. Sacks (Department of Pathology, Brigham and Women's Hospital, Harvard Medical School) as an N-terminal GST fusion containing a thrombin cleavage site in a pGEX-2T vector.

Both SseI C178A and Cdc42 N61E point mutants were generated using a Quick Change site directed mutagenesis kit (Stratagene).

6.2 Protein purification

All constructs were transformed into the BL21 (DE3) strain and grown in LB medium at 37ºC to stationary phase followed by induction with 0.1 mM IPTG at 20ºC overnight. The cells were harvested by centrifugation, re-suspended in 50 mM Tris pH 8.0, 10 mM Imidazole, 200 mM NaCl, 1 mM PMSF and lysed using a homogenizer (Avestin). To digest DNA and decrease viscosity, 0.1 mg/ml DNaseI and 5 mM $MgCl_2$ were added to the lysate.

After centrifugation the lysates from His-tagged constructs were loaded onto a gravity flow column containing Ni-NTA resin (Qiagen) equilibrated in the lysis buffer (without PMSF). This was followed by a 10 column volume (CV) wash (lysis buffer + 30 mM Imidazole) and protein elution with 4 CV of elution buffer (lysis buffer + 250 mM Imidazole). GST-tagged lysates were loaded onto Glutathione-sepharose resin (GE Healthcare) equilibrated in 50 mM Tris pH 8.0, 200 mM NaCl, 5mM DTT, washed with equilibration buffer + 500 mM salt, and eluted in equilibration buffer + 10 mM reduced glutathione. The Rho GTPases were purified with additional 10 mM $MgCl_2$ in all buffers.

Following overnight dialysis into 50 mM Tris pH 8.0, 200 mM NaCl, 2 mM DTT and cleavage with 3C protease or thrombin (1:75), the proteins were run back over either the Ni-NTA or Glutathione resins to remove the tags. The flow-through was then concentrated and further purified by gel filtration chromatography on Superdex 75 & 200 columns (GE Healthcare) into 20 mM Tris pH 8.0, 200 mM NaCl, 2 mM DTT. The purified proteins were stored at -80ºC.

SeMet substituted SseI 137-322 was made by expression in the Met $^-$ *E. coli* strain 834. Cells were grown for 6 hrs at 37ºC after induction and the protein purified as before except that 10mM DTT rather than 2 mM was used in all the buffers to prevent Se oxidation.

6.3 Limited proteolysis

Purified protein at ~10mg/mL (+ added 5mM $CaCl_2$) was incubated for 30 mins at RT and 4°C with increasing ratios (1:10 to 1:1000) of the non-specific protease subtilisin (Sigma). The reactions were stopped by addition of SDS-PAGE sample buffer and heating at 95°C for 10 mins. The samples were run out on SDS gels and transferred onto a PVDF membrane in the presence of 10 mM CAPS, 10% methanol, pH 11.0 transfer buffer at 60 mV overnight. Protein bands were visualized by Ponceau (Sigma) staining and bands corresponding to stable fragments persisting during proteolysis were sent to the Proteomics Resource Center at Rockefeller for Edman N-terminal sequencing (data reported for the first 5 residues).

6.4 Reductive methylation

Purified SseI 137-322 was dialyzed into 50 mM HEPES pH 7.6, 150 mM NaCl and diluted down to 0.4 mg/mL (300 mL). 40 μL of 1M Formaldehyde (methanol free) and 20 μL of 1 M Dimethyl-amine-borane complex (DMBA, Aldrich) was slowly added per mL of protein. After shaking for 2 hrs in the dark at 4°C, the previous step was repeated. This was followed by the addition of a further 10 μL of DMBA per mL of protein and shaking overnight for 18 hrs. The reaction was quenched by addition of 25 mL of 1 M $(NH_4)_2SO_4$, after which the protein was concentrated and exchanged into 20 mM Tris pH 8.0, 2 mM DTT over an SD75 (GE Healthcare) column. Samples were sent for total mass MALDI-TOF analysis at the W.M. Keck Foundation Biotechnology Resource Laboratory at Yale University to determine the total amount of lysine modification.

6.5 Crystallization

Crystallization screens were conducted with SseI constructs 1-125, 1-138, 24-125, 24-134, 137-289, and 137-322, using 1:1 sitting drops at RT and 4°C. SseI 1-125, 1-138, 24-125, and 137-289 did not lead to repeatable hits. SseI 24-134 (~15 mg/mL) gave fine clusters of crystals in 1.8 M Na/K phosphate pH 6.9 (RT), but these could not be further optimized to give single diffracting crystals. The complete catalytic C-terminal domain SseI 137-322 (~20 mg/mL) gave large rectangular crystals in 1.15 M $LiSO_4$, 0.1 M NaOAc, pH 4.6 (4°C) when optimized (1:1 hanging drops). However these were very

fragile and consisted of multiple lattices overlaid on each other as observed in their diffraction patterns.

SseI 137-322 was then reductively methylated at its lysines and rescreened. It gave multiple hits in a number of conditions containing MES, Tris, and HEPES buffers pH 6.0 - 9.0, and 1.6 to 2.0 M $(NH_4)_2SO_4$. They ranged from bowties to spherical clusters of needles. To obtain single crystals we microseeded crushed crystals from 0.1 M MES pH 7.1, 1.8 M $(NH_4)_2SO_4$, into pre-equilibrated (2 hrs) drops containing protein/reservoir solution (0.1 M MES pH 5.5, 1.6 M $(NH_4)SO_4$). This procedure gave single rectangular crystals but they were still very small. To obtain larger crystals we macroseeded single crystals obtained in the previous step again into pre-equilibrated drops, giving large, thin rectangular plates that were well-diffracting. Crystals of SeMet-substituted methylated 137-322 were obtained under the same conditions as the native crystals but did not require seeding to give large single diffracting crystals. Crystals were then cryoprotected by transfer in small increments into mother liquor supplemented with 25% glycerol.

6.6 Structure determination

Data was collected at beamlines X3A and X29 at the NSLS (National Synchroton Light Source) and the structure solved by single-wavelength anomalous dispersion (SAD) (Laskowski 2009) to 1.70 Å using the Se peak (0.97900 Å) of the SeMet substituted crystals. Crystals belonged to space group $P4_1\,2_1\,2$ (a = b = 52.256 Å, c = 132.146 Å, $\alpha = \beta = \gamma = 90°$), with a Mathews coefficient of 2.1 $Å^3$/Da (42 % solvent) and one monomer

in each asymmetric unit. Data was processed using HKL2000 (Otwinowski 1997) and phased using PHENIX (Adams et al 2010). A preliminary model was built into the obtained electron density by Arp/Warp (Perrakis et al 1999) and then completely built by repeated cycles of model-building in COOT (Emsley & Cowtan 2004) followed by refinement in REFMAC5 (Murshudov et al 1997). The final model contains 166 amino acids spanning residues 145 – 313, and 192 water molecules, with an R/R_{free} of 20.6/24.2. There was no density observed for 12 residues at the N terminus (137-144 and 4 vector encoded residues), 9 residues at the C terminus (314-322) and residues 264-266 (loop). All figures were generated using PyMol (www.pymol.org).

The native crystals only diffracted to 2.05 Å, and belonged to space group $P2_1\ 2_1\ 2_1$ (a = 54.048 Å, b = 63.189 Å, c = 110.618 Å, $\alpha = \beta = \gamma = 90°$), with a Mathews coefficient of 1.90 $Å^3$/Da (35% solvent). Data was processed using HKL2000 (Otwinowski 1997) and phased by molecular replacement in PHASER (McCoy et al 2005) using the SeMet-solved structure as a starting model. The solved model contained two molecules of the protein per asymmetric unit, and was further refined in REFMAC5 (Murshudov et al 1997), giving an R/R_{free} of 22.5/27.5.

6.7 Protease assays

Two separate assays from Invitrogen were used to check for protease activity. The EnzChek Peptidase/Protease Kit (E33758) uses a mixture of small peptides that are labeled at one terminus with a fluorophore and a quencher at the other. Cleavage results

in separation of the fluorophore from the quencher and an increase in fluorescence (excitation 502 nm, emission 528 nm). Samples were tested at concentrations between 0.04 and 4 mg/mL and were added in a 50 μL volume to 50 μL of substrate in 10 mM Tris pH 7.8. The EnzChek Protease Kit (E6638) uses casein modified with the green dye BODIPY FL. Cleavage results in the release of highly fluorescent peptides (excitation 505 nm, emission 598 nm). Samples were tested at concentrations between 0.12 and 12 mg/mL and were added in a 50 μL volume to 50 μL of substrate in 10 mM sodium bicarbonate pH 7.8. Papain (Sigma) was used as a positive control with both assays. The effects of added 10 mM DTT and 0.2 mM E-64 (a cysteine protease inhibitor) (Sigma) were also tested.

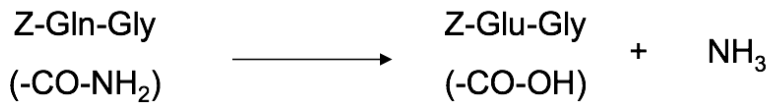
6.8 Ammonia release assay

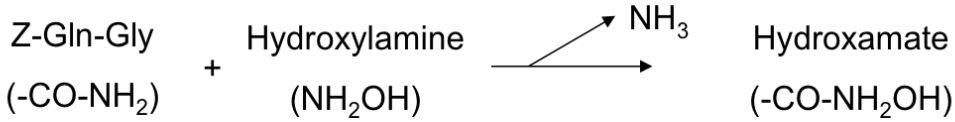
Ammonia released by modification of Gln residues was detected using a kit from Sigma. Ammonia reacts with α-ketoglutaric acid (3.4 mM) and NADPH (0.23 mM) in the presence of L-glutamate dehydrogenase (~12U) to form L-glutamate and oxidized $NADP^+$, which can be measured by the decrease in absorbance at 340 nm. In the cuvette version of the assay 100 μL of sample was added to 1 mL of test solution while in a modified microplate version, 25 μL of sample was added to 250 μL of test solution.

6.9 Deamidase/Transglutaminase assay

For Gln deamidation, 30 mM Z-Gln-Gly (Sigma) in 200 mM HEPES pH 7.2, 10 mM $CaCl_2$, 10 mM DTT was mixed with 9 μM enzyme and incubated at 37°C. 100 μL aliquots were taken out at certain intervals and added to 20 μL of 80% TCA to stop the reaction. These samples were then centrifuged to remove precipitated protein and the supernatant tested in the ammonia assay described previously. Both full length SseI (36 μg/mL) and 137-322 (18 μg/mL) were tested.

For transglutamination, 100 mM hydroxylamine was also present in the reaction mix and ammonia release detected as above. Additional 100 μL aliquots at each time point were also taken out and added to 100 μL of stop solution (0.37 M $FeCl_3$, 0.67 M HCl, 20 % TCA), which detects the presence of hydroxamate formation between Gln and NH_2OH. Samples were centrifuged and the absorbance of the supernatant at 525 nm read. Pig liver transglutaminase (Sigma) at 3 μM was used as a control in all the assays.

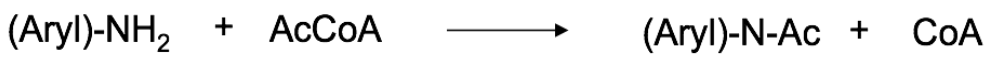




N,N-dimethyl casein (Sigma) (that was methylated at its lysines to prevent cross-linking by and protein precipitaion during transglutamination) was used as a further substrate for deamidation and transglutamination. 1 mM casein in 20 mM Tris, pH 8.0, 50 mM NaCl was incubated with 9 μM enzyme at 37°C. Aliquots were taken out and tested for ammonia release as before. The hydroxamate assay could not be used with casein as a substrate since the stop solution contains TCA which precipitates protein and in this case would also precipitate any modified casein, unlike the small-molecule substrate used previously.

6.11 N-acetyl transferase assay

The acetlyation of arylamines results in hydrolysis of Acetyl Coenzyme A to give free CoA which has an exposed sulfhydryl group. This –SH group can be detected by 5,5"-dithio-bis(2-nitrobenzoic) acid (DTNB). 10 μL of enzyme was added to 90 μL of 20 mM Tris pH 8.0, 400 μM AcCoA (Sigma) and 500 μM arylamine substrate at 37°C. After 1 hr, 25 μL of 6.4 M Guanidine HCl, 100 mM HEPES pH 7.2, 5 mM DTNB (Sigma) was added to the reaction and the absorbance at 405 nm read. Four different arylamine substrates were tested: Sulfamethazine, 4-aminobenzoic acid, 4-anisidine, and Isoniazid. Both FL (4 μg) and C-terminal SseI (2 μg) were tested and human cytosolic NAT2 (Sigma) was used as a positive control. Note that the enzymes had to be transferred into buffer without DTT using a PD10 column before use in the assay.



Another way of following NAT activity is to look for the disappearance of free-amines as the substrate is acetylated. Reactions were set up as above but the reaction mix also contained 10 mM LiK Acetyl phosphate (Sigma) and 32 U phosphotransacetylase (Sigma). The LiK acetyl phosphate and phosphotransacetylase recycle AcCoA in order to drive the reaction forward. 100 μL of 20 % TCA was added to the samples after 1 hr to stop the reaction instead of DTNB. They were then centrifuged and 100 μL of the supernatant added to 200 μL of 5 % (w/v) dimethylaminobenzaldehyde (DMAB). The absorbance at 450 nm was read.

6.12 Rho GTPase loading

Purified Cdc42, Rac1, and RhoA were loaded with GTP as follows. Protein at ~5 mg/mL was incubated with 10 mM EDTA, and 10 mM GTP for 30 mins at RT and then $MgCl_2$ added to a final concentration of 30 mM. After 30 mins the sample was exchanged into 20 mM Tris pH 8.0, 200 mM NaCl, 2 mM DTT, 2 mM $MgCl_2$ using a PD-10 column (GE Healthcare). Loading with non-hydrolyzable GTPγS was done using alkaline phosphatase to degrade GDP to GMP, for which the Rhos have a very low affinity, allowing exchange at lower nucleotide concentrations (Smith & Rittinger 2002). 10 mg of protein (1 mL) in 20 mM Tris pH 8.0, 200 mM $(NH_4)SO_4$, 10 μM $ZnCl_2$, 1 mM GTPγS was incubated with 30 U of agarose-bound alkaline phosphatase (Sigma) for 2 hrs. This was followed by removal of the alkaline phosphatase by centrifugation and

buffer exchange into 20 mM Tris pH 8.0, 200 mM NaCl, 2 mM DTT, 2 mM $MgCl_2$ using a PD10 column.

6.13 Testing of Rho GTPase deamidation/transglutamination

Deamidation/transglutamination of the Rhos were tested in a coupled assay. To start the reaction 5 μL of enzyme was added to 50 μL of either Cdc42, Rac1, or RhoA, and 35μL of a mix containing α-ketoglutaric acid (3.4 mM), NADPH (0.23 mM), and L-glutamate dehydrogenase (~12U). Ammonia release from either deamidation or transglutamination (additional 2 mM NH_2OH present in the well) reactions was followed by fluorescence at 460 nm (excitation at 340 nm). Both FL and C-terminal SseI, and GDP and GTPγS forms of the Rhos were tested. Proteins concentrations used were: SseI 1-322 and 137-322 (200 μM), Cdc42 (194 μM), Rac1 (300 μM), and RhoA (120 μM). 50 μL of 100 μM $(NH_4)SO_4$ was added to just the assay mix without any Rhos as an internal control for the magnitude of signal expected. Reactions were followed for 30-60 mins at 37°C.

6.14 Co-expressions of SseI and RhoGTPases

Full-length SseI and the full-length C178A mutant were co-expressed separately with Cdc42, Rac1, and RhoA. Cells were grown and induced as before and the Rhos were then purified out using the normal GST purification protocol. 3C protease was then added (1:75) and the proteins incubated overnight at 4°C before running them on a SDS-

PAGE gel. The gel was stained with Commassie Blue and the bands corresponding to each Rho (with tag cleaved off) was sent for trypsin-digested peptide mass analysis by ESI at the Proteomics Resource Center at Rockefeller. Another co-expression was done with Cdc42 and SseI 137-322 (as well as its mutant), but in this case the purified GST-fusion protein was sent for analysis without tag cleavage.

6.15 Native gel electrophoresis

Rho samples from the co-expression experiments were treated with 3C protease and the GST tag removed by binding to Glu-sepharose. These were then analyzed by native gel electrophoresis to check if their mobilities had changed upon co-expression with SseI. 12% and 15% acrylamide running gels were used and run at 25 mA for 2 hrs. Unmodified Cdc42 and Cdc42 N61E were used as standards.

CHAPTER 7 - RESULTS AND DISCUSSION

7.1 Domain determination

As described previously, SseI shows homology at its N terminus to the STE translocation domain of TTSS-2 secreted effectors (Miao & Miller 2000), but does not have significant sequence homologs to its putative catalytic C terminus. To determine its domain structure for further crystallographic analyses we used secondary structure prediction (Rost et al 2004) to design two constructs, 1-138 and 137-322, that were both very soluble. Together with full-length 1-322, these were subjected to limited proteolysis (Koth et al 2003) with the non-specific protease subtilisin and stable fragments persisting during digestion identified using SDS-PAGE and Edman Sequencing (Figure 9 & 10). Constructs 1-125, 24-125, 24-134, and 137-311, were then made based on these results and tested for solubility to further delineate domain boundaries. The protein appears to have three distinct regions: a secretion signal from 1-24, the STE domain from 24-130 and a single catalytic domain from 137-322.

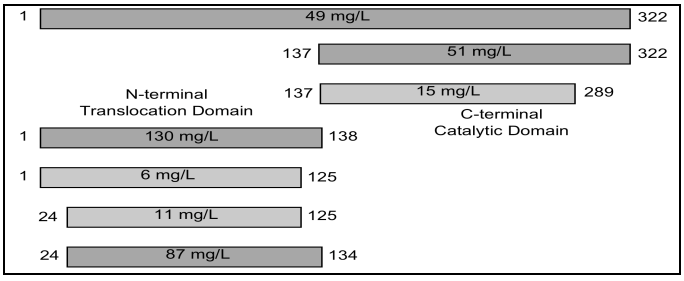


Figure 9: Domain structure. Yields are shown as mg of soluble His-tagged protein eluted off the Ni-NTA column per L of *E. coli* grown

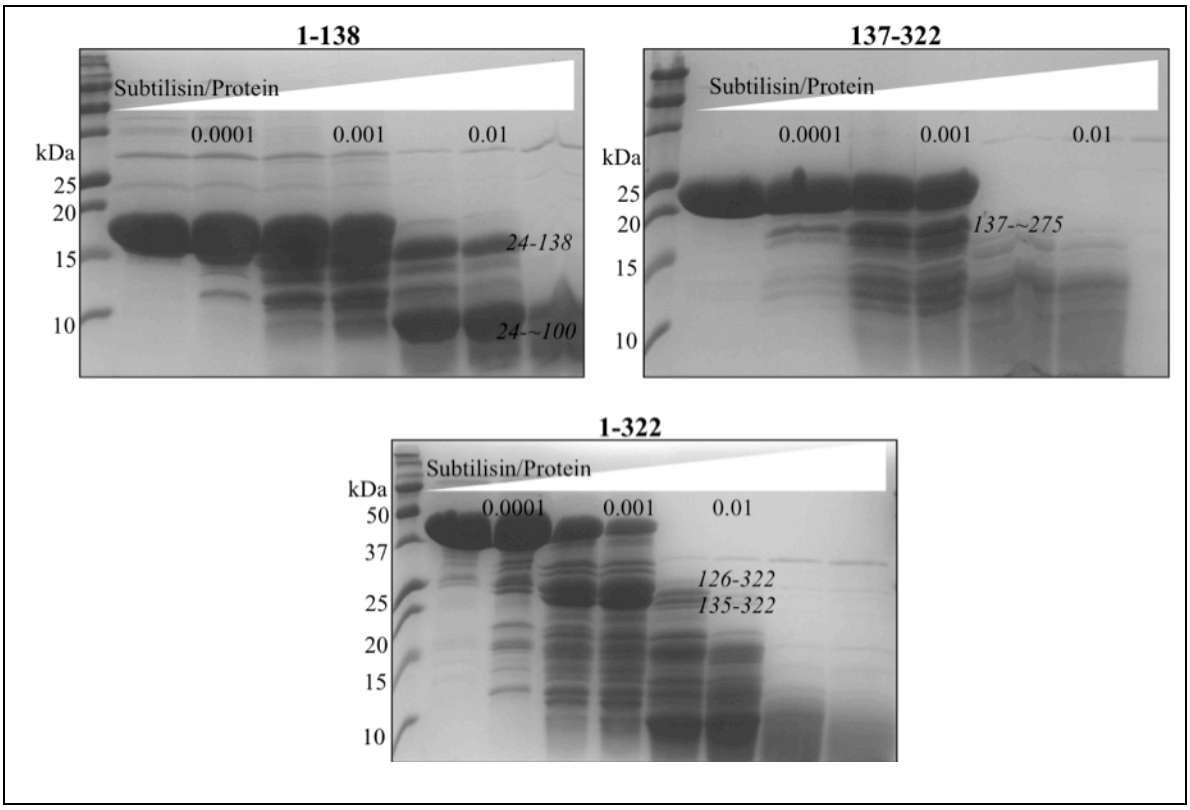


Figure 10: Limited proteolysis. 1-138, 137-322, and 1-322, were incubated with increasing amounts of subtilisin at RT and stable fragments *(italics)* analyzed by N-terminal Edman sequencing. C-terminal positions were estimated based on molecular weights of the fragment.

7.2 Crystallization

24-134 and 137-322 were screened exhaustively at RT and 4˚C. 24-134 gave fine clusters of crystals in 1.8 M Na/K phosphate pH 6.9 (RT), but these could not be further optimized to give single diffracting crystals. The complete catalytic C-terminal domain, 137-322, gave large rectangular crystals in 1.15 M $LiSO_4$, 0.1 M NaOAc, pH 4.6 (4°C). However these were very fragile and consisted of multiple lattices overlaid on each other as observed in their diffraction patterns (Figure 11). There is a string of 4 lysines within

the last 11 amino acids of the C terminus, which might be interfering with crystal packing, so the protein was reductively methylated (Rayment 1997) and rescreened. Mass spectrometry indicated that 12 out of 13 (12 Lys and N-terminal) possible sites were modified. The methylated protein gave multiple hits in a number of conditions containing MES, Tris, and HEPES buffers, pH 6.0 to 9.0, and 1.6 to 2.0 M $(NH_4)_2SO_4$. They ranged from bowties to spherical clusters of needles. To obtain single crystals we microseeded crushed crystal clusters from 0.1 M MES pH 7.1, 1.8 M $(NH_4)_2SO_4$ into 0.1 M MES pH 5.5, 1.6 M $(NH_4)_2SO_4$, to obtain single but small rectangular crystals. Macroseeding these small crystals under the same conditions gave large, thin rectangular plates that diffracted well (Figure 11). Surprisingly, SeMet-substituted crystals grew under the same final conditions as the native crystals but did not require seeding.

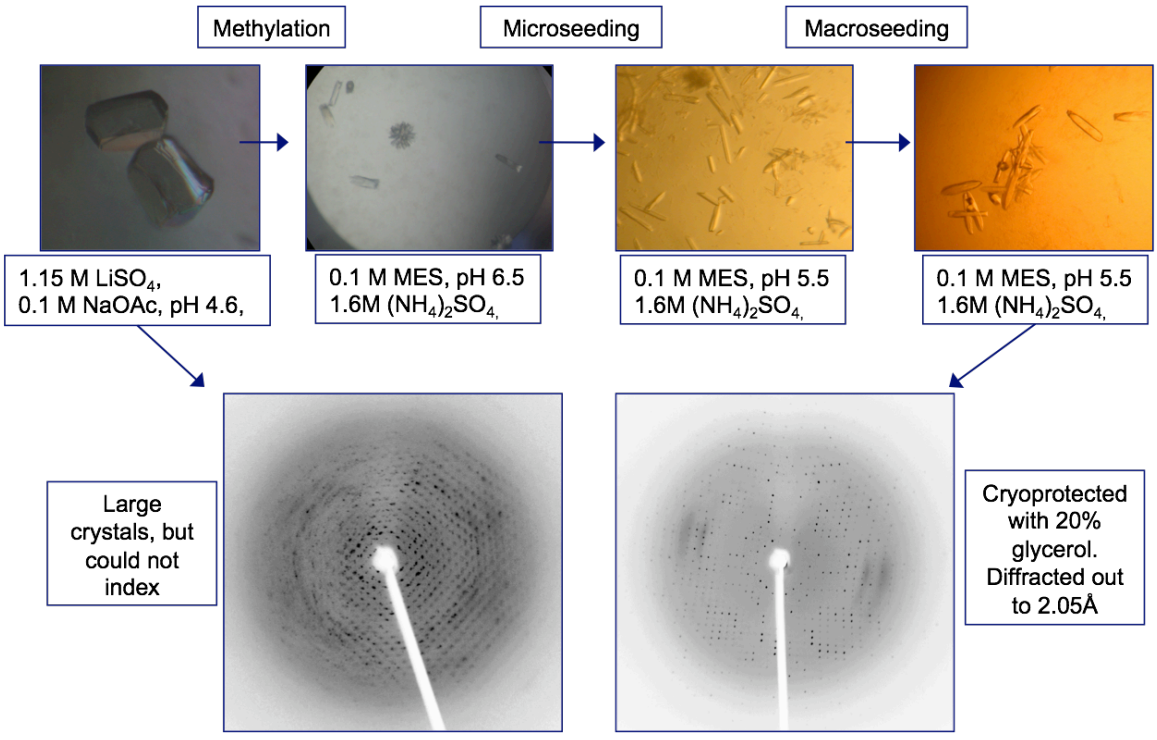


Figure 11: Steps needed to optimize native crystals for diffraction.

7.3 Overall structure

Data collected on the SeMet crystals was sufficient to solve the structure of the C-terminal catalytic domain by single-wavelength anomalous dispersion (SAD) to 1.70 Å. The final model contains 166 amino acids spanning residues 145 – 313, and 121 waters molecules, with an R/R_{free} of 20.6/24.2. There was no density observed for 12 residues at the N terminus (137-144 and 4 vector encoded residues), 9 residues at the C terminus (314-322) and 3 residues in loop 264-266. The Ramachandran plot showed that 90.4% of the residues were in the core region, 9.6% in the allowed region, and 0% in the generously allowed and disallowed regions (PDBSum, (Laskowski 2009)).

Table 10: Crystallographic statistics. Values in parentheses are for the highest resolution shell.

	SeMet SAD	**Native**
Data Collection		
Space group	$P4_22_12$	$P2_12_12_1$
Cell dimension		
a (Å)	52.256	54.048
b (Å)	52.256	63.189
c (Å)	132.146	110.618
$\alpha = \beta = \gamma$ (°)	90	90
Wavelength (Å)	0.9790	1.0809
Resolution (Å)	50.0 - 1.70	50.0 - 2.05
R_{sym}	2.7 (55.9)	2.9 (51.8)
I/σI	88.7 (2.1)	45.9 (2.1)
Completeness	99.5 (95.0)	99.5 (98.9)
Redundancy	15.2 (8.1)	4.4 (3.7)
FOM	0.841	
Refinement		
No. unique reflections	20,990	24,051
R/R_{free}	20.6/24.2	22.5/27.5
Average B-factor	26.12	38.6
R.m.s. deviations		
Bond lengths (Å)	0.013	0.021
Bond angles (°)	1.337	1.829

The overall structure and its topology are shown in Figures 12 & 13. The single catalytic domain consists of a core of 6 anti-parallel beta sheets and 8 surrounding helices. It shows a remarkable similarity to the structure of members of the Cysteine Protease Superfamily (discussed in Section 7.4), which allowed us to identify SseI's catalytic triad as being Cys178, His216, and Asp231.

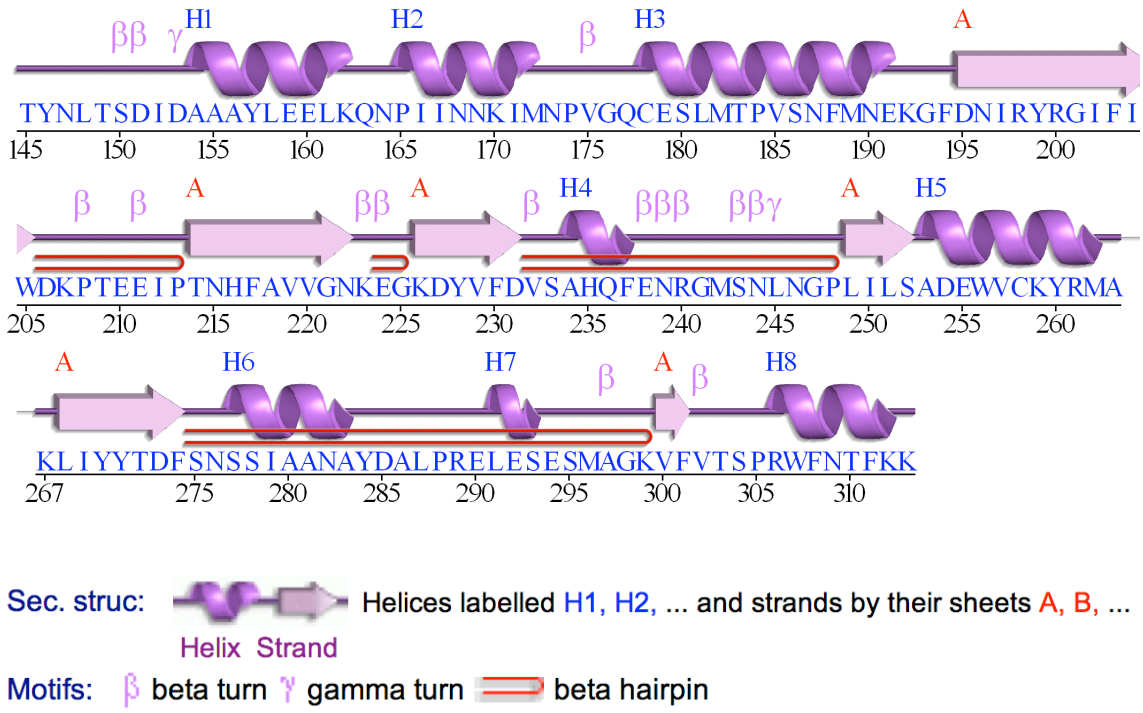


Figure 12: Secondary structure alignment with the sequence of the catalytic domain (Generated by PDBsum).

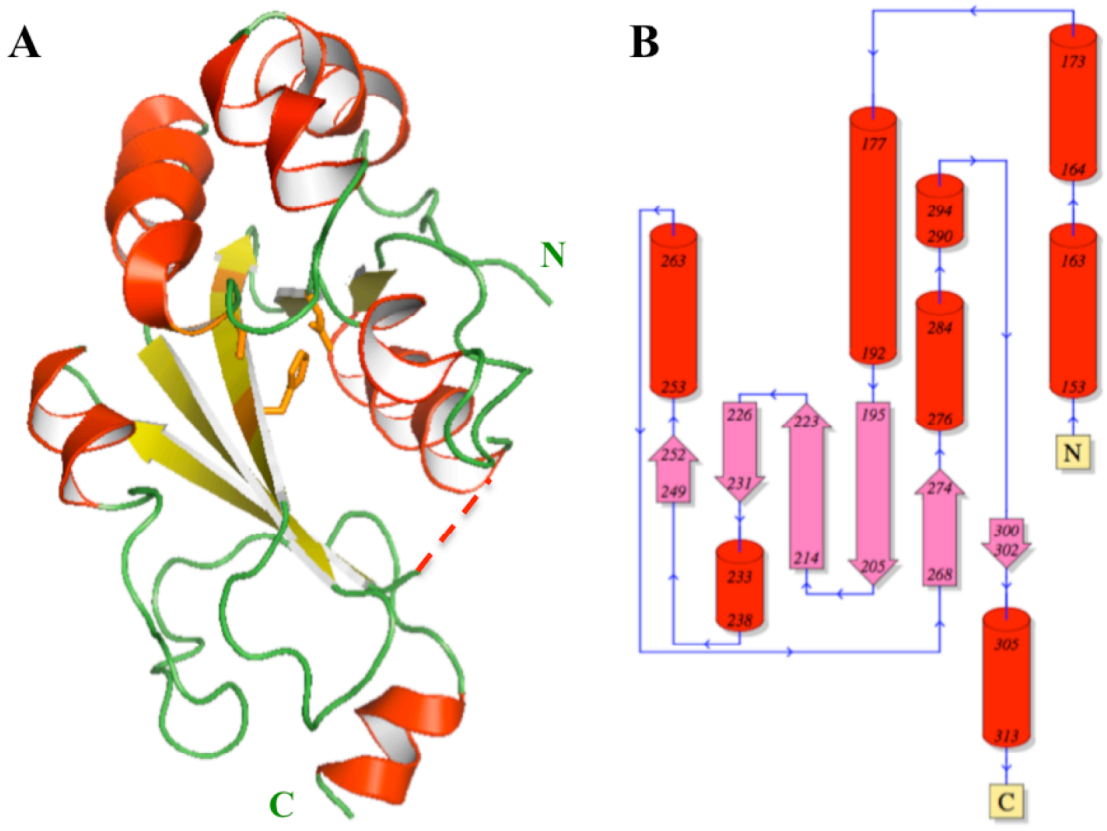


Figure 13: Structure of the catalytic domain. (A) Cys 178, His216, and Asp231 are shown in stick format (orange) and the position of the missing loop (264-266) is shown by a dashed line (red). (B) Secondary structure topology of the domain (generated by PDBsum).

The native crystals only diffracted to 2.05 Å and had a different space group from the SeMet crystals. They also had a significantly higher crystal mosaicity (0.946 vs 0.495). The native structure was solved by Molecular Replacement, using the SeMet model, and showed two molecules of the protein per asymmetric unit, R/R_{free} of 22.5/27.5. Each molecule within the dimer was linked by a disulphide bond between Cys258 on each chain, on the opposite face of the molecule from the catalytic Cys178 (Figure 14). The dimer is a crystal artifact since SseI purifies as a monomer on size exclusion

chromatography. SeMet protein was purified in 10 mM DTT instead of 2 mM DTT to prevent Se oxidation, which could explain why it didn't crystallize as a disulphide-linked dimer. It might also be the reason seeding was not required to obtain large diffracting SeMet crystals. The chain backbone in the native structure did not show any large deviations from the SeMet structure, with an r.m.s.d. of 0.282 Å between the two models. The higher resolution SeMet structure is used for all subsequent structural analyses.

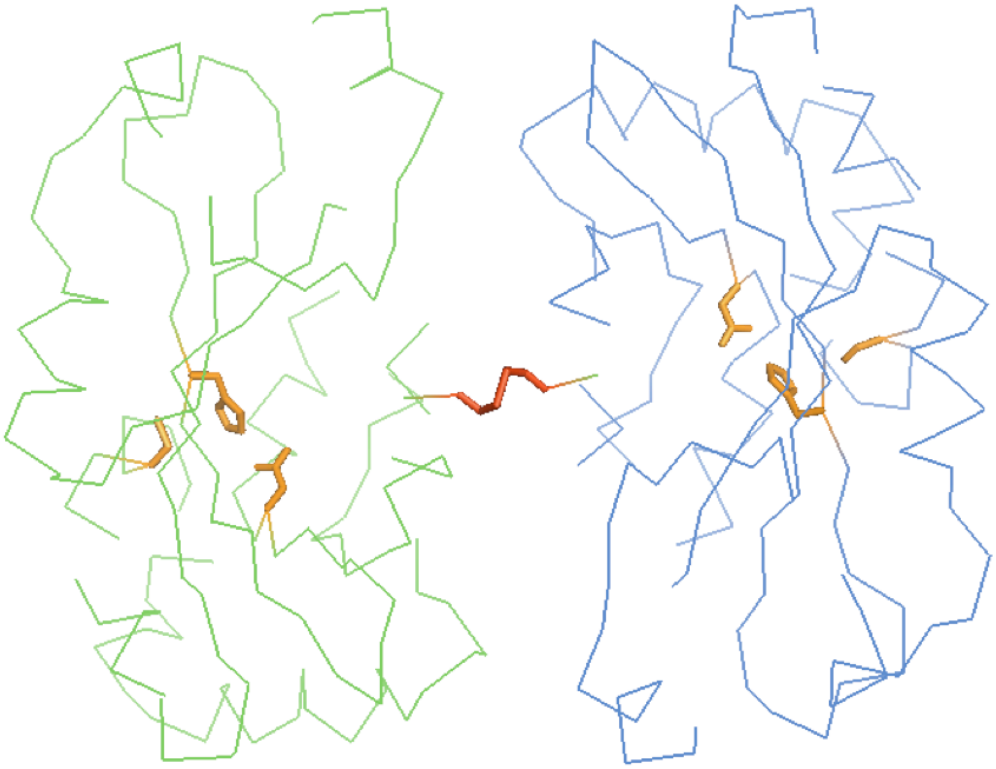


Figure 14: Structure of the native dimer form. The disulphide bond between Cys258 is shown in red and the catalytic triads in orange.

The electrostatic surface potential map of the catalytic domain is shown in Figure 15. There are multiple acidic and basic patches distributed around the molecule that could form binding surfaces for regulatory or substrate proteins. Looking at the region

surrounding the active site it is not obvious what type of substrate would be likely to bind, as there is both a positively charged and a negatively charged region on opposite sides of the solvent-accessible catalytic Cys178.

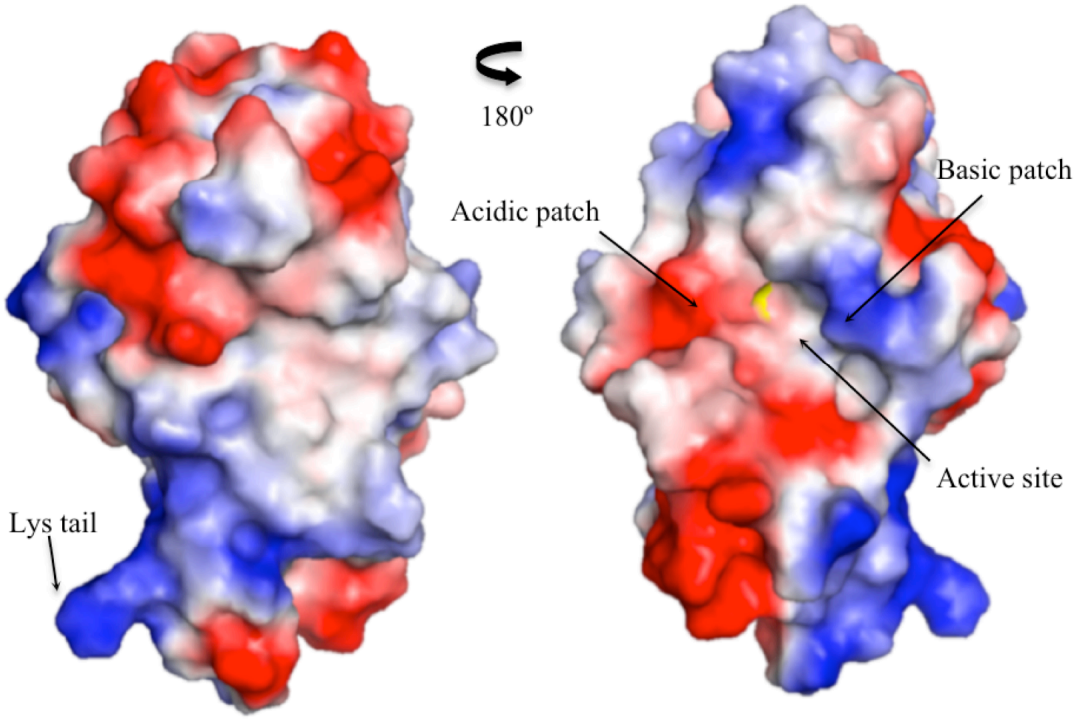


Figure 15: Electrostatic surface potentials of the catalytic domain. Two views rotated by 180º are shown. The active site Cys178 is colored yellow.

Behind Cys178, there is a small cavity with four water molecules (Figure 16). They form hydrogen bonds to the backbone of several residues surrounding them: Leu181, Cys178, Phe 217, Ser233, and Ala234. There is an additional hydrogen bond to the side chain of Asp231. There is no room in this cavity for additional molecules, and there are no other cavities from the surface that lead towards the active site residues, making it likely that a small molecule such as water is involved in SseI's catalytic mechanism. It is interesting

that the positively charged C-terminal tail containing the methylated lysines is pointing away from the core of the protein and has minimal interactions with it. This could be important for localization or binding to other regulatory proteins. It also explains why the non-methylated crystals were so fragile poorly diffracting.

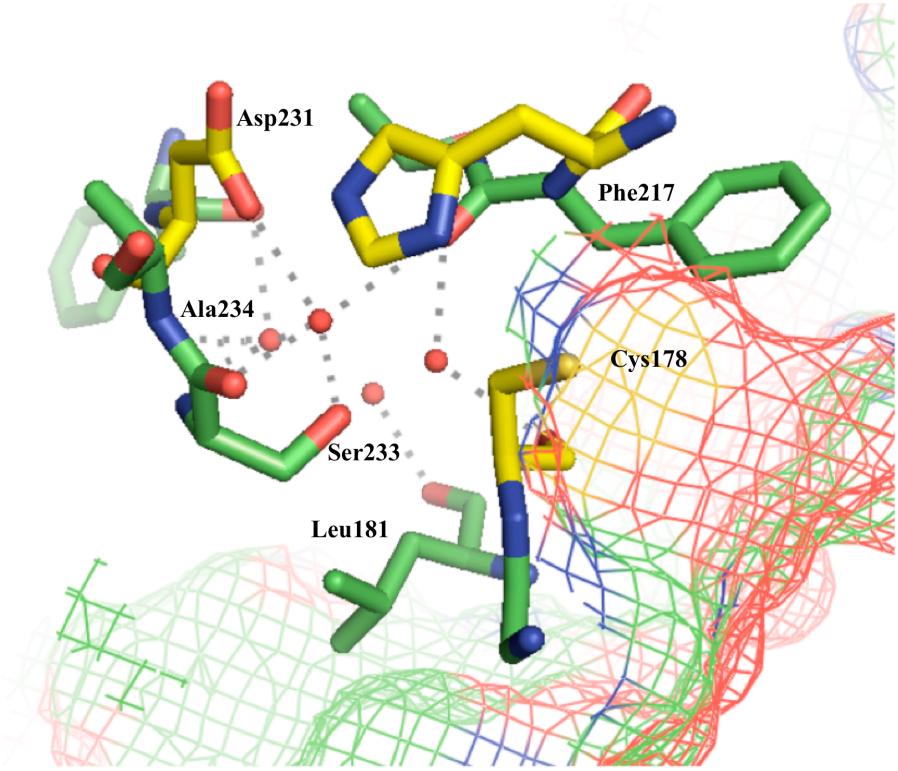


Figure 16: Residues making contact with the four water molecules in the active site. The catalytic residues are in yellow and the mesh shows the surface of the molecule.

7.4 Structural homologs

To use the catalytic domain structure as the basis of further investigation into the biochemical function of SseI, it was submitted to the DALI server (Distance alignment matrix, (Holm et al 2008)), which searches through the PDB database to find structural

homologs based on backbone Cα-atom alignments. Table 11 shows the obtained structural matches, as well as the length of the structural alignment, and sequence identity within each alignment.

Table 11: Structural matches from DALI. Z-scores above 2 are considered significant.

PDB ID	Z-Score	rmsd	No. of residues	Length of alignment	% sequence similarity	Description
2ebf	7.5	3.3	711	126	20	Toxin from *Pasteurella multicoda*
3c9q	6.6	3.6	195	111	9	Uncharacterized human C8ORF32
2z8t	6.5	3.4	185	109	15	Protein-glutaminase, *Cryseobacterium ploteolyticum*
2bsz	6.1	3.8	267	117	11	Bacterial arylamine N-acetyltransferase
1e2t	6.0	3.6	274	107	11	Bacterial arylamine N-acetyltransferase
1ukf	5.1	2.8	188	86	9	AvrPpHB, *Pseudomonas syringae* TTSS effector - protease
1gx3	5.1	3.5	276	111	11	Bacterial arylamine N-acetyltransferase
2vfb	5.0	3.7	272	112	11	Bacterial arylamine N-acetyltransferase
1x3w	4.8	3.1	320	98	19	Yeast Peptide N-Glycase
1w4t	4.8	3.6	276	108	7	Bacterial arylamine N-acetyltransferase
2pqt	4.7	3.7	295	116	13	Bacterial arylamine N-acetyltransferase
2pfr	4.6	3.6	290	108	14	Bacterial arylamine N-acetyltransferase
2hly	4.6	3.5	205	99	9	Uncharacterized *Agrobacterium tumefaciens* Atu2299
1g0d	4.6	4.1	666	108	9	Fish transglutaminase
3d9w	4.3	3.6	284	112	9	Bacterial arylamine N-acetyltransferase
2f4m	4.1	3.3	295	97	16	Mouse Peptide N-Glycase
1fl3	4.1	4.0	721	105	10	Mammalian Transglutaminase
2qsg	4.0	3.7	502	108	9	Tranglutaminase-like domain in Rad4, Yeast DNA repair protein
2iho	4.0	3.4	292	89	4	Mushroom lectin domain, *Marasmius oreades*
1l9m	3.8	4.4	681	106	13	Mammalian Transglutaminase
2q3z	3.2	3.7	655	99	9	Mammalian Transglutaminase
1kv3	3.1	4.1	651	94	11	Mammalian Transglutaminase
1cv8	3.0	2.9	173	82	7	Staphopain, protease, *Staphylococcus aureus*

All of the structural matches were from members of Clan CA of the Cysteine Protease Superfamily (EC 3.4.22), which are a large family spread across prokaryotes, eukaryotes, and viruses. They have a multitude of acyl hydrolysis or acyl transfer functions such as endo/exo-peptidases and proteases, synthetases, deamidases, deubiquitinases, transglutaminases, and acetyl transferases (Barrett & Rawlings 2001, Rawlings et al 2009). The active site His deprotonates Cys, which then acts as a nucleophile attacking acyl (C=O) bonds at the C atom. The acyl bond is reformed by displacement of the Cys by water (in the case of hydrolyis), or another nucleophile (in the case of acyl transfer). A third residue (Asn/Gln/Asp/Glu) acts to stabilize the His during Cys deprotonation. These enzymes all have a common fold, consisting of a core of antiparallel β-sheets that is surrounded by helices.

Interestingly, all the closest matches to SseI are for members of this family who have a Cys-His-Asp catalytic triad, as opposed to the classical cysteine proteases such as papain, which have Cys-His-Asn catalytic triads. The only match from this sub-family was staphopain, a protease from *Staphylococcus aureus*, having a Z-score of 3 compared to 7.5 for the top match, PMT.

The one outlier was a match to the transglutaminase-like domain of Rad4 (Z-score of 4.0), a DNA repair protein from yeast, which has a Glu (274) in place of the Asp. However, the functionality of this domain remains unclear (Min & Pavletich 2007)

The structural homologs could be categorized into five functional families: protease, Gln deamidase, transglutaminase (TG), arylamine N-acetyl transferase (NAT), and peptide N-glycanase (PNGase). It is interesting that all of these activities rely on either hydrolysis or formation of an amide moiety. An alignment of the catalytic triads from representative members of each group is shown in Figure 17, as well as from *Pasteuralla multocida* toxin (PMT), the closest structural match, whose biochemical function was unknown at the time. The measured Cys to His and His to Asp bond distances of 3.5 Å and 2.8 Å respectively fit well with that expected for this type of catalytic triad.

All three residues align closely between each structure, except for the catalytic Cys1165 of PMT, which is disulphide bonded to Cys1159 of a neighboring helix in its WT structure. The toxin is activated when this bond is broken in the reducing environment of the cytosol. If Cys1159 is mutated, the structure of the resulting mutant protein shows a significant shift in the loop containing Cys1165, which moves it into a position more closely resembling that of a classical catalytic triad. (Kitadokoro et al 2007).

PMT is also interesting in that it shows the highest sequence similarity to SseI (20% over the 126 aa structural alignment). Although not obvious in the context of a 1285 aa protein, recently another group was able to find the same region of sequence identity to PMT by an iterative PSI-BLAST search, also identifying SseI as a putative cysteine protease (McLaughlin et al 2009). This suggests an evolutionary relationship between the two proteins, not just one of structural mimicry.

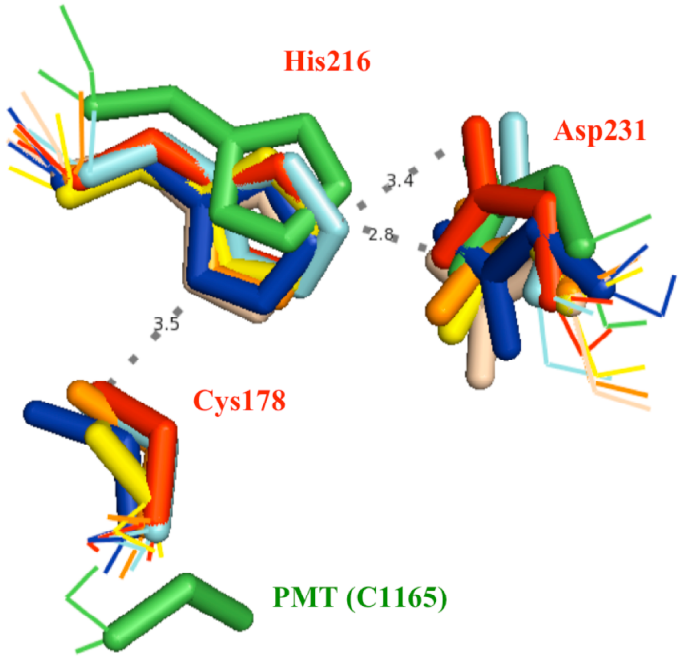


Figure 17: Alignment of the catalytic triad from representative structural matches. The distances between the SseI triad residues, C178, H216, and D231 are shown in Å. Structures shown are: SseI (red), PMT (green), Protein-glutaminase from *Cryseobacterium ploteolyticum* (blue), Arylamine N-acetyltransferase from *S. typhimurium* (yellow), Protease AvrPpH3 from *Pseudomonas syringae* (cyan), Fish transglutaminase (cream), and Yeast Peptide N-glycanase (orange).

7.5 *In vitro* testing for activity

Based on the DALI matches and functional categorizations we set out to test if SseI had any of the activities displayed by its structural homologs, namely: protease, Gln deamidase, transglutaminase, arylamine N-acetyltransferase, or Peptide N-glycanase. Small molecule and peptide substrates were used for the *in vitro* assays as SseI's substrate(s) was not known. Both full-length (FL) and C-terminal SseI were tested, since

it is possible that the N-terminal STE domain could be binding to the catalytic domain and inhibiting its activity if not in the presence of its correct binding partners, as was shown in the case of the TTSS-2 effector SspH2 (Quezada et al 2009). Furthermore, the last 180 aa of PMT, corresponding to its cysteine protease domain, has been shown to be sufficient for its activity when transfected into mammalian cells (Aminova et al 2008).

7.5.1 Protease

Since all the structural homologs were members of the Cysteine Protease Superfamily, this activity was tested first. The closest protease match is AvrPphB, a TTSS effector from the plant pathogen *Pseudomonas syringae* (Z-score 5.1) (Zhu et al 2004). AvrPphB undergoes autoprocessing (removing the first 62 aa) before becoming active, and specifically cleaves the serine/threonine kinase PBS1. It is part of the YopT family of TTSS-translocated cysteine proteases, all of which are processed in the same way (Shao et al 2002). However, there is no evidence of autoprocessing in SseI. A structural alignment of AvrPphB with SseI is shown in Figure 18. The cores fit well except for an extended helix protruding out from AvrpH3 that is not present in SseI.

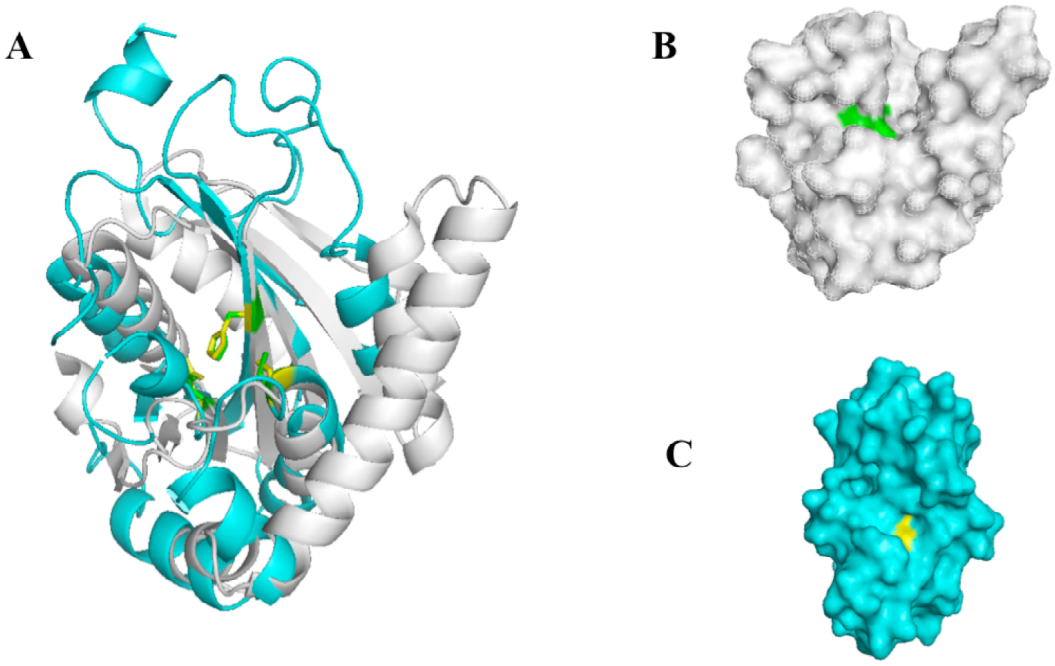


Figure 18: Structural alignment of AvrPphB and SseI. (A) AvrPphB (PDB: 1ukf) is in white with the catalytic triad C98/H212/D227 (green) and SseI is in blue with its catalytic triad in yellow. Surface representations are also shown of: (B) AvrPphB and (C) SseI.

Two separate assays from Invitrogen were used to check for protease activity. The EnzChek Peptidase/Protease Kit (E33758) uses a mixture of small peptides that are labeled at one terminus with a fluorophore and a quencher at the other. Cleavage results in separation of the fluorophore from the quencher and an increase in fluorescence at 528 nm. The EnzChek Protease Kit (E6638) uses casein modified with the green dye BODIPY FL. Cleavage results in the release of highly fluorescent peptides emitting at 598 nm. The effect of added 10 mM DTT (in case the cysteine needed to be reduced before becoming active) was also tested. Papain, a general cysteine protease found in papaya, was used as positive control. Neither FL nor C-terminal SseI showed any activity in these assays (Figure 19 & 20), nor did added 10 mM DTT have an activating

effect. However, it is still possible that SseI could be a protease with high substrate selectivity.

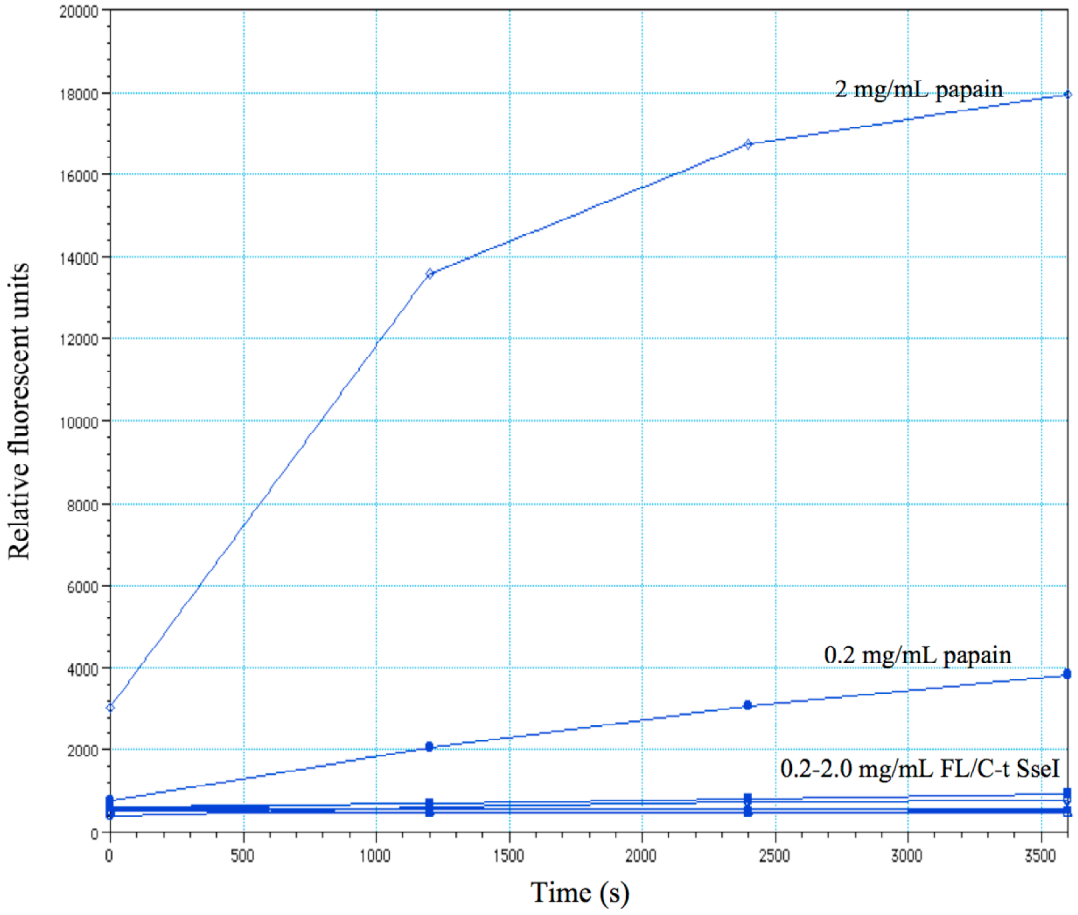


Figure 19: Protease assay using fluorescent peptide. 0.2-2.0 mg/mL of FL and C-terminal SseI were used, with papain as a control. Addition of 10 mM DTT did not alter the activity of SseI. Additon of 0.2 mM E-64 completely ablated papain activity.

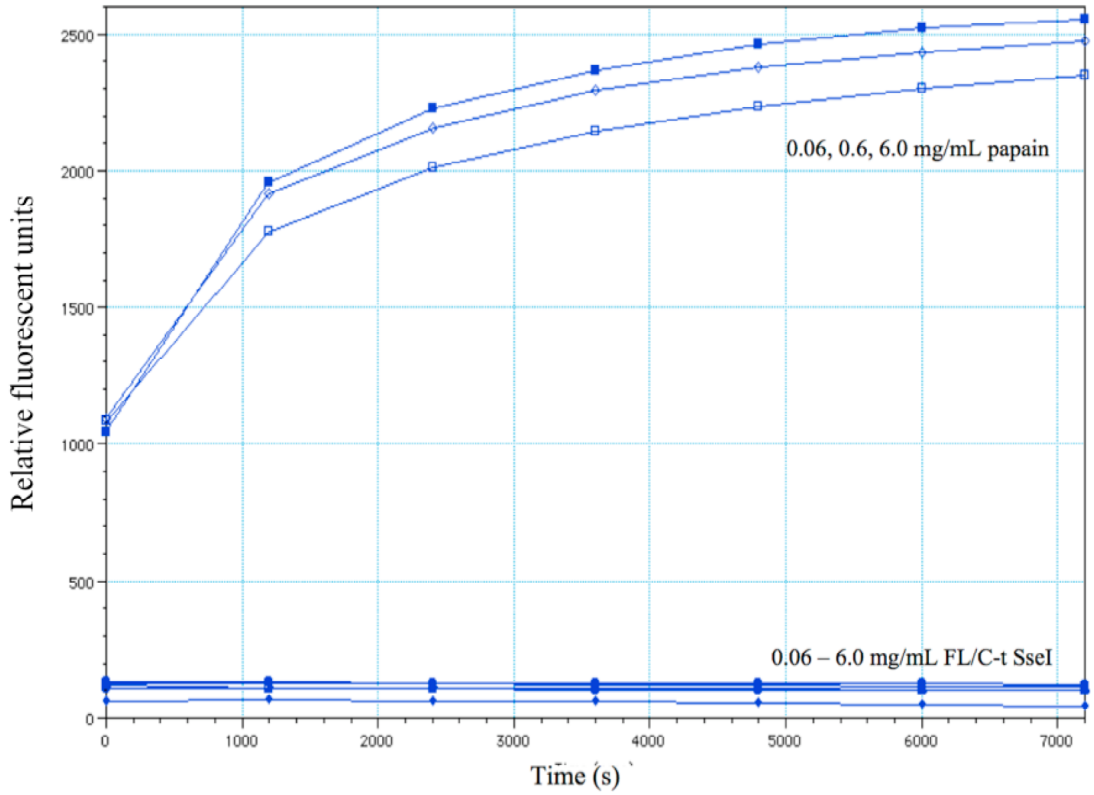


Figure 20: Protease assay using fluorescent casein. 0.06-6.0 mg/mL of FL and C-terminal SseI were used, with papain as a control. Addition of 10 mM DTT did not alter the activity of SseI. Additon of 0.2 mM E-64 completely ablated papain activity.

7.5.1. Gln Deamidase

Protein-glutaminase from *Cryseobacterium proteolyticum* is a deamidase that converts glutamine residues to glutamate in both peptides and proteins (Yamaguchi et al 2001). A structural alignment of the two proteins (Z-score of 6.5) is shown in Figure 21. The cores fit well but protein-glutaminase contains two additional beta strands on the outside of the core that are not present in SseI.

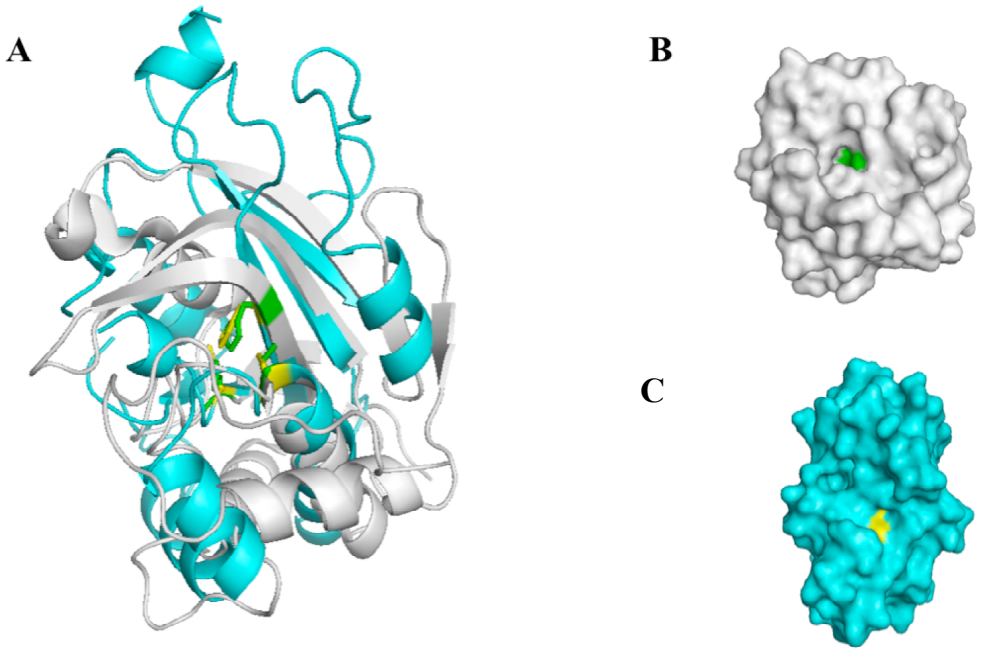


Figure 21: Structural alignment of Protein-glutaminase and SseI. (A) Protein-glutaminase (PDB: 12zk9) is in white with the catalytic triad C42/H83/D103 (green) and SseI is in blue with its catalytic triad in yellow. Surface representations are also shown of: (B) Protein-glutaminase and (C) SseI.

To test for deamidase activity, SseI was incubated at 37°C with Z-Gln-Gly, a small peptide substrate, and aliquots of the reaction tested for ammonia release using another enzymatic assay. The second assay uses the transfer of ammonia to α-ketoglutaric acid in the presence of NADPH and L-glutamate dehydrogenase. The decrease in NADPH concentration can be followed at 340 nm (Figure 22). Casein was also tested as a substrate (Figure 23). Using casein additionally checks for asparagine deamidation, an activity which is also commonly found as a protein modification and follows a similar chemistry to gln deamidation. Since no deamidases were commercially available, transglutaminase from pig liver (+ hydroxylamine), which also releases ammonia from

Gln during its reaction, was used a positive control. Neither FL nor C-terminal SseI showed any activity with Z-Gln-Gly or casein.

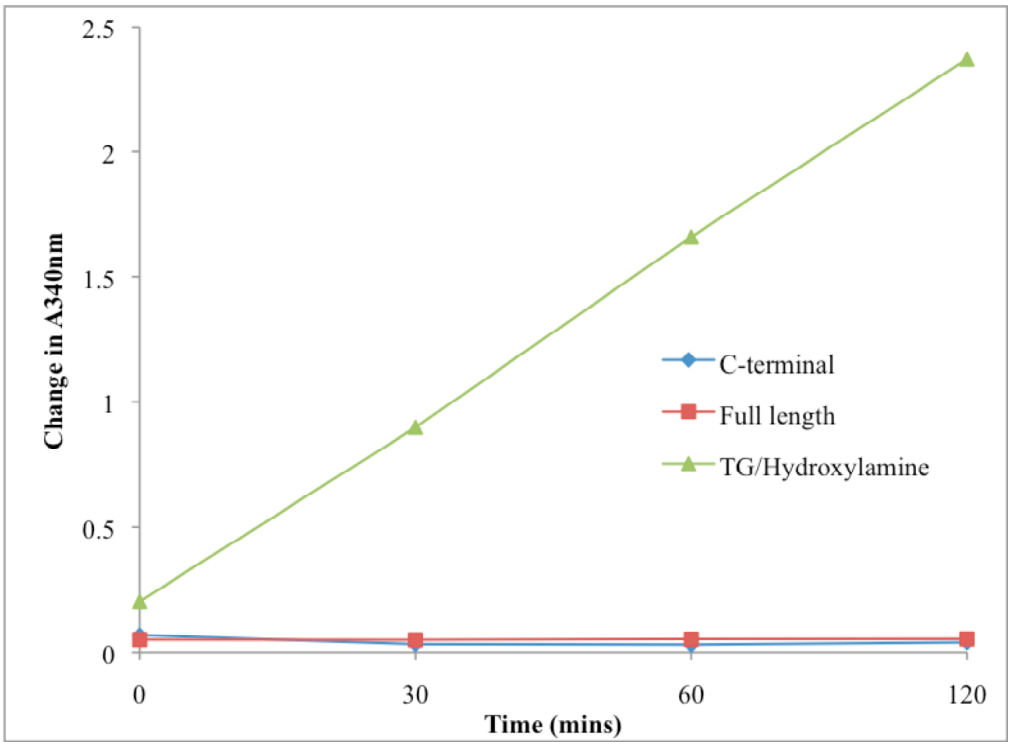


Figure 22: Ammonia release from 30 mM Z-Gln-Gly. 9 μM of FL or C-terminal SseI was used, with 3 μM pig liver transglutaminase (+100 mM NH_2OH) as a control.

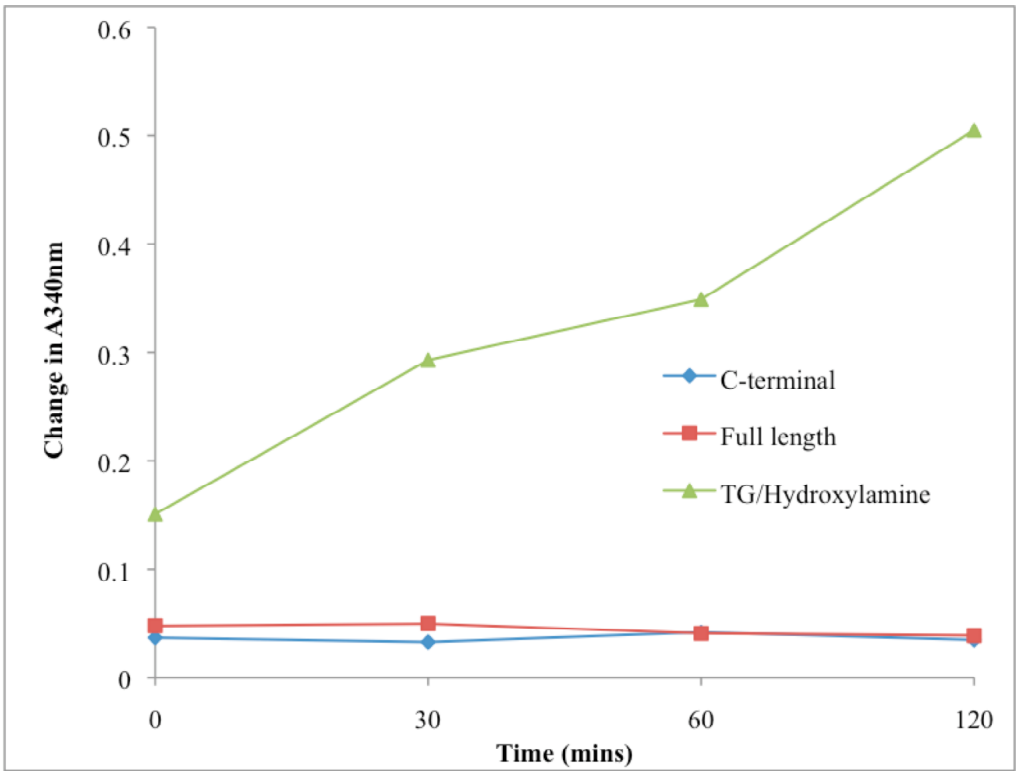


Figure 23: Ammonia release from 1 mM casein. 9 μM of FL or C-terminal SseI was used, with 3 μM pig liver transglutaminase (+100 mM NH_2OH) as a control.

7.5.3 Transglutaminase

The transglutaminases (TGs) are a widespread group of Ca^{2+}-dependent enzymes that catalyze the cross-linking of Gln to Lys residues between proteins, or of Gln to free amines in the cell cytosol and extracellular matrix. They function as molecular glue, polymerizing proteins and stabilizing them in tissues, and have been implicated in wound healing, apoptosis, angiogenesis, and cellular regeneration (Griffin et al 2002). A structural alignment of SseI and the cysteine protease domain from Red Sea Bream (a fish) TG is shown in Figure 24 (Z-score of 4.6). TG has three domains in addition to the

catalytic domain, which bind Ca^{2+} and have regulatory functions on activity, as well as separate sites binding sites for the acyl donor and the Gln acceptor (Noguchi et al 2001).

To test for TG activity, SseI was incubated with Z-Gln-Gly or casein and hydroxylamine (NH_2OH) at 37°C, and ammonia released during the reaction monitored. As a further test in the case of Z-Gln-Gly, additional aliquots were removed to check specifically for hydroxamate formation between Gln and hydroxylamine. The hydroxamate test was not possible when using casein as a substrate since the assay stop solution also precipitated out casein (unlike the smaller peptide Z-Gln-Gly). Pig liver TG was used a positive control in all the assays. Neither FL nor C-terminal SseI displayed TG activity leading to ammonia release or hydroxamate formation (Figure 25, 26, & 27).

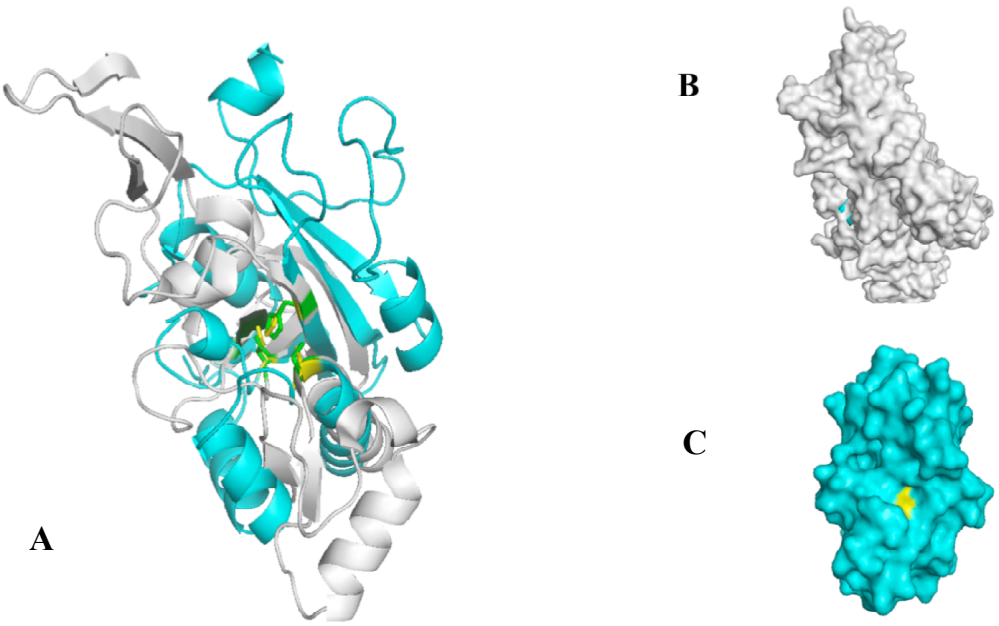


Figure 24: Structural alignment of TG catalytic domain and SseI. (A) TG (PDB: 1g0d) is in white with the catalytic triad C272/H332/D355 (green) and SseI is in blue with its catalytic triad in yellow. Surface representations are also shown of: (B) TG and (C) SseI.

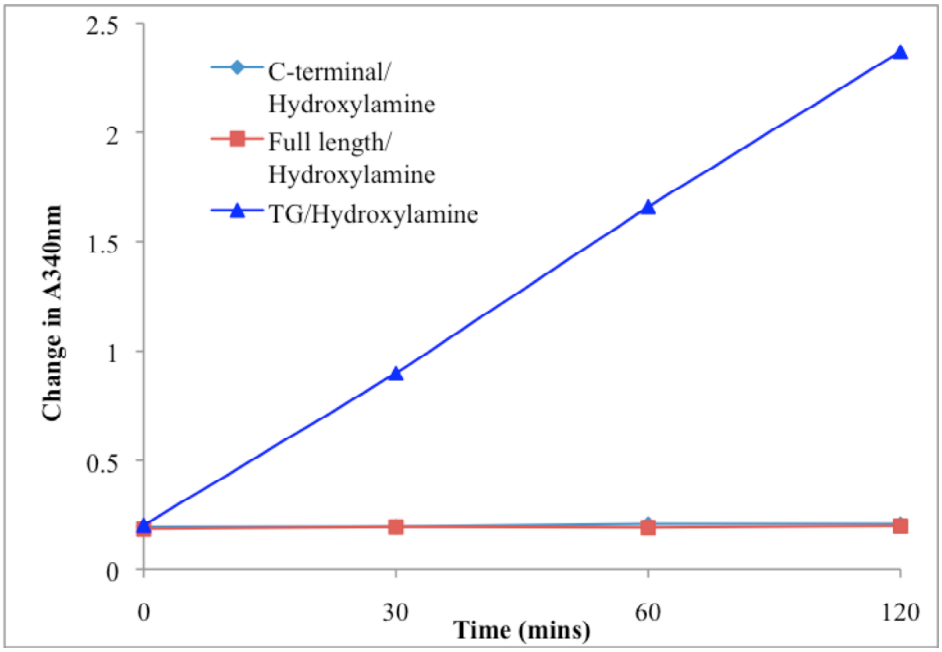


Figure 25: Ammonia release from 30 mM Z-Gln-Gly + 100 mM hydroxylamine. 9 μM of FL or C-terminal SseI was used, with 3 μM pig liver TG as a control.

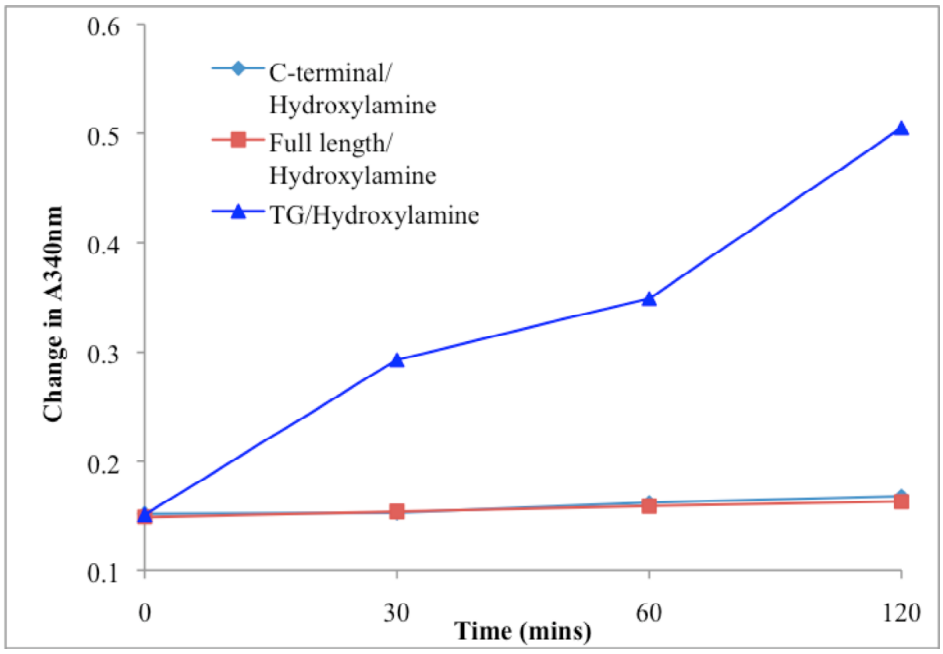


Figure 26: Ammonia release from 1 mM casein + 100 mM hydroxylamine. 9 μM of FL or C-terminal SseI was used, with 3 μM pig liver TG as a control.

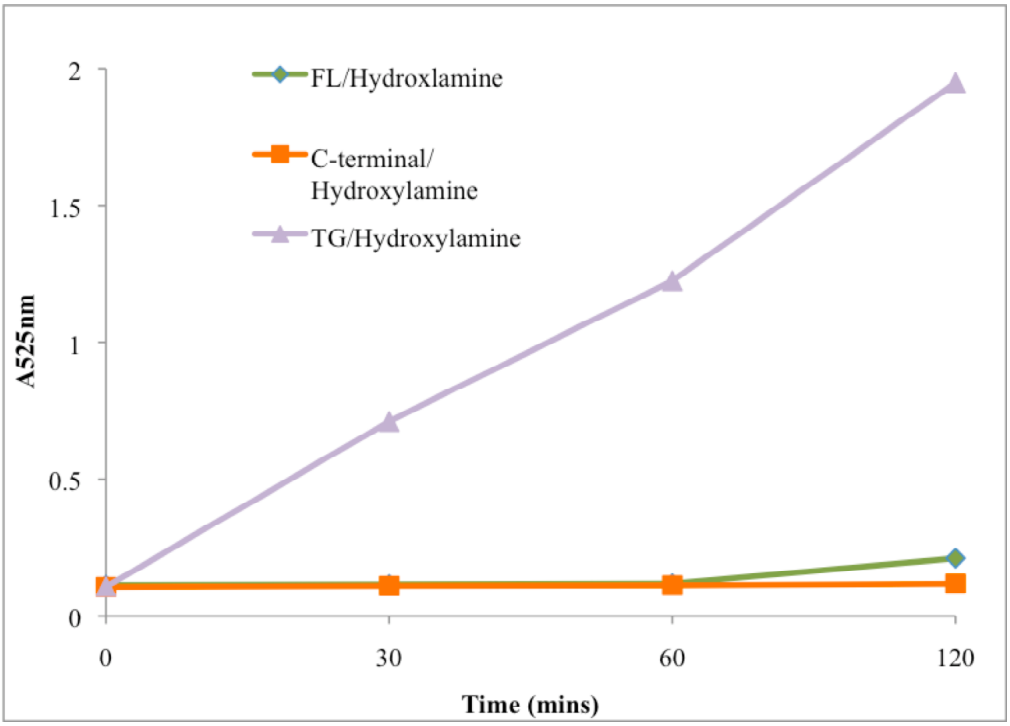


Figure 27: Hydroxamate formation from 30 mM Z-Gln-Gly + 100 mM hydroxylamine. 9 μM of FL or C-terminal SseI was used, with 3 μM pig liver TG as a control.

7.5.4. Arylamine N-acetyl transferase

The NATs catalyze the transfer of an Acetyl group from AcetylCoA to the terminal nitrogen of a wide range of hydrazine and arylamine compounds (Upton et al 2001). In prokaryotes they are usually involved in inactivating antibiotics (first identified in isoniazid-resistant *Mycobacterium tuberculosis*), while in mammals they play a role in drug metabolism and foreign substance carcinogenesis by the liver. A structural alignment of SseI with the cysteine protease domain of NAT from *S. typhimurium* is shown in Figure 28 (Z-score of 6.0) (Sinclair et al 2000). The cores of the two cysteine

protease domains match very well but NAT contains multiple helical insertions, creating an extra lobe in the protein that is not present in SseI.

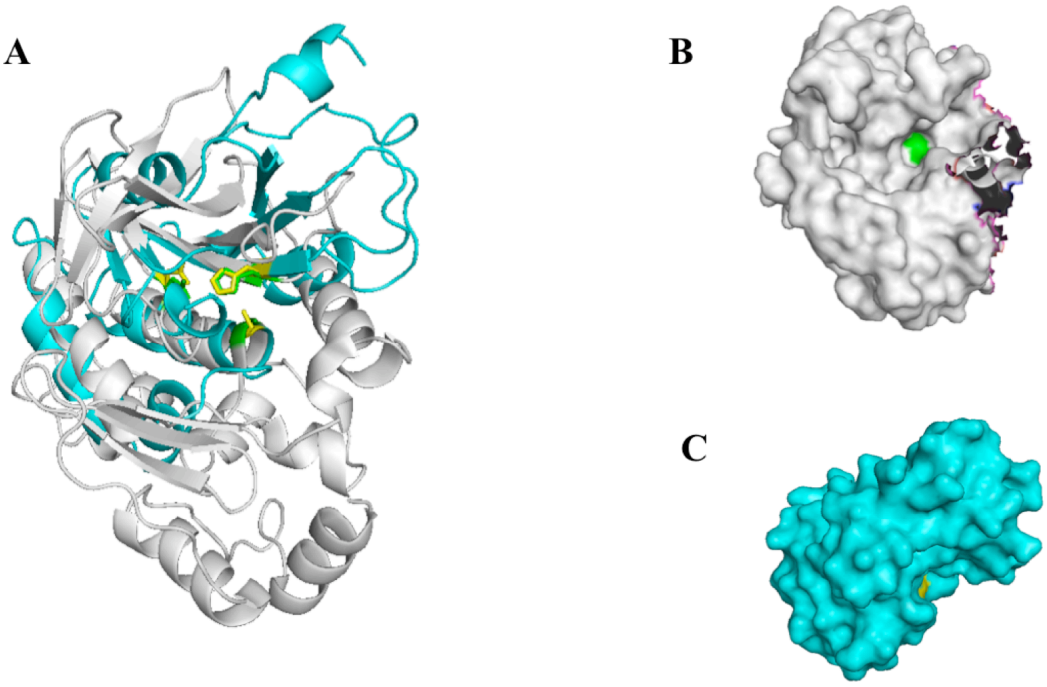


Figure 28: Structural alignment of NAT catalytic domain and SseI. (A) TG (PDB: 1e2t) is in white with the catalytic triad C69/H107/D122 (green) and SseI is in blue with its catalytic triad in yellow. Surface representations are also shown of: (B) NAT and (C) SseI.

To look for NAT activity, we tested four representative arylamine substrates: Sulfamethazine, 4-aminobenzoic acid, 4-anisidine, and Isoniazid (Brooke et al 2003). Two different assays were done. The first looked for appearance of free sulfhydryl groups as CoA is released from AcetylCoA (detected by 5,5”-dithio-bis(2-nitrobenzoic) acid (DTNB)) and the other detected the disappearance of free amine groups as the substrate was used up (detected by dimethylaminobenzaldehyde (DMAB)). The second assay also contained LiK Acetyl phosphate and phosphotransacetylase in the reaction mix

to recycle AcetylCoA and drive the reaction forward. Human cytosolic NAT2 was used as a positive control. However, neither FL nor C-terminal SseI showed activity in either assay (Figure 29 & 30).

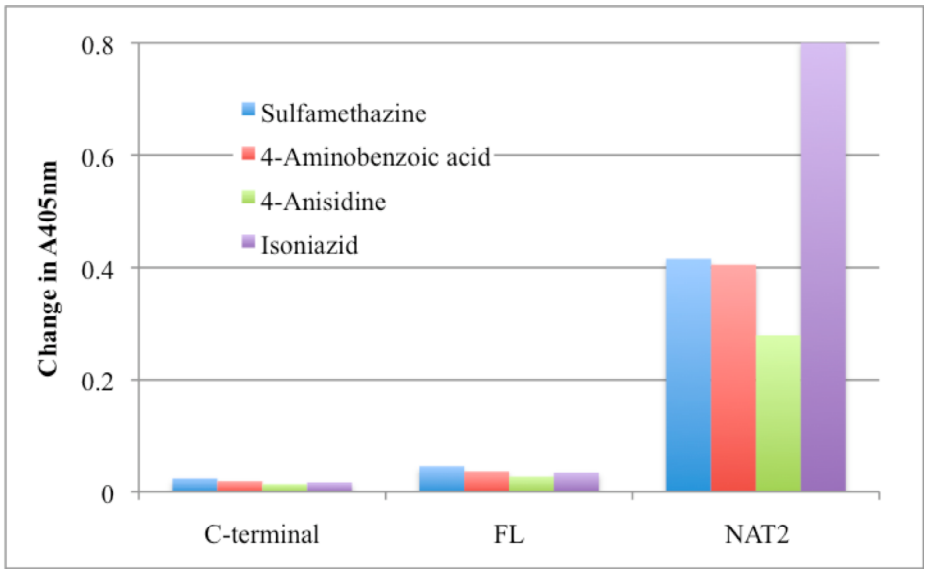


Figure 29: Sulfhydryl release from 400 µM AcetylCoA and 500 µM substrate. Concentrations used were: SseI FL (40 µg/mL), C-terminal (20 µg/mL), and NAT2 (20 µg/mL)

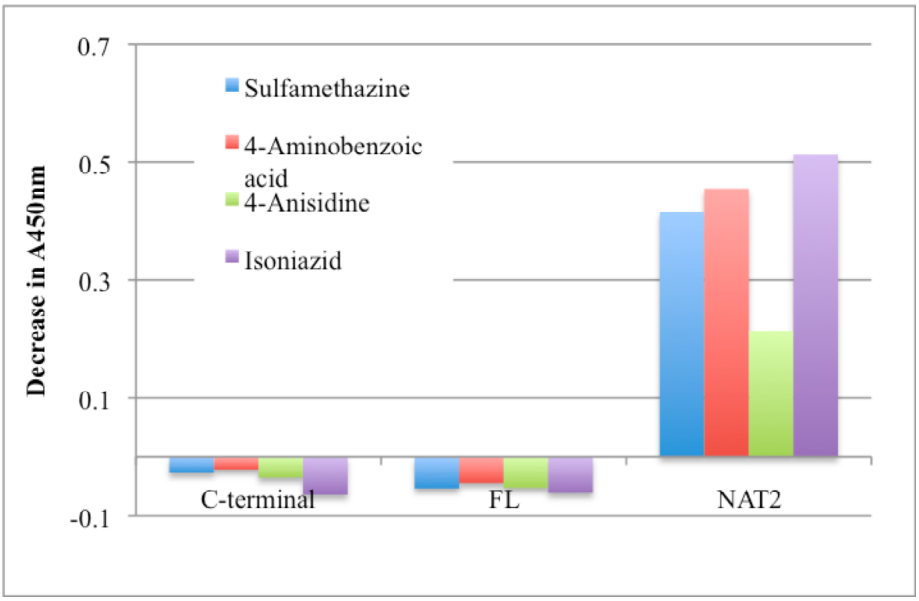


Figure 30: Amine disappearance measured in 400 µM AcetylCoA and 500 µM substrate. Concentrations used were: SseI FL (40 µg/mL), C-terminal (20 µg/mL), and NAT2 (20µg/mL)

7.5.5 Peptide N-glycanase

Peptide N-glycanase is involved in the deglycosylation of misfolded glycoproteins before they can be degraded by the proteosome, and is widely conserved in eukaryotes. It specifically cleaves the bond between the Asn residue and first GlcNAc residue of the glycan chain in N-linked glycoproteins (Suzuki 2005). A structural alignment of SseI and the cysteine protease domain of yeast PNGase is shown in Figure 31 (Z-score of 4.8) (Lee et al 2005). The cores of the two proteins match but a distinguishing feature of PNGase is a deep binding cleft formed between the catalytic domain and an adjacent domain, along which the unfolded glycoprotein can bind. Such a binding cleft is missing from SseI, making it unlikely that it is acting as a PNGase. Also, by this time a biochemical function for PMT had been determined (Orth et al 2009), which modified our search for SseI's function. As a result, an assay for PNGase activity was not done.

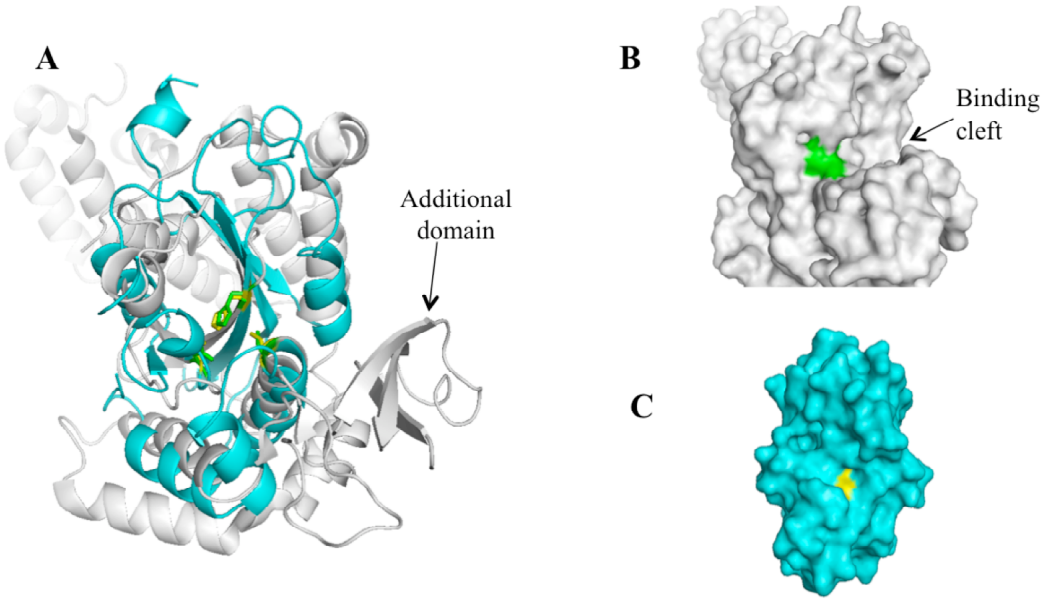


Figure 29: Structural alignment of the catalytic domain of PNGase and SseI. (A) PNGase (PDB: 3esw) is in white with the catalytic triad C191/H218/D235 (green) and SseI is in blue with its catalytic triad in yellow. Surface representations are also shown of: (B) PNGase and (C) SseI.

7.5 *Pastereulla multocida* toxin is a Gln deamidase

When SseI's structure was solved, its closest structural homolog PMT was known to bind the Gα subunits of heterotrimeric GTPases and affect their downstream signaling (Orth et al 2008, Orth et al 2005), but its exact biochemical mechanism was not known. A recent report was able to determine that its catalytic domain (1105-1285) was a Gln deamidase, specifically converting Gln205 in $G\alpha_i$ and $G\alpha_q$ to glutamate (Orth et al 2009). They were able to show this enzymatic activity by co-expressing PMT with $G\alpha_{i,}$ in *E. coli*, and analyzing purified $G\alpha_i$ by trypsin digestion and mass spectrometry. $G\alpha_q$ modification was shown by treating mouse embryonic fibroblasts with PMT overnight and then isolating out $G\alpha_q$ in detergent extracts of the membrane. Purified $G\alpha_q$ showed a shift in 2-D gel electrophoresis consistent with deamidation and a decrease in pI. Modification of Gln205 stops the intrinsic GTPase activity of the Gα subunit, keeping it in the GTP-bound form once it has separated from the $G_{\beta\gamma}$ complex, and leading to constitutive signaling through its downstream partners.

Based on the high structural similarity with the catalytic domain of PMT (Figure 32, Z-score of 7.5), their common evolutionary lineage (shown by sequence similarity to regions within this domain) (McLaughlin et al 2009), and the small active site cavity (consistent with water being the attacking nucleophile during its catalysis), we hypothesized that SseI was also a Gln deamidase. Even though it had not shown activity in the *in vitro* Z-Gln-Gly and casein ammonia release assays, this could be explained by a high substrate specificity rather than lack of deamidating functionality.

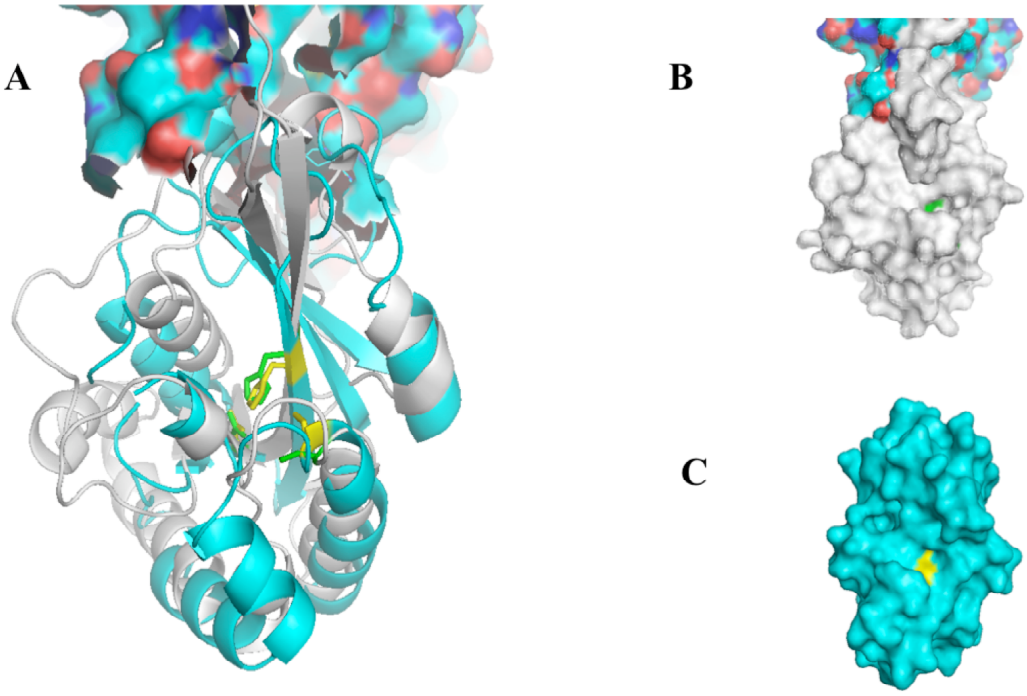


Figure 30: Structural alignment of the catalytic domain of PMT and SseI. (A) PMT (PDB: 2ebf) is in white with the catalytic triad C1165/H1205/D1220 (green) and SseI is in blue with its catalytic triad in yellow. Surface representations are also shown of: (B) PMT and (C) SseI.

7.6 Rho GTPases as possible substrates

Furthermore, based on the localization of SseI to regions of active actin polymerization, and its involvement in cell motility and adhesion, we hypothesized that its substrate was a member of the Rho GTPase family rather than a Gα protein. The Rho GTPases (~20) behave as molecular switches during signal transduction, switching between an active GTP-bound form and an inactive GDP bound form (Bustelo et al 2007, Heasman & Ridley 2008). The rate of intrinsic GTPase activity determines the length of the 'on' signal, but signaling can also be modified by three classes of regulatory proteins.

Guanine nucleotide exchanges factors (GEFs) promote release of GDP and binding of GTP, turning the signal on. GTPase activating proteins (GAPs) increase the Rho's intrinsic GTPase activity, turning the signal off. Finally, the guanine nucleotide dissociation inhibitors (GDIs) bind to the C-terminal prenyl tails of some Rhos and sequester them in the cytoplasm away from their targets and regulators at the cell membrane, effectively shutting down the signaling pathway.

In particular, Cdc42, Rac1, and RhoA, all have direct effects on cellular movement, cytoskeletal rearrangement, cellular adhesion, and cell polarity. Moreover, these three Rhos are also targeted by other bacterial toxins (Lemonnier et al 2007). For instance, all three are deamidated by CNF1 (cytotoxic necrotizing factor) from *E. coli* and transglutaminated with small cytosolic amines by DNT (dermonecrotic toxin) from *Bordetella*, specifically at Gln61 (Cdc42/Rac1) or Gln63 (RhoA). The Rhos contain two 'switch' regions (28-44, 59-97, in RhoA numbering), which change their conformation based on whether GTP or GDP is bound to the protein. When a GTP molecule replaces GDP, the instrinsic GTPase activity of the protein is turned on. However, if Gln61/63 is modified, the enzyme is no longer able to correctly position a water molecule for hydrolytic attack on the γ-phosphate group of the nucleotide. Hence there is no conversion back to GDP, and the Rho is kept in the 'active' GTP form (Hakoshima et al 2003).

We tested Cdc42, Rac1, and RhoA in a modified version of the glutaminase assay. The ammonia detection reagent was combined with each Rho and SseI in one reaction mix

and the decrease in fluorescence (at 460nm) due to a decrease in NADPH concentration from the coupled reaction as ammonia is released was measured in a microplate. Both the GDP and GTP-bound (pre-loaded with the non-hydrolyzable analog GTPγS) forms of each Rho were tested, and SseI was added as an equimolar mix of FL and C-terminal protein. Hydroxylamine was also added to the mix to test for possible transglutamination activity since the reaction mechanisms are similar, with water being the attacking group in one and a small amine in the other. The deamidation of RhoA by CNF1 in a similar assay proceeds to completion within 5 mins (Lerm et al 1999), so the reaction was monitored for 30 mins, after which added $(NH_4)_2SO_4$ was used as a standard to calculate the expected signal from complete deamidation of each Rho (a decrease of ~500 RFUs).

SseI did not show deamidation or transglutamination activity in the assays. There was some background decrease in signal over time, but this was due to degradation/ bleaching of the NADH and not SseI's enzymatic activity, since the Rhos on their own also showed the same trend (Figure 32, 32, & 34). Interestingly, *in vitro* assays with PMT and purified Gαs also did not show deamidase activity (Orth et al 2009), raising the possibility that an activating component from the bacterium or the host cell is needed for full activity of both SseI and PMT.

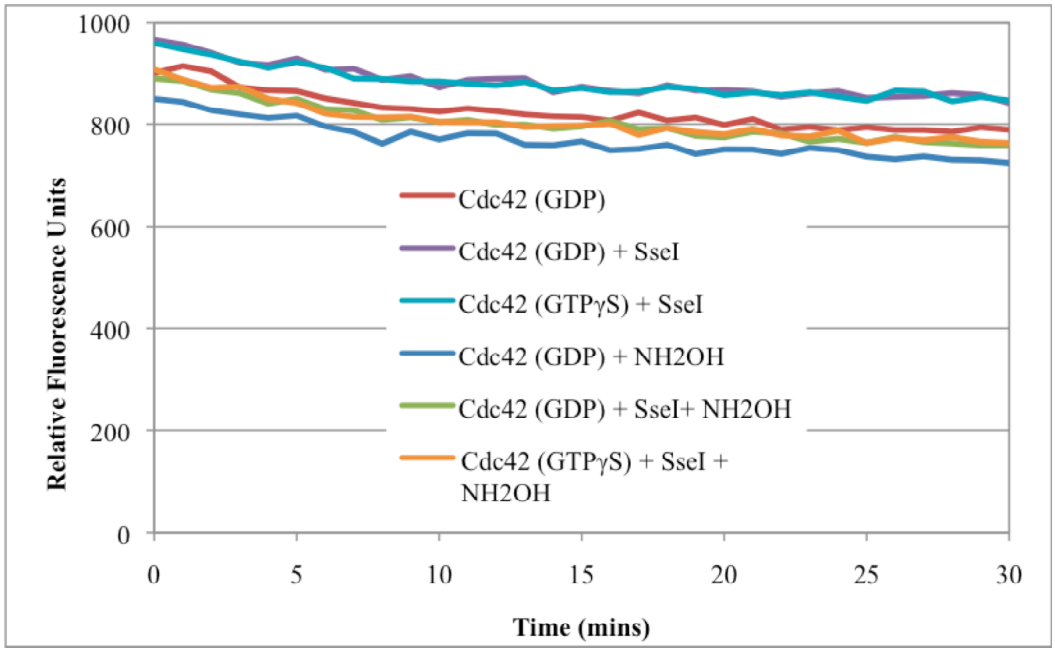


Figure 31: Ammonia release from deamidation and transglutamination of Cdc42.
Concentrations used were: SseI (200 μM mix of FL & C-terminal), Cdc42 (194 μM), Rac1 (300 μM), and RhoA (120 μM).

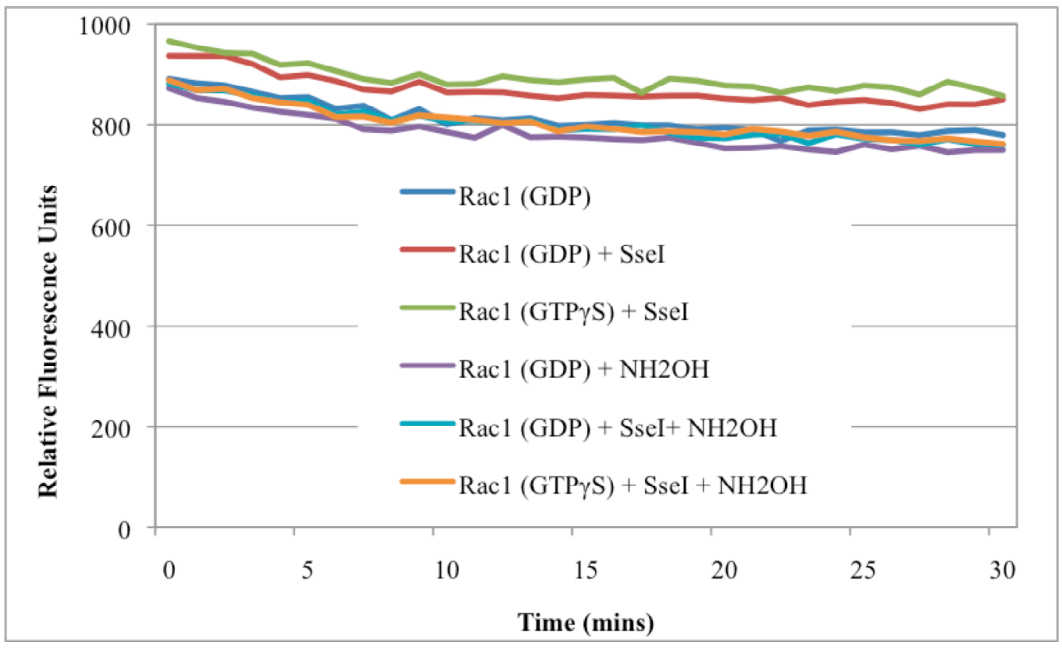


Figure 32: Ammonia release from deamidation and transglutamination of Rac1.
Concentrations used were: SseI (200 μM mix of FL & C-terminal), Cdc42 (194 μM), Rac1 (300 μM), and RhoA (120 μM).

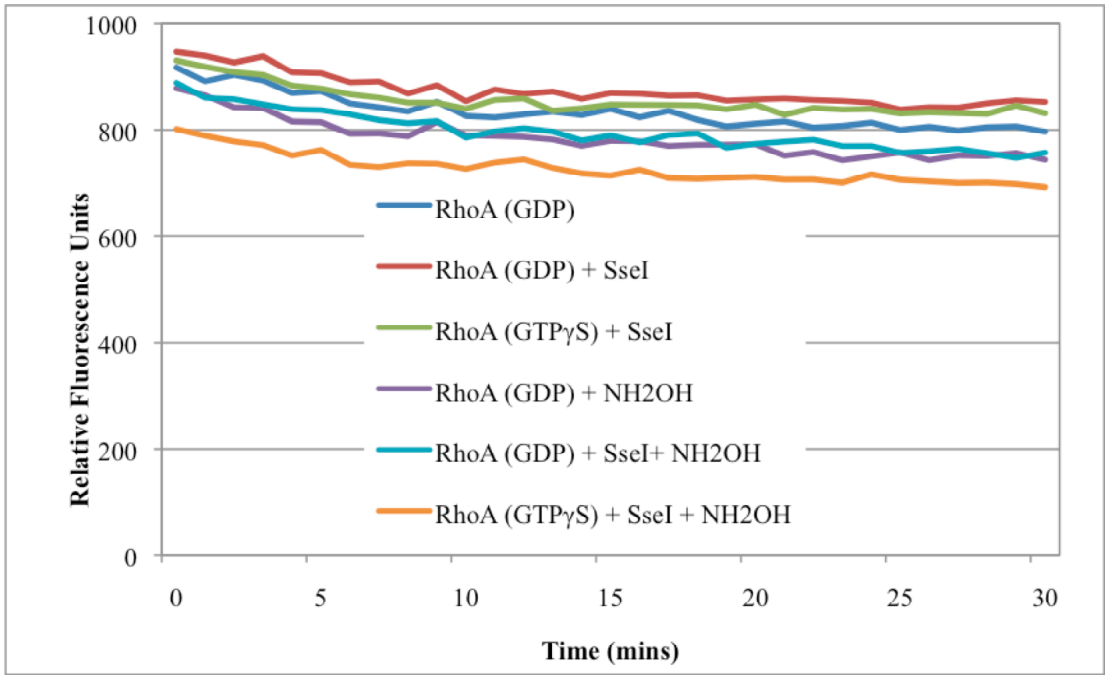


Figure 33: Ammonia release from deamidation and transglutamination of RhoA. Concentrations used were: SseI (200 μM mix of FL & C-terminal), Cdc42 (194 μM), Rac1 (300 μM), and RhoA (120 μM).

7.7 Co-expression of SseI and Rhos

The deamidase activity of PMT had originally been shown through co-expression with $G\alpha_i$ in *E. coli*, and since the *in vitro* assays did not show deamidation or transglutamination of the Rhos under the conditions used, we decided to follow a similar approach. Co-expressions of Cdc42, Rac1, and RhoA with FL SseI and FL SseI C178A mutant were done. The purified Rhos (as SDS-PAGE gel bands after Glutathione sepharose purification) were sent for trypsin digestion followed by ESI mass spectrometry of the eluted peptides. Deamidation of the Rhos was checked for by shifts in the mass spectra of the Gln61/63-containing peptides of each Rho. In each spectra shown there are multiple peaks corresponding to the peptide being analyzed. The first

peak is the monoisotopic (theoretical) peak, with C13, O18, and higher isotope peaks following. Since ESI works using the native charge of the molecule being analyzed, the observed m/z of each peptide is actually the total theoretical peptide mass divided by the number of charged residues present in the peptide.

Rac1 and RhoA did not show any modification of the Gln61/63 containing peptides on co-expression with FL SseI, as compared to co-expression with the FL C178A mutant (Figures 37 & 38). However, Cdc42 showed a small enrichment where the deamidated form of the peptide is expected to be (Figure 36). Deamidation leads to loss of NH_2 but addition of OH, a total mass increase of 1 a.m.u. Since the Gln61-containing peptide has 4 charges on it, a total mass difference of 1 a.m.u. would appear as an enrichment of the peak 0.25 a.m.u larger than the monoisotopic peak, as was observed in the case of Cdc42 (Figure 36, red arrows). The change is not large so a further co-expression with C-terminal SseI and C-terminal C178A was done in case the N-terminal STE domain was inhibiting the activity of the enzyme. Again Cdc42 showed a slight enrichment in the expected position, suggesting that some amount of deamidation was occurring.

The changes observed by mass spectrometry were very small, so we decided to use native gel electrophoresis as a way of confirming that the peak enrichments seen were actually due to deamidation of Gln61. A Gln to Glu conversion changes the theoretical pI of Cdc42 from 4.92 to 4.81, which should result in a mobility difference in native gel electrophoresis, as was reported for deamidated RhoA and Rac1 (Flatau et al 2000). A Cdc42 N61E mutant was used as a positive control to find conditions that would

differentiate the two forms of Cdc42. However, initial attempts have not yet found an observable separation, and we are now trying to use isoelectric focusing to confirm deamidation of Cdc42.

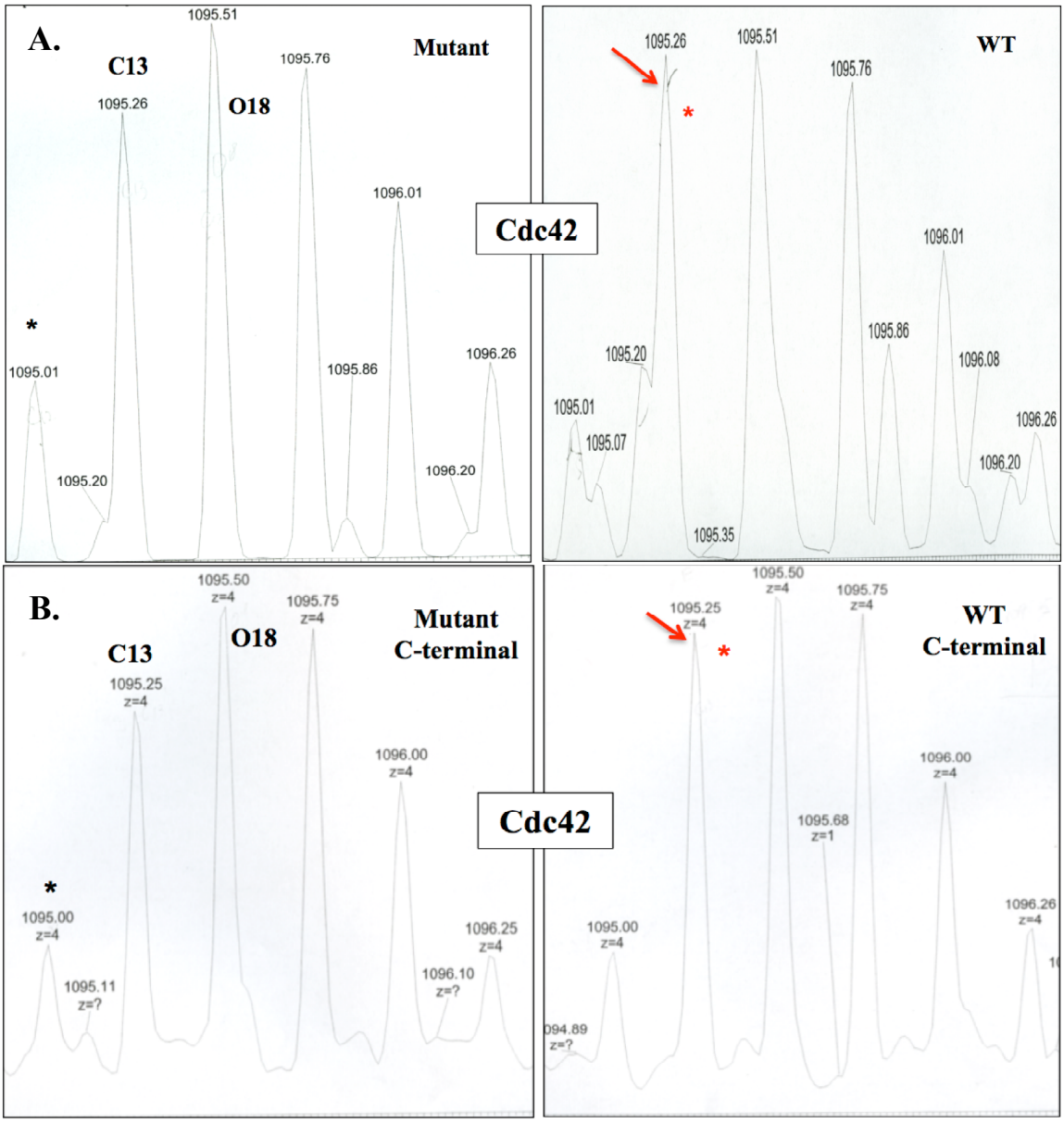


Figure 34: ESI-MS peptide masses from co-expression of Cdc42 and: (A) WT/mutant FL SseI or (B) WT/mutant C-terminal SseI.

The peptide being detected is, FPSEYVPTVFDNYAVTVMIGGEPYTLGLFDTAGQEDYDR, with theoretical m/z = 1095.01, z = 4. The monoisotopic peak (*) and position of the deamidated peak (*) (which has the same mass as the C13 unmodified peak) are highlighted.

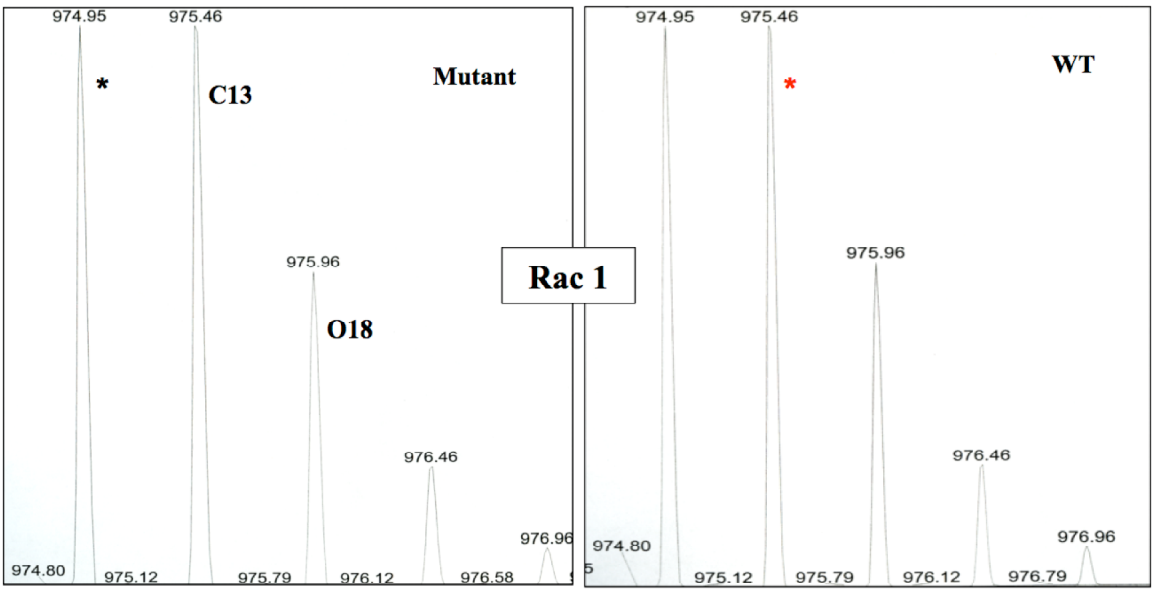


Figure 35: ESI-MS peptide masses from co-expression of Rac1 and WT or mutant FL SseI. The peptide being detected is, PVNLGLWDTAGQEDYDR, with theoretical m/z = 974.95, z =2. The monoisotopic peak (*) and position of the deamidated peak (*) (which has the same mass as the C13 unmodified peak) are highlighted.

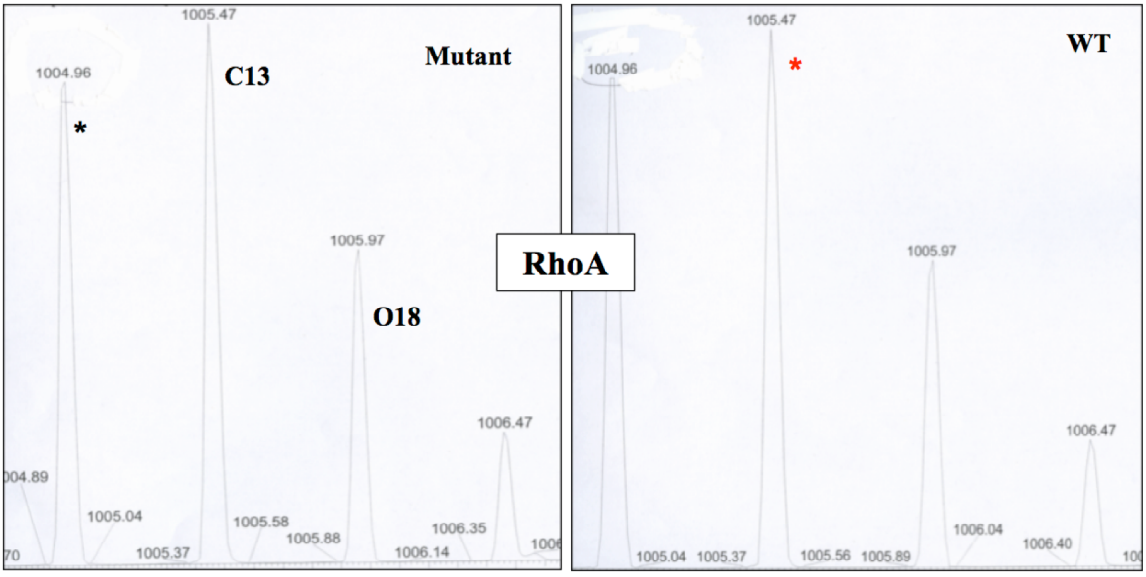


Figure 36: ESI-MS peptide masses from co-expression of RhoA and WT or mutant FL SseI. The peptide being detected is, QVELALWDTAGQEDYDR, with theoretical m/z = 1004.96, z = 2. The monoisotopic peak (*) and position of the deamidated peak (*) (which has the same mass as the C13 unmodified peak) are highlighted.

7.8 Possible activators of SseI

The slight amount of Cdc42 modification seen in the co-expression experiments suggests that SseI is still not fully activated in *E. coli* cells. It might be that if the enzyme has both active and inactive forms, through thermal motion there will still be a small amount of active enzyme present in equilibrium with the inactive form during co-expression, even in the absence of activating conditions. The large amounts of expressed protein, long incubation period (24 hrs), and close proximity to Cdc42 in the *E. coli* cytosol could then lead to some observable deamidation, unlike the *in vitro* assays which were monitored for only 30 mins.

The requirement for an additional bacterial or host factor for activation of cysteine protease domains (CPDs) has been shown in other virulence effectors. For instance, the *Vibrio cholerae* toxin, RTX (a member of the multifunctional MARTX family of toxins), has a CPD that is activated by inositol hexakisphosphate ($InsP_6$) once it is has been transported into the host cytosol. $InsP_6$ binding leads to toxin autoproteolysis by the CPD and release of its other effector domains. The structure of the $InsP_6$/RTX complex showed that the $InsP_6$ allosteric binding site was distinct from the CPD catalytic site, indicating that it was not acting as a co-factor (Lupardus et al 2008).

Two possibilities for this activating role with SseI are: TRIP6, which was identified as a possible interaction partner through a yeast two-hybrid screen (but not proven biochemically) (Worley et al 2006), and IQGAP1, for which a direct interaction with SseI

was shown both in macrophages and *in vitro* (McLaughlin et al 2009). We tried expressing human full length TRIP6 and 276-476 (which contains the minimal interaction domain shown by the screen) in *E. coli* cells, but the constructs were extremely insoluble and toxic to cells. Using pLysS strains (tighter induction control) and Rosetta strains (containing genes for 'rare' codon tRNAs) of *E. coli*, or co-expression with SseI constructs, did not increase yields. In addition, the group that demonstrated the biochemical interaction of IQGAP1 with SseI was not able to recapitulate the interaction with TRIP6, casting some doubt as to whether it is an actual binding partner *in vivo*. As a result, we did not pursue the TRIP6 purification any further.

IQGAP1 is a good candidate as an activating partner of SseI since, it has been shown to bind both Cdc42 and Rac1 (Owen et al 2008), is a central regulator of actin polymerization dynamics (Brandt & Grosse 2007), and has been directly implicated in *Salmonella* invasion of cells (Brown et al 2007). Its role as a scaffolding protein might be key, localizing both SseI and Cdc42 together at regions of active actin polymerization. We are in the process of testing purified IQGAP1 as a possible activator of SseI deamidation activity with the generalized substrates Z-Gln-Gly and casein, and with the Rhos. Whole cell lysate from macrophages and macrophage-derived cell lines will also be investigated for an activating effect on SseI in these assays.

The physiological relevance of Cdc42 as a possible deamidation substrate of SseI in the context of *Salmonella* infection remains to be confirmed. One way to do so would be to transfect primary macrophages or macrophage-like cell lines with SseI and then isolate

out cellular Cdc42 using immobilized antibodies or downstream binding proteins such PAK protein kinase (Pellegrin & Mellor 2008). Any deamidation of Gln61 can then be ascertained by mass spectrometry. One caveat is that in cells with constitutively active Rho GTPases (mutant transfections or toxin treatment), the activated Rhos are quickly ubiquitin-tagged and degraded by the proteosome (Doye et al 2006). However, if a depletion of Cdc42 is observed in WT but not mutant SseI transfections, it would still provide indirect evidence that such a modification is occurring.

CHAPTER 8 - CONCLUSIONS

In this work were able to determine the domain organization of SseI and obtain a crystal structure of its catalytic C-terminal domain (reductively methylated, residues 137-322), by SAD to 1.70 Å. Based on its structural homologs, SseI is a member of a particular subgroup (Clan CA) of the Cysteine Protease Superfamily. These enzymes have acyl hydrolase and acyl transferase activities against a wide range of amide-derived moieties. Based on the structural alignments, SseI's catalytic triad was also identified as being Cys178, His216, and Asp231. Furthermore, categorizations within the structural matches identified five possible biochemical activities for further testing, namely: protease, Gln deamidase, transglutaminase, arylamine N-acetyltransferase, and Peptide N-glycanase.

In vitro assays based on small molecule and peptide substrates were done for the activities described (except PNGase), but neither the full-length enzyme nor the isolated catalytic domain showed enzymatic activity in any of the assays. Recent work (Orth et al 2009) has shown that the closest structural (and sequence) match, PMT, has deamidase activity against Gln205 of the Gα subunits of heterotrimeric G-proteins. Conversion of Gln to Glu disrupts the intrinsic GTPase activity of these enzymes, keeping them constitutively active. Based on its cellular localization and phenotype, we hypothesized that SseI might be acting similarly against members of the Rho GTPase family. We tested Cdc42, Rac1, and RhoA, for possible deamidation, since they directly link cellular signalling pathways to changes in cellular motility, cell adherence, cell polarity, and

cytoskeletal rearrangement. That other bacterial toxins such as CNF1 and DNT have been shown to attack these three proteins in the same way *in vivo*, modifying Gln 61/63 to Glu and making the Rhos constitutively active, also supported this hypothesis. Moreover, there are four water molecules in a small cavity behind the surface accessible Cys178, suggesting that water (and not a larger molecule) is acting as the attacking group during catalysis.

No evidence of deamidation of the Rhos was seen in the *in vitro* ammonia release assays. However, co-expressing the Rhos with SseI in *E. coli* cells, followed by mass spectrometric analysis, showed a small modification of Cdc42, specifically in the peptide containing Gln61 (in both full-length and C-terminal SseI co-expressions). Co-expressed Rac1 and RhoA did not show any evidence of deamidation. Since a Gln to Glu conversion should also lead to a pI shift and altered protein mobility in an electric field, we are now in the process of confirming this modification by native gel electrophoresis.

It is possible that an additional host or *Salmonella* factor is required to convert SseI into a fully active conformation, as has been shown for some auto-processing toxins that contain cysteine protease domains. We are testing one of its previously identified interaction partners, IQGAP1, as a likely candidate for this role. Whole cell lysates from macrophage and macrophage-derived cell lines will also be tested for an activating effect on SseI.

Finally, the physiological relevance of Cdc42 deamidation by SseI during *Salmonella* infection remains to be confirmed.

PUBLICATIONS

Bhaskaran SS, Stebbins CE. (2010). The Salmonella secreted virulence effector SseI is a structural member of the Cysteine Protease Superfamily. *(manuscript in preparation)*

Esquerra RM, Jensen RA, **Bhaskaran S**, Pillsbury ML, Mendoza JL, Lintner BW, Kliger DS, Goldbeck RA. (2008). The pH dependence of heme pocket hydration and ligand rebinding kinetics in photodissociated carbonmonoxymyoglobin. *J Biol Chem*. 283(20): 14165-75.

Velloso LM, **Bhaskaran SS**, Schuch R, Fischetti VA, Stebbins CE. (2008). A structural basis for the allosteric regulation of non-hydrolysing UDP-GlcNAc 2-epimerases. *EMBO Rep*. 9(2): 199-205. *(co-author)*

Bhaskaran SS, Stebbins CE. (2007). Designer bugs: structural engineering to build a better mouse model. *Cell Host Microbe.* (4): 241-3.

Goldbeck RA, **Bhaskaran S**, Ortega C, Mendoza JL, Olson JS, Soman J, Kliger DS, Esquerra RM. (2006). Water and ligand entry in myoglobin: assessing the speed and extent of heme pocket hydration after CO photodissociation. *Proc Natl Acad Sci U S A*. 103(5): 1254-9.

BIBLIOGRAPHY

Abrahams GL, Hensel M. 2006. Manipulating cellular transport and immune responses: dynamic interactions between intracellular Salmonella enterica and its host cells. *Cell Microbiol* 8: 728-37

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, et al. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66: 213-21

Akeda Y, Galan JE. 2005. Chaperone release and unfolding of substrates in type III secretion. *Nature* 437: 911-5

Aminova LR, Luo S, Bannai Y, Ho M, Wilson BA. 2008. The C3 domain of Pasteurella multocida toxin is the minimal domain responsible for activation of Gq-dependent calcium and mitogenic signaling. *Protein Sci* 17: 945-9

Badger J, Sauder JM, Adams JM, Antonysamy S, Bain K, et al. 2005. Structural analysis of a set of proteins resulting from a bacterial genomics project. *Proteins* 60: 787-96

Barrett AJ, Rawlings ND. 2001. Evolutionary lines of cysteine peptidases. *Biol Chem* 382: 727-33

Beyer W, Turnbull PC. 2009. Anthrax in animals. *Mol Aspects Med* 30: 481-9

Brandt DT, Grosse R. 2007. Get to grips: steering local actin dynamics with IQGAPs. *EMBO Rep* 8: 1019-23

Brooke EW, Davies SG, Mulvaney AW, Pompeo F, Sim E, Vickers RJ. 2003. An approach to identifying novel substrates of bacterial arylamine N-acetyltransferases. *Bioorg Med Chem* 11: 1227-34

Brown MD, Bry L, Li Z, Sacks DB. 2007. IQGAP1 regulates Salmonella invasion through interactions with actin, Rac1, and Cdc42. *J Biol Chem* 282: 30265-72

Brown MD, Sacks DB. 2006. IQGAP1 in cellular signaling: bridging the GAP. *Trends Cell Biol* 16: 242-9

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, et al. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54: 905-21

Bustelo XR, Sauzeau V, Berenjeno IM. 2007. GTP-binding proteins of the Rho/Rac family: regulation, effectors and functions in vivo. *Bioessays* 29: 356-70

Campbell RE, Mosimann SC, Tanner ME, Strynadka NC. 2000. The structure of UDP-N-acetylglucosamine 2-epimerase reveals homology to phosphoglycosyl transferases. *Biochemistry* 39: 14993-5001

Cieslak TJ, Eitzen EM, Jr. 1999. Clinical and epidemiologic principles of anthrax. *Emerg Infect Dis* 5: 552-5

Collier RJ. 2009. Membrane translocation by anthrax toxin. *Mol Aspects Med* 30: 413-22

Collier RJ, Young JA. 2003. Anthrax toxin. *Annu Rev Cell Dev Biol* 19: 45-70

Cornelis GR. 2006. The type III secretion injectisome. *Nat Rev Microbiol* 4: 811-25

Cybulski RJ, Jr., Sanz P, O'Brien AD. 2009. Anthrax vaccination strategies. *Mol Aspects Med* 30: 490-502

Deane JE, Abrusci P, Johnson S, Lea SM. 2010. Timing is everything: the regulation of type III secretion. *Cell Mol Life Sci* 67: 1065-75

Doye A, Boyer L, Mettouchi A, Lemichez E. 2006. Ubiquitin-mediated proteasomal degradation of Rho proteins by the CNF1 toxin. *Methods Enzymol* 406: 447-56

Driks A. 2009. The Bacillus anthracis spore. *Mol Aspects Med* 30: 368-73

Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60: 2126-32

Engel J, Balachandran P. 2009. Role of Pseudomonas aeruginosa type III effectors in disease. *Curr Opin Microbiol* 12: 61-6

Evans PR. 1997. *Data Reduction*. Presented at Proceedings of CCP4 Study Weekend

Figueroa-Bossi N, Bossi L. 1999. Inducible prophages contribute to Salmonella virulence in mice. *Mol Microbiol* 33: 167-76

Fischetti VA. 2008. Bacteriophage lysins as effective antibacterials. *Curr Opin Microbiol* 11: 393-400

Flatau G, Landraud L, Boquet P, Bruzzone M, Munro P. 2000. Deamidation of RhoA glutamine 63 by the Escherichia coli CNF1 toxin requires a short sequence of the GTPase switch 2 domain. *Biochem Biophys Res Commun* 267: 588-92

Foley SL, Lynne AM. 2008. Food animal-associated Salmonella challenges: pathogenicity and antimicrobial resistance. *J Anim Sci* 86: E173-87

Fouet A. 2009. The surface of Bacillus anthracis. *Mol Aspects Med* 30: 374-85

Franz DR. 2009. Preparedness for an anthrax attack. *Mol Aspects Med* 30: 503-10

Galan JE, Wolf-Watz H. 2006. Protein delivery into eukaryotic cells by type III secretion machines. *Nature* 444: 567-73

Galen JE, Pasetti MF, Tennant S, Ruiz-Olvera P, Sztein MB, Levine MM. 2009. Salmonella enterica serovar Typhi live vector vaccines finally come of age. *Immunol Cell Biol* 87: 400-12

Goossens PL. 2009. Animal models of human anthrax: the Quest for the Holy Grail. *Mol Aspects Med* 30: 467-80

Griffin M, Casadio R, Bergamini CM. 2002. Transglutaminases: nature's biological glues. *Biochem J* 368: 377-96

Hakoshima T, Shimizu T, Maesaki R. 2003. Structural basis of the Rho GTPase signaling. *J Biochem* 134: 327-31

Hansen-Wester I, Stecher B, Hensel M. 2002. Analyses of the evolutionary distribution of Salmonella translocated effectors. *Infect Immun* 70: 1619-22

Haraga A, Ohlson MB, Miller SI. 2008. Salmonellae interplay with host cells. *Nat Rev Microbiol* 6: 53-66

He SY, Nomura K, Whittam TS. 2004. Type III protein secretion mechanism in mammalian and plant pathogens. *Biochim Biophys Acta* 1694: 181-206

Heasman SJ, Ridley AJ. 2008. Mammalian Rho GTPases: new insights into their functions from in vivo studies. *Nat Rev Mol Cell Biol* 9: 690-701

Hinderlich S, Stasche R, Zeitler R, Reutter W. 1997. A bifunctional enzyme catalyzes the first two steps in N-acetylneuraminic acid biosynthesis of rat liver. Purification and characterization of UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase. *J Biol Chem* 272: 24313-8

Holm L, Kaariainen S, Rosenstrom P, Schenkel A. 2008. Searching protein structure databases with DaliLite v.3. *Bioinformatics* 24: 2780-1

Hudson MJ, Beyer W, Bohm R, Fasanella A, Garofolo G, et al. 2008. Bacillus anthracis: balancing innocent research with dual-use potential. *Int J Med Microbiol* 298: 345-64

Hugh-Jones M, Blackburn J. 2009. The ecology of Bacillus anthracis. *Mol Aspects Med* 30: 356-67

Kawamura T, Ishimoto N, Ito E. 1979. Enzymatic synthesis of uridine diphosphate N-acetyl-D-mannosaminuronic acid. *J Biol Chem* 254: 8457-65

Kawamura T, Kimura M, Yamamori S, Ito E. 1978. Enzymatic formation of uridine diphosphate N-acetyl-D-mannosamine. *J Biol Chem* 253: 3595-601

Keim P, Gruendike JM, Klevytska AM, Schupp JM, Challacombe J, Okinaka R. 2009. The genome and variation of Bacillus anthracis. *Mol Aspects Med* 30: 397-405

Keppler OT, Hinderlich S, Langner J, Schwartz-Albiez R, Reutter W, Pawlita M. 1999. UDP-GlcNAc 2-epimerase: a regulator of cell surface sialylation. *Science* 284: 1372-6

Kiser KB, Bhasin N, Deng L, Lee JC. 1999. Staphylococcus aureus cap5P encodes a UDP-N-acetylglucosamine 2-epimerase with functional redundancy. *J Bacteriol* 181: 4818-24

Kitadokoro K, Kamitani S, Miyazawa M, Hanajima-Ozawa M, Fukui A, et al. 2007. Crystal structures reveal a thiol protease-like catalytic triad in the C-terminal region of Pasteurella multocida toxin. *Proc Natl Acad Sci U S A* 104: 5139-44

Koehler TM. 2009. Bacillus anthracis physiology and genetics. *Mol Aspects Med* 30: 386-96

Koth CM, Orlicky SM, Larson SM, Edwards AL. 2003. *Use of limited proteolysis to identify protein domains suitable for structural analysis*: Elsevier Inc.

Kuhn HM, Meier-Dieter U, Mayer H. 1988. ECA, the enterobacterial common antigen. *FEMS Microbiol Rev* 4: 195-222

Lan R, Reeves PR, Octavia S. 2009. Population structure, origins and evolution of major Salmonella enterica clones. *Infect Genet Evol* 9: 996-1005

Laskowski RA. 2009. PDBsum new things. *Nucleic Acids Res* 37: D355-9

Lawley TD, Chan K, Thompson LJ, Kim CC, Govoni GR, Monack DM. 2006. Genome-wide screen for Salmonella genes required for long-term systemic infection of the mouse. *PLoS Pathog* 2: e11

Lee JH, Choi JM, Lee C, Yi KJ, Cho Y. 2005. Structure of a peptide:N-glycanase-Rad23 complex: insight into the deglycosylation for denatured glycoproteins. *Proc Natl Acad Sci U S A* 102: 9144-9

Lee SH, Galan JE. 2004. Salmonella type III secretion-associated chaperones confer secretion-pathway specificity. *Mol Microbiol* 51: 483-95

Lemonnier M, Landraud L, Lemichez E. 2007. Rho GTPase-activating bacterial toxins: from bacterial virulence regulation to eukaryotic cell biology. *FEMS Microbiol Rev* 31: 515-34

Lerm M, Schmidt G, Goehring UM, Schirmer J, Aktories K. 1999. Identification of the region of rho involved in substrate recognition by Escherichia coli cytotoxic necrotizing factor 1 (CNF1). *J Biol Chem* 274: 28999-9004

Leslie AGW. 1992. *Recent changes to the MOSFLM package for processing film and image plate data.* . Presented at Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography

Li L, Bin LH, Li F, Liu Y, Chen D, et al. 2005. TRIP6 is a RIP2-associated common signaling component of multiple NF-kappaB activation pathways. *J Cell Sci* 118: 555-63

Lilic M, Stebbins CE. 2004. Re-structuring the host cell: up close with Salmonella's molecular machinery. *Microbes Infect* 6: 1205-11

Lilic M, Vujanac M, Stebbins CE. 2006. A common structural motif in the binding of virulence factors to bacterial secretion chaperones. *Mol Cell* 21: 653-64

Lupardus PJ, Shen A, Bogyo M, Garcia KC. 2008. Small molecule-induced allosteric activation of the Vibrio cholerae RTX cysteine protease domain. *Science* 322: 265-8

MacPherson GG, Jenkins CD, Stein MJ, Edwards C. 1995. Endotoxin-mediated dendritic cell release from the intestine. Characterization of released dendritic cells and TNF dependence. *J Immunol* 154: 1317-22

Maldonado-Arocho FJ, Fulcher JA, Lee B, Bradley KA. 2006. Anthrax oedema toxin induces anthrax toxin receptor expression in monocyte-derived cells. *Mol Microbiol* 61: 324-37

Marlovits TC, Kubori T, Lara-Tejero M, Thomas D, Unger VM, Galan JE. 2006. Assembly of the inner rod determines needle length in the type III secretion injectisome. *Nature* 441: 637-40

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: 1040-2

Marlovits TC, Stebbins CE. 2010. Type III secretion systems shape up as they ship out. *Curr Opin Microbiol* 13: 47-52

McClelland M, Sanderson KE, Clifton SW, Latreille P, Porwollik S, et al. 2004. Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of Salmonella enterica that cause typhoid. *Nat Genet* 36: 1268-74

McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, et al. 2001. Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. *Nature* 413: 852-6

McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. 2005. Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* 61: 458-64

McGhie EJ, Brawn LC, Hume PJ, Humphreys D, Koronakis V. 2009. Salmonella takes control: effector-driven manipulation of the host. *Curr Opin Microbiol* 12: 117-24

McLaughlin LM, Govoni GR, Gerke C, Gopinath S, Peng K, et al. 2009. The Salmonella SPI2 effector SseI mediates long-term systemic infection by modulating host cell migration. *PLoS Pathog* 5: e1000671

Miao EA, Brittnacher M, Haraga A, Jeng RL, Welch MD, Miller SI. 2003. Salmonella effectors translocated across the vacuolar membrane interact with the actin cytoskeleton. *Mol Microbiol* 48: 401-15

Miao EA, Miller SI. 2000. A conserved amino acid sequence directing intracellular type III secretion by Salmonella typhimurium. *Proc Natl Acad Sci U S A* 97: 7539-44

Min JH, Pavletich NP. 2007. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* 449: 570-5

Moayeri M, Leppla SH. 2009. Cellular and systemic effects of anthrax lethal toxin and edema toxin. *Mol Aspects Med* 30: 439-55

Mock M, Fouet A. 2001. Anthrax. *Annu Rev Microbiol* 55: 647-71

Morgan PM, Sala RF, Tanner ME. 1997. Eliminations in the Reactions Catalyzed by UDP-N-Acetylglucosamine 2-Epimerase. *J Am Chem Soc* 119: 10269 -77

Morita-Ishihara T, Ogawa M, Sagara H, Yoshida M, Katayama E, Sasakawa C. 2006. Shigella Spa33 is an essential C-ring component of type III secretion machinery. *J Biol Chem* 281: 599-607

Mueller CA, Broz P, Cornelis GR. 2008. The type III secretion system tip complex and translocon. *Mol Microbiol* 68: 1085-95

Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53: 240-55

Noguchi K, Ishikawa K, Yokoyama K, Ohtsuka T, Nio N, Suzuki E. 2001. Crystal structure of red sea bream transglutaminase. *J Biol Chem* 276: 12055-9

Orth JH, Fester I, Preuss I, Agnoletto L, Wilson BA, Aktories K. 2008. Activation of Galpha (i) and subsequent uncoupling of receptor-Galpha(i) signaling by Pasteurella multocida toxin. *J Biol Chem* 283: 23288-94

Orth JH, Lang S, Taniguchi M, Aktories K. 2005. Pasteurella multocida toxin-induced activation of RhoA is mediated via two families of G{alpha} proteins, G{alpha}q and G{alpha}12/13. *J Biol Chem* 280: 36701-7

Orth JH, Preuss I, Fester I, Schlosser A, Wilson BA, Aktories K. 2009. Pasteurella multocida toxin activation of heterotrimeric G proteins by deamidation. *Proc Natl Acad Sci U S A* 106: 7179-84

Otwinowski M. 1997. Processing of X-ray Diffraction Data Collected in Oscillation Mode. In *Macromolecular Crystallography, Part A*, pp. 307-26

Owen D, Campbell LJ, Littlefield K, Evetts KA, Li Z, et al. 2008. The IQGAP1-Rac1 and IQGAP1-Cdc42 interactions: interfaces differ between the complexes. *J Biol Chem* 283: 1692-704

Parsot C, Hamiaux C, Page AL. 2003. The various and varying roles of specific chaperones in type III secretion systems. *Curr Opin Microbiol* 6: 7-14

Pellegrin S, Mellor H. 2008. Rho GTPase Activation Assays. In *Current Protocols in Cell Biology*: Wiley Interscience

Perrakis A, Morris R, Lamzin VS. 1999. Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6: 458-63

Pohanka M, Skladal P. 2009. Bacillus anthracis, Francisella tularensis and Yersinia pestis. The most important bacterial warfare agents - review. *Folia Microbiol (Praha)* 54: 263-72

Quezada CM, Hicks SW, Galan JE, Stebbins CE. 2009. A family of Salmonella virulence factors functions as a distinct class of autoregulated E3 ubiquitin ligases. *Proc Natl Acad Sci U S A* 106: 4864-9

Raffatellu M, Chessa D, Wilson RP, Tukel C, Akcelik M, Baumler AJ. 2006. Capsule-mediated immune evasion: a new hypothesis explaining aspects of typhoid fever pathogenesis. *Infect Immun* 74: 19-27

Rawlings ND, Barrett AJ, Bateman A. 2009. MEROPS: the peptidase database. *Nucleic Acids Res* 38: D227-33

Rayment I. 1997. *Reductive alkylation of lysine residues to alter crystallization properties of proteins*: Elsevier Inc.

Read TD, Peterson SN, Tourasse N, Baillie LW, Paulsen IT, et al. 2003. The genome sequence of Bacillus anthracis Ames and comparison to closely related bacteria. *Nature* 423: 81-6

Remaut H, Waksman G. 2004. Structural biology of bacterial pathogenesis. *Curr Opin Struct Biol* 14: 161-70

Roselin LS, Lin MS, Lin PH, Chang Y, Chen WY. 2010. Recent trends and some applications of isothermal titration calorimetry in biotechnology. *Biotechnol J* 5: 85-98

Rost B, Yachdav G, Liu J. 2004. The PredictProtein server. *Nucleic Acids Res* 32: W321-6

Rowe B, Ward LR, Threlfall EJ. 1997. Multidrug-resistant Salmonella typhi: a worldwide epidemic. *Clin Infect Dis* 24 Suppl 1: S106-9

Rychlik I, Karasova D, Sebkova A, Volf J, Sisak F, et al. 2009. Virulence potential of five major pathogenicity islands (SPI-1 to SPI-5) of Salmonella enterica serovar Enteritidis for chickens. *BMC Microbiol* 9: 268

Saier MH, Jr. 2004. Evolution of bacterial type III protein secretion systems. *Trends Microbiol* 12: 113-5

Salyers AA, Whitt DD. 2002. *Bacterial Pathogenesis: A Molecular Approach*: American Society of Microbiology

Samuel J, Tanner ME. 2004. Active site mutants of the "non-hydrolyzing" UDP-N-acetylglucosamine 2-epimerase from Escherichia coli. *Biochim Biophys Acta* 1700: 85-91

Schuch R, Fischetti VA. 2009. The secret life of the anthrax agent Bacillus anthracis: bacteriophage-mediated ecological adaptations. *PLoS One* 4: e6532

Schwartz M. 2009. Dr. Jekyll and Mr. Hyde: a short history of anthrax. *Mol Aspects Med* 30: 347-55

Scorpio A, Chabot DJ, Day WA, O'Brien D K, Vietri NJ, et al. 2007. Poly-gamma-glutamate capsule-degrading enzyme treatment enhances phagocytosis and killing of encapsulated Bacillus anthracis. *Antimicrob Agents Chemother* 51: 215-22

Self AJ, Hall A, eds. 1995. *Purification of Recombinant Rho/Rac/G25K from Escherichia coli*, Vols. 256: Academic Press. 3-10 pp.

Shao F, Merritt PM, Bao Z, Innes RW, Dixon JE. 2002. A Yersinia effector and a Pseudomonas avirulence protein define a family of cysteine proteases functioning in bacterial pathogenesis. *Cell* 109: 575-88

Sinclair JC, Sandy J, Delgoda R, Sim E, Noble ME. 2000. Structure of arylamine N-acetyltransferase reveals a catalytic triad. *Nat Struct Biol* 7: 560-4

Smith SJM, Rittinger K. 2002. Preparation of GTPases for Structural and Biophysical Analysis. In *GTPase Protocols: The Ras Superfamily*, pp. 13-24: Springer Protocols

Stasche R, Hinderlich S, Weise C, Effertz K, Lucka L, et al. 1997. A bifunctional enzyme catalyzes the first two steps in N-acetylneuraminic acid biosynthesis of rat liver. Molecular cloning and functional expression of UDP-N-acetyl-glucosamine 2-epimerase/N-acetylmannosamine kinase. *J Biol Chem* 272: 24319-24

Stavrinides J, Ma W, Guttman DS. 2006. Terminal reassortment drives the quantum evolution of type III effectors in bacterial pathogens. *PLoS Pathog* 2: e104

Stebbins CE, Galan JE. 2001. Structural mimicry in bacterial virulence. *Nature* 412: 701-5

Suzuki T. 2005. A simple, sensitive in vitro assay for cytoplasmic deglycosylation by peptide: N-glycanase. *Methods* 35: 360-5

Tang WJ, Guo Q. 2009. The adenylyl cyclase activity of anthrax edema factor. *Mol Aspects Med* 30: 423-30

Tanner ME. 2002. Understanding nature's strategies for enzyme-catalyzed racemization and epimerization. *Acc Chem Res* 35: 237-46

Tierrez A, Garcia-del Portillo F. 2005. New concepts in Salmonella virulence: the importance of reducing the intracellular growth rate in the host. *Cell Microbiol* 7: 901-9

Tonello F, Montecucco C. 2009. The anthrax lethal factor and its MAPK kinase-specific metalloprotease activity. *Mol Aspects Med* 30: 431-8

Tournier JN, Rossi Paccani S, Quesnel-Hellmann A, Baldari CT. 2009. Anthrax toxins: a weapon to systematically dismantle the host immune defenses. *Mol Aspects Med* 30: 456-66

Turk BE. 2007. Manipulation of host signalling pathways by anthrax toxins. *Biochem J* 402: 405-17

Upton A, Johnson N, Sandy J, Sim E. 2001. Arylamine N-acetyltransferases - of mice, men and microorganisms. *Trends Pharmacol Sci* 22: 140-6

van der Goot G, Young JA. 2009. Receptors of anthrax toxin and cell entry. *Mol Aspects Med* 30: 406-12

Vojtova J, Kamanova J, Sebo P. 2006. Bordetella adenylate cyclase toxin: a swift saboteur of host defense. *Curr Opin Microbiol* 9: 69-75

Worley MJ, Ching KH, Heffron F. 2000. Salmonella SsrB activates a global regulon of horizontally acquired genes. *Mol Microbiol* 36: 749-61

Worley MJ, Nieman GS, Geddes K, Heffron F. 2006. Salmonella typhimurium disseminates within its host by manipulating the motility of infected cells. *Proc Natl Acad Sci U S A* 103: 17915-20

Xu J, Lai YJ, Lin WC, Lin FT. 2004. TRIP6 enhances lysophosphatidic acid-induced cell migration by interacting with the lysophosphatidic acid 2 receptor. *J Biol Chem* 279: 10459-68

Yamaguchi S, Jeenes DJ, Archer DB. 2001. Protein-glutaminase from Chryseobacterium proteolyticum, an enzyme that deamidates glutaminyl residues in proteins. Purification, characterization and gene cloning. *Eur J Biochem* 268: 1410-21

Zhang S, Kingsley RA, Santos RL, Andrews-Polymenis H, Raffatellu M, et al. 2003. Molecular pathogenesis of Salmonella enterica serotype typhimurium-induced diarrhea. *Infect Immun* 71: 1-12

Zhu M, Shao F, Innes RW, Dixon JE, Xu Z. 2004. The crystal structure of Pseudomonas avirulence protein AvrPphB: a papain-like fold with a distinct substrate-binding site. *Proc Natl Acad Sci U S A* 101: 302-7