Identification of an IcsB-like Acyltransferase Family

Waad Bajunaid

Thesis submitted to the University of Ottawa

in partial Fulfillment of the requirements for the

Master's degree in Chemistry

Department of Chemistry and Biomolecular Sciences

Faculty of Science

University of Ottawa

Candidate	**Supervisor**
Waad Bajunaid	Professor Francois-Xavier Campbell-Valois

Acknowledgment

Foremost, I would like to thank François-Xavier Campbell-Valois for his invaluable support and guidance throughout my project. I would also like to thank the members of my thesis committee Dr. John Pezacki, and Dr. Christopher Boddy, for their insightful comments and feedback. I would like to thank all members of Host-Microbe Interaction Lab for making work in the lab enjoyable and memorable.

I am grateful for The Ministry of Higher Education in Saudi Arabia and the Saudi Arabian Cultural Bureau in Canada for supporting my graduate study.

Finally, I would like to express gratitude to my parents, siblings, and friends for their endless support and motivation. Without their encouragement, this work would have been much more challenging. I am grateful to have you as part of my life.

.

Abstract

Shigella flexneri is a Gram-negative enteropathogen and one of the main causes of diarrheal diseases. *Shigella* uses its type three secretion system (T3SS) to secrete effector proteins into host cells, thus allowing their invasion. Through vacuole rupture, Shigella cells are transferred into the host cell cytosol whereby eliciting actin-based movement they disseminate into neighboring cells, a process known as cell-to-cell spread. IcsB is one of the secreted proteins that is crucial for *Shigella* to help escaping the vacuole formed during cell-to-cell spread. Recently, it was proven that IcsB has acyltransferase activity that is essential to vacuole escape. Surprisingly, using a bioinformatic approach, we were able to identify proteins that share homology with IcsB when we queried against the catalytic domain. Those proteins have high homology with IcsB that is limited in most cases to a central segment of their primary structure that corresponds to the catalytic domain. Furthermore, the three catalytic residues Histidine 145, Aspartate 195 and Cystine 306 are completely conserved in all homologous proteins. Interestingly, most of those homologs have not been studied yet. Thus, we hypothesized that these proteins might be forming an IcsB-like acyltransferase family. Based on previous findings that IcsB expression is toxic in budding yeast, we established a yeast cytotoxicity to test whether the putative IcsB homologs triggered the same phenotype. The results demonstrated that seven out of 11proteins were toxic to yeast. Mutating the catalytic histidine rescued the toxicity observed in their wildtype counterpart. The results in this thesis are the first step toward confirming the relation between IcsB and the homologous proteins we discovered. Additional experiments from our lab spurred by my data have confirmed that several of the toxic homologs are acyl transferase, while those that were not toxic are not. Future work needs to be conducted to determine whether the acylation activity of these proteins contributes to pathogenesis.

Table of Contents

List of figures

List of tables:

List of abbreviations:

T3SS	Type three secretion system
T3SA	Type three secretion apparatus
AMP	Ampicillin
WB	Western Blot
HRP	horseradish peroxidase
SD-His	Yeast single dropout yeast media/ without histidine
SD- Ura	Yeast single dropout media/ without uracil
DDO-His-Ura	Yeast double dropout media/ without histidine and uracil
EV	Empty vector
H1	*Burkholderia pseudomallei* protein
H2	*Escherichia marmotae* protein
H3	*Chromobacterium sp. LK11* protein
H4	*Achromobacter arsenitoxydans* protein
H5	*Sodalis praecaptivus* protein
H6	*Achromobacter sp. Rta* protein
H7	*Endozoicomonas montiporae* protein
H8	*Robbsia andropogonis* protein
H9	*Achromobacter denitrificans* protein
H10	*Desulfonatronospira thiodismutans* protein
H11	*Ideonella sakaiensis* protein

Chapter I Introduction

According to the World Health Organization (WHO), diarrheal diseases were collectively the 5th most frequent causes of deaths worldwide in 2016 (WHO, 2018). There are various causative agents of diarrhea, such as viral, parasitic, and bacterial infections. One of the main cause of bacteria-induced diarrhea is infection of the large intestine by *Shigella flexneri*. It is a Gram-negative enteropathogenic bacteria that causes gastroenteritis or so-called shigellosis. Its infection starts by invading and replicating on the lining, also called the mucosae, of the large intestine. It is characterized by dysentery a form of diarrhea containing mucus and also quite frequently blood; this main symptom is also often accompanied by nausea, fever, and abdominal cramps (CDC, 2019). *Shigella* is responsible for 164.7 million cases worldwide, and 1.1 million deaths annually (Kotloff et al.,1999). The disease is transmitted through the fecal-oral route from person-to-person or by consuming contaminated food and water, which happens mainly in developing countries due to the lack of sanitation practices.

Shigella genus belongs to the class of gammaproteobacteria from the *Enterobacteriaceae* family, and it is classified into four serogroups based on biochemical characteristics and antigenic structure. These four species are *S. flexneri, S. dysenteriae, S. boydii, and S. sonnei.* The four species are responsible for 69% of diarrheal infection, and 61% rate of death resulted from shigellosis. The majority of infection caused by shigella occurs in children under five years. Even though most of *Shigella* infection found in developing countries due to the lack of sanitation, *Shigella* infections are also encountered in developed countries. The average distribution of *Shigella's* isolates in the developed countries gives the highest percentage of 77% to *S. sonnei*, while *S. flexneri* scored 60% in the developing countries (Kotloff et al.,1999).

Due to the excessive usage of antibiotics to treat bacterial infections, many bacteria have developed high resistance to drugs, thus complicating the treatment of infections. According to the Center for Disease Control, in most cases *Shigella*'s infection, on the other hand, can be treated without the need for antibiotics. Nevertheless, antibiotic drugs, such as azithromycin and others, are used to treat severe cases in elder, infants and immunosuppressed patients. In other case, antibiotics treatments were also shown to decrease the disease period and the symptoms associated with the infection. *Shigella* similarly to other *Enterobacteriaceae* has experienced a dramatic increase in the prevalence of antibiotics resistance to drugs over the years, particularly since 2013 (CDC, 2019). The World Health Organization has recently included *Shigella* in the group of bacteria for which the needs of novel antibiotics are urgent. Pharmaceutical companies and other stakeholders were asked to prioritize their research objective to find new and effective antibiotic drugs for antibiotic-resistant bacteria including *Pseudomonas aeruginosa, Staphylococcus aureus, Salmonella spp., Streptococcus pneumoniae, Shigella spp.,* and others (WHO, 2017).

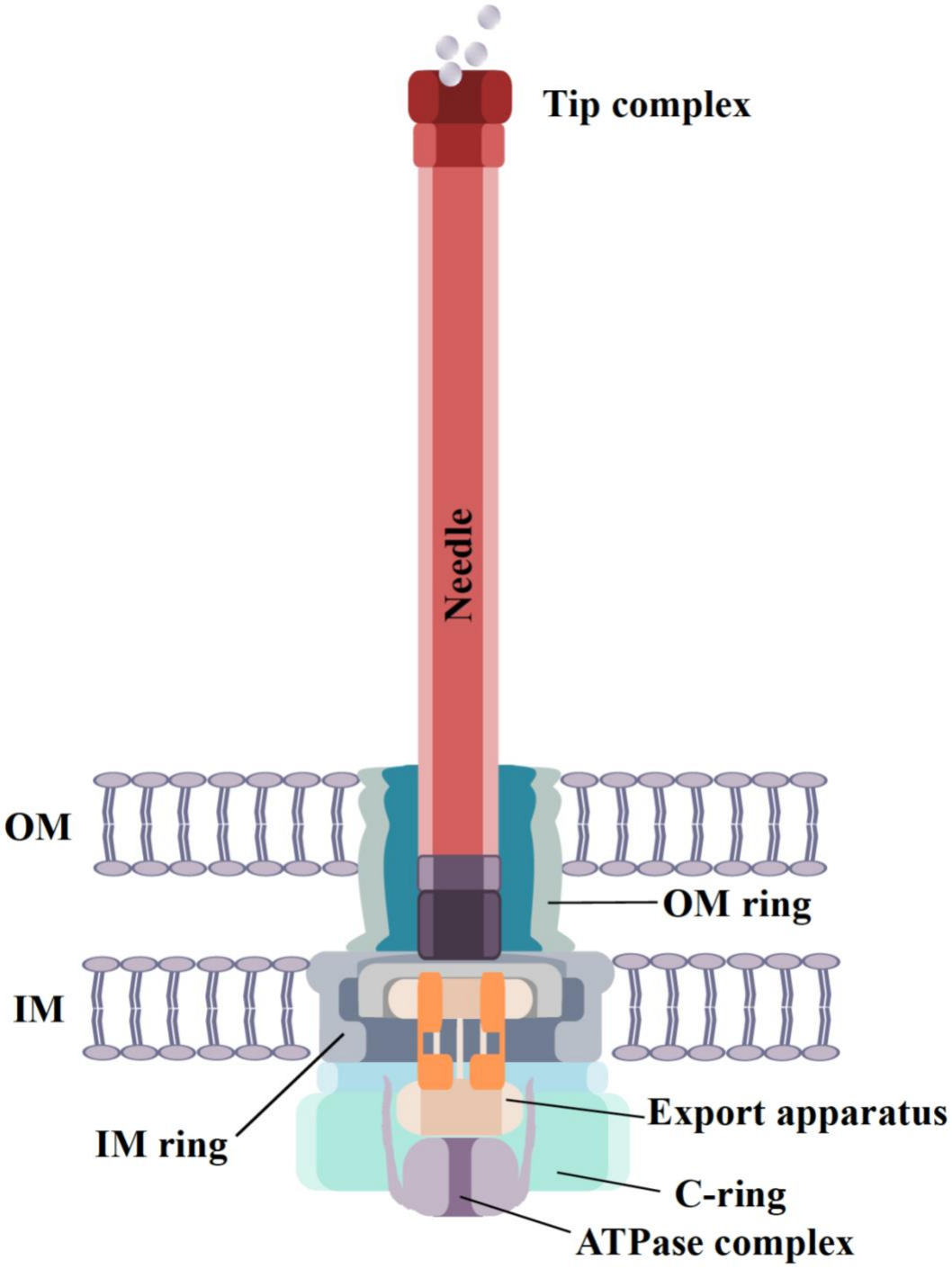
The pathogenicity of *Shigella* species relies on a large virulence plasmid (pWR100) that is 200 kb encoding some of the effector proteins secreted by the bacteria and genes that are essential for its invasion. The virulence plasmid is also found in pathogenic *E. coli* strain K-12 that is considered to be closely related to the bacteria *Shigella* (Sansonetti et al., 1983). It was found that when the virulence plasmid is lost in *Shigella* spp., the bacteria loses its invasiveness making the virulence plasmid crucial for its pathogenicity (Maurelli et al., 1985). It was shown that a segment of 31 kb of the virulence plasmid known as the entry region is critical for the bacterial uptake by epithelial cell. This is because the components of the type three secretion

system and approximately 25 of its secreted proteins are encoded within the entry region (Buchrieser et al., 2000).

Shigella as well as other Gram-negative pathogenic bacteria like *Salmonella*, and *Chlamydia* use type three secretion system (T3SS). This T3SS is used by the bacteria to inject translocator and effector proteins through its needle into the host cell (figure 1.1). These effectors can modulate host immune response and interrupt the cell cytoskeleton during the infection (Deng et al., 2017). The core structure of the type III secretion system composed of different protein parts exhibiting in the needle complex, cytoplasmic ring, basal body, and ATPase complex (Hu et al., 2015 and Campbell-Valois & Pontier, 2016). The expression of the T3SS components and effectors are mostly regulated at the transcriptional level. The type III secretion assembly and expression are regulated by temperature (Maurelli & Sansonetti, 1988). VirB regulates the transcription of some genes located in the entry region of the virulence plasmid including MxiE and IpgC, and its expression is regulated in a temperature-dependent manner (Durand et al., 2000). MxiE gene also regulates transcription of some of *Shigella* genes, such as VirA and IpaH9.8 genes (Mavris et al., 2002). In addition, the secretion is also regulated at the post-translational level by specific T3SS chaperons, and it modulates the effectors that are expressed before the secretion is activated (Campbell-Valois et al., 2016).

Figure 1.1. Structure of type III secretion apparatus.

Adopted from (Deng et al., 2017)

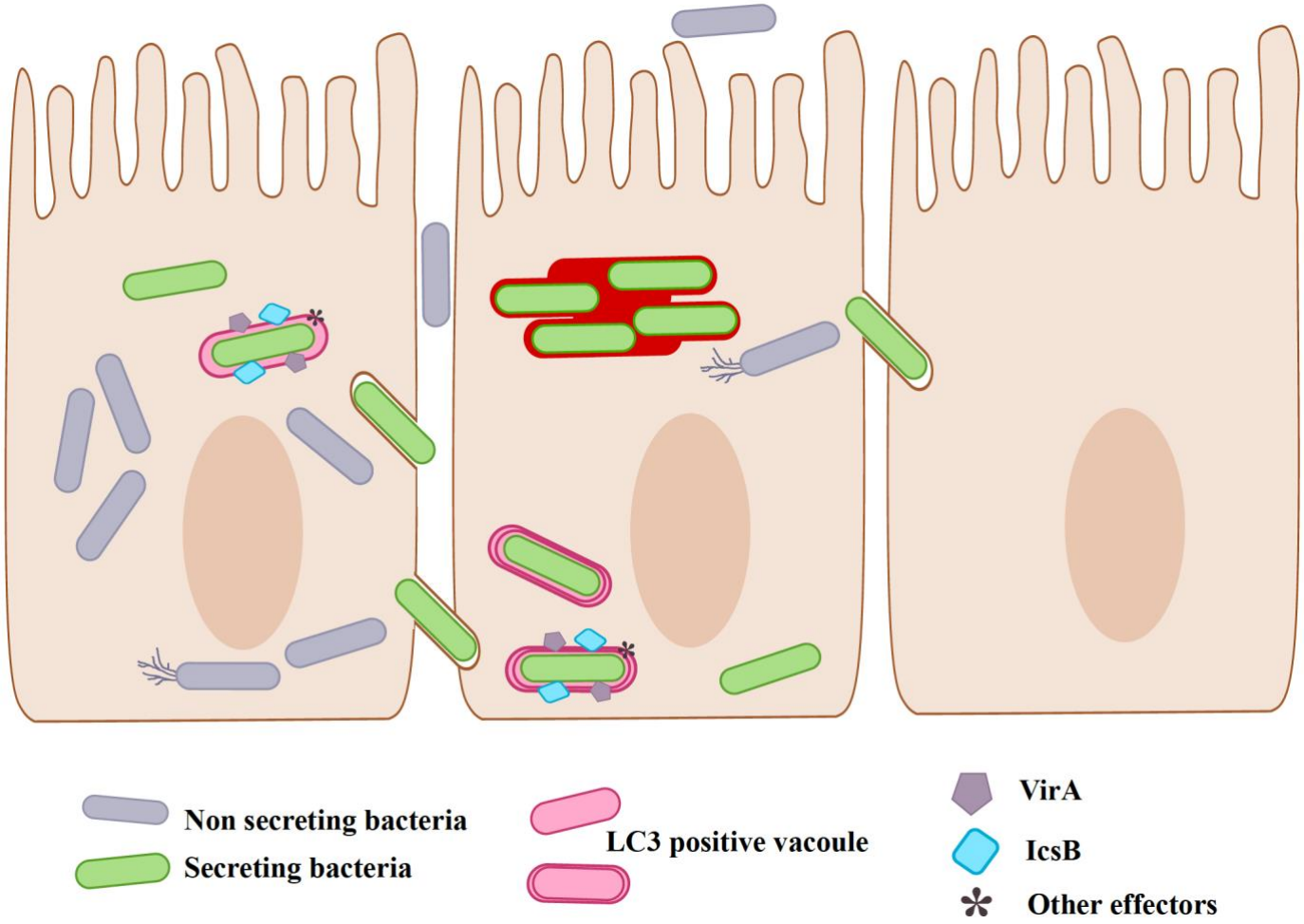


The infection cycle of *Shigella* consists of different steps (figure 1.2). It starts with the initial uptake of the bacteria by human epithelial cells in a process similar to phagocytosis after the release of *Shigella*'s T3S effectors. IcsA protein is responsible of the adhesion of the bacteria to the host cells. After the bacteria is inside the cell, the lysis of its vacuole is stimulated by released effectors and translocators. The bacteria in this stage start to replicate in the host cytoplasm, and its movement is enhanced by the actin comets at one pole induced by the action of the IcsA protein. This bacterial movement allows *Shigella* to spread to neighbouring cells. The cell-to-cell spread is initiated when bacteria inside the host hit the plasma membrane (PM). Thus, leading to the extension of the PM from the infected cell to the neighbouring cell making the bacteria trapped in a protrusion. At the stage where the bacteria arrive in the adjacent cell, a dissemination vacuole is formed, trapping the bacteria inside two membranes. This vacuole is again ruptured with the help of the released translocators IpaB and IpaC. The process is repeated all along, leading to progressive invasion of cells in the monolayer or within an infection foci *in vivo* (Arena et al., 2015 and Campbell-Valois & Pontier, 2016). IcsB, the protein secreted by the type III secretion system, is crucial for *Shigella* to escape the dissemination vacuole. It requires the chaperone IpgA for its production, secretion and stabilization in the bacterial cytosol. Also, the mutant IpgA strain resulted in a reduction of IcsB secretion (Ogawa et al., 2003). Furthermore, SicP, a protein from *Salmonella*, shared homology with IpgA from *S. flexneri* and it was found to be necessary for the SptP protein stabilization, and the SicP mutant strain showed increasing in SptP degradation (Fu & Galán, 1998). The results from Ogawa et al. suggested that both the T3SS chaperone IpgA and SicP are from the same family (Ogawa et al., 2003). Another chaperone that shares homology with both IpgA from *Shigella* and SicP from

Salmonella is the chaperone BicP from *Burkholderia*. It is required to prevent the degradation of the effector BopA (Vander Broek & Stevens, 2017).

Figure 1.2. Schematic of the role of IcsB during the infection of Shigella flexneri

Adopted from (Campbell-Valois et al., 2015).



Autophagy is the process in which the degradation of cytoplasmic contents (e.g. proteins and organelles) occurs in a double membrane compartment. When this is triggered by nutrient or serum starvation, it is often referred to canonical autophagy. On the other hand, xenoautophagy is a type of autophagy that targets foreign particles, such as bacteria and viruses, that have invaded host cells. LC3 is a microtubule-associated protein and a prominent autophagosome marker in both canonical and xenoautophagy. In its canonical form, autophagy is a process in which the cellular components are to be recycled following nutrient or growth factor starvation. Different types of autophagy are responsible for the degradation of different organelles. During the infection cycle, the marker of autophagosome membranes ATG8/LC3 was found to be recruited around actively secreting or recently secreted *Shigella flexneri* during entry and cell-to-cell spread (Ogawa et al., 2005 and Campbell-Valois et al., 2015).

In *Shigella*, there is evidence of many effectors implicated in the vacuole escape process, such as IcsB, VirA and IpaH9.8. Guanylate-binding proteins (GBPs), however, play an essential role in preventing *Shigella* from spreading to neighbouring cells in the presence of Interferon-gamma, only. This phenomenon takes place when the recognition of *Shigella* by GBPs partly or completely inhbits motility and thus cell-to-cell spread. IpaH9.8, a member of the E3 ubiquitin ligase family, reverses the host defence action by degrading GBPs, allowing the bacteria to continue its spreading (Li et al., 2017 and Wandel et al., 2017). In addition, it was suggested that the role VirA in escape autophagy is by activating the GTPase activity of Rab1. The effect of VirA on GTPase activity acts against the autophagy defence and ensure the survival of the intracellular bacteria (Dong et al., 2012).

Furthermore, the outer membrane protein IcsA, also known as VirG, was reported to trigger the host defence by binding Atg5, a core member of the autophagy machinery. It was

proposed that the binding of IcsA to Atg5 is inhibited when IcsB is secreted since the IcsB mutant strain resulted in entrapping the bacteria in autophagy vacuole, and therefore blocking LC3 recruitment (Ogawa et al., 2003 and 2005). Another model, it was suggested that IcsB and VirA act together in allowing the bacteria to escape its LC3 positive vacuole during the cell-to-cell spread process. The recruitment of LC3 was found to be much higher in IcsB and VirA mutant strain (Campbell-Valois et al., 2015).

IcsB, as stated previously, is a protein secreted by the type III secretion system during *Shigella*'s infection. The protein is 494 amino acids in length and its predicted molecular weight is approximately 56 kDa. On top of properties descrived above, the ectopic expression of IcsB inhibits the growth of yeast cells (Slagowski et al., 2008). The use of yeast cells to understand the function of bacterial effectors is not without precedent and is in fact routinely used because of the exquisite sensitivity of this organism to cellular stress (Popa et al., 2016). Furthermore, the study done by Pei & Grishin predicted that IcsB might be functioning as a protease or an acyltransferase (Pei & Grishin, 2009). Moreover, it was recently proven that IcsB possesses acyltransferase activity that has a role in the host cell's cytoskeleton interruption. IcsB was able to acylate the amine group of lysine residues in the polybasic region. The acyltransferase activity was diminished when the catalytic triad histidine-145, aspartate-195, and cysteine-306 (Pei & Grishin, 2009 and Liu et al., 2018) were mutated. It has been demonstrated that the trafficking proteins Rho-GTPase are targeted by IcsB as well as CHMP5, which is implicated in the ESCRT-III complex. Thus, the acyltransferase activity of IcsB helps *Shigella flexneri* to escape the host defence mechanism. Also, in the absence of CHPM5, no LC3 was found around *Shigella* even in the strain of *ΔIcsB* (Liu et al., 2018). This observation contradicts

with the model suggested by Ogawa in 2005 where they suggested that when IcsA is exposed in the cytosol, it induces its recognition by the autophagy protein Atg5.

Furthermore, cell-to-cell spread is encountered in different human pathogens, such as *Listeria monocytogenes* and *Burkholderia pseudomallei* (Sitthidet et al., 2008). As discussed above, *B. pseudomallei* secrete a T3SS effector protein, named BopA, that plays a crucial role to help the bacteria escape its vacuole. Indeed, LC3 was found to be recruited around WT *Burkholderia* and BopA mutant cells displayed elevated level of LC3 in their vicinity (Gong et al., 2011), similarly to what was previously observed in *Shigella*. In an earlier paper, BopA and IcsB were shown to share homology with the Rho GTPases inactivation domain (RID) of MARTX toxins found in *Vibrio cholerae* and related species (Pei & Grishin, 2009). To our knowledge, however, there were no other proteins reported to share homology and possess acyltransferase activities akin to IcsB. Therefore, this observation spurred my thesis work.

In this project, using bioinformatic approaches, 10 uncharacterized proteins from poorly studied bacteria were found to share homology with IcsB. Aligning the sequences of these 11 proteins revealed that they have conservations especially in a region corresponding to the catalytic domain of IcsB and BopA. In addition, the catalytic triad of IcsB (Histidine 145, Aspartic acid 195, and cystine 306) (Pei & Grishin, 2009 and Liu et al., 2018) was found to be perfectly conserved. We hypothesized that these proteins could be forming an IcsB-like acyl transferase family. Since IcsB is toxic to yeast cells, yeast cytotoxicity assay was established to demonstrate if these homologs have the same toxicity phenotype previously observed in IcsB. Furthermore, seven out of the 11 homologs found to kill yeast cells upon expression and their expression was detected through western blot. Mutating the histidine catalytic residue of IcsB and the toxic homologous proteins diminish the toxicity observed in the wild type protein.

Chapter II IcsB Homolog Proteins Materials and Methods

2.1 Gathering information and data about IcsB homologs

2.1.1 PSI-BLAST search:

Position-specific Iterated BLAST (PSI-BLAST) uses the regular BLAST search output to make a profile, called a position-specific scoring matrix, obtained by extracting the sequence information embedded in multiple sequence alignment of proteins with E-value better than the fixed threshold. Afterward, this profile is used to initiate a new search to include more sequences. As the search continues, more and more sequences are added to the profile to expand the search (Bhagwat & Aravind, 2007).

The experimentally determined catalytic domain of IcsB from *Shigella flexneri* (gi| 32307025), residues 128-319 (Pei & Grishin, 2009), was queried against NCBI non-redundant protein database (April 20, 2018). After the first iteration, many proteins appeared by PSI-BLAST using IcsB catalytic domain but only eight were novel. Those eight were picked to initiate a new search (second iteration). After that, the process was repeated for one last time (third iteration) including all proteins from first and second rounds before the search stopped. The algorithm parameters were kept as default (expect threshold: 10, max matches in the query range: 0, matrix: Blosum62, Gap costs: existence 11 Extension: 1, conditional compositional score matrix adjustment, PSI BLAST threshold: 0.005, and no filters nor masks were checked), but for the maximum target sequences which was fixed at 500 to retrieve a higher number of homologous sequences.

2.1.2 TXSScan

TXSScan is a computational tool based on MacSyFinder and used to classify and characterize the type of secretion system encoded by a bacterial genome based on the similarities between its genes and genes that are typical of the corresponding secretion system. It can be used for searching known systems. In this case for instance, the TXSScan was set to identify key components of the type three secretion system (Abby, et al., 2014). These core structural components of T3SS are SctC, SctF, SctJ, SctN, SctS, SctR, SctQ, SctV, SctU, and SctT (Deng et al., 2017). SctC is forming the outer membrane ring in the structure, while SctF is the needle part. StcV, StcU are part of the export apparatus and act respectively as the autoprotease, and the export gate. StcR, StcS, and StcT are also part of the export apparatus and they act as inner membrane component. StcQ is the T3SS ATPase which is part of the cytoplasmic ring (Deng et al., 2017). The full genome sequences were obtained from NCBI or ENSEMBL bacteria database if it was not found in the former. TXSScan was accessed through Galaxy Pasteur (https://galaxy.pasteur.fr).

2.1.3 Manual search for homology with T3SS components:

Manual search for homology with type three secretion components was performed against homolog species that could not be identified by TXSScan using BLAST software. To do the homology search, different proteins from T3SS structural component of *Shigella* were picked. These genes are MxiH, MxiG, MxiD, MxiA, Spa47, Spa33, and IpaB. The protein sequences were obtained from either the virulence plasmid of *Shigella* (pWR100) or UniProt website.

2.1.4 Pairwise sequence alignment and phylogenetic tree:

The full protein coding sequences were obtained from NCBI database and aligned using Clustal Omega version 1.2.4 (Sievers et al., 2011). After aligning all protein sequences together, the catalytic residues Histidine 145, Asparagine 195 and Cysteine 306 were highlighted (Liu et al., 2018). The software identifies fully and semi-conservation of amino acids among the provided sequences.

In addition, a phylogenetic tree of IcsB homologs was generated from Phylogeny FR https://www.phylogeny.fr using the catalytic domain of IcsB and its homologs in FASTA format. The default "One click" mode was used for tree reconstruction. The software uses different tools to perform the phylogeny analysis, such as multiple sequence analysis, curation, phylogeny and tree reconstruction. These tools are MUSCLE, Gblocks, PhyML, and TreeDyn. When providing the sequences in FASTA format, the program starts analyzing and at the end, it provides a graphical tree image with branch length or branch support values. (Dereeper et al., 2008). The branch support value indicates the repetition number out of 100 of same branch when the tree was reconstructed.

2.1.5 Type Three secretion effector protein predictions algorithm:

There are several algorithms used to predict type three secretion effector proteins. In this case, algorithms picked are pEffect, EffectiveDB, and BEAN. A control test was performed for accuracy purposes. Two control proteins were used to benchmark each algorithm. First, IpaB was used as a positive control; it is a translocator protein secreted by *Shigella*'s T3SS. The algorithm should be able to identify IpaB as T3SS substrate. Second, RecA, was used a negative control; it is a protein that plays a role in homologous recombination and DNA repair process in

E. coli and is not evolutionary related to the T3SS, and thus should not be identified as one of its substrate.

The prediction done using pEffect software (https://rostlab.org/services/peffect/) actually combines both PSI-BLAST search and *de novo* prediction using machine learning technique to predicts T3SS effectors (Goldberg, et al., 2016). Furthermore, it provides a score for each predicted query. The higher score the more likely to be a type three secretion effector.

In addition, Bacterial Effector Analyzer (BEAN) software (http://systbio.cau.edu.cn/bean/) and EffectiveDB software (https://effectors.csb.univie.ac.at) are also a machine learning based algorithm that predict effectors released by the type three secretion system. They have similar working method in which BEAN uses the secretion signals from known type three secretion effector proteins to apply it on the provided sequence and tells if the protein is predicted as type three secretion effector or if the protein is considered as non-effector (Dong, et al., 2015). Similarly, EffectiveDB based on the N- terminal signal identification of the query protein in comparison with known type III secretion effectors using machine learning tools (Eichinger et al., 2015). Furthermore, the software provides prediction score to indicate the protein possibilities of being type III secretion effector.

Chapter II Results

2.2 search results:

2.2.1 Sequence similarity searches for IcsB catalytic domain

The initial attempts to identifying homologs of IcsB in the database by Psi-BLAST using the full protein coding sequence of IcsB as query failed. Therefore, we decided to use only the catalytic domain of IcsB as query (Pei & Grishin, 2009), which immediately appeared more promising. After the first iteration of PSI-BLAST, more than two hundred hits were retrieved. Hits that showed identity between 20% and 75% were selected to perform the second round of PSI-BLAST. The reason for multiple iterations was to expand the search to obtain as many sequences as possible. The hits showed up from the second iteration were mostly proteins with sequence identity lower than 20% in addition to many MARTX toxins. The search was initiated for a third time, and MARTX toxins from *Vibrio cholerae* and related species came out as predominant new hits in this round. Interestingly, a recent study demonstrated that MARTX toxins possess fatty acylation activity of Rho GTPases proteins (Zhou et al., 2017) similarly to IcsB. Nonetheless, because MARTX toxins sequence identity with IcsB ranges between 15-19% and are not associated with the T3SS (Boardman & Fullner Satchell, 2004), we stopped the Psi-BLAST at the third iteration and discarded the MARTX toxins from our study. Despite this, our search revealed more than 200 proteins sharing homology with the catalytic domain of IcsB. Some proteins were redundant, like in the case of IcsB, there were more than 10 proteins annotated as IcsB with almost 100% sequence identity. By excluding proteins sharing homology above 75% and lower than 20% with IcsB and one another, 11 homolog proteins, including BopA, were selected for further study (table 2.1). Nonetheless, after obtaining the full sequences of the new homologs, BLAST was run for each one of the proteins against the full length IcsB

(table 2.1). Consequently, Pairwise-BLAST was used to identify regions of homology that could indicate functional similarity between different bacterial species.

For practical reasons, we renamed arbitrarily the homologous proteins H1 to H11 (presented in table 2.1). As for the sequence identity between IcsB and each homolog (table 2.1), it is noteworthy to mention that BLAST sequence identity represents the percentage of shared homology between two sequences excluding the non-shared parts. These proteins are BopA from *Burkholderia pseudomallei* (H1), and it scored 31% identity with IcsB. A hypothetical protein from *Escherichia marmotae* (H2) that scored 53% identity when queried against IcsB and it is the highest among the group. A hypothetical protein from *Chromobacterium sp. LK11* (H3) that showed 31% sequence identity with IcsB. A protein annotated as the BopA from *Achromobacter arsenitoxydans SY8* (H4), which showed 27% identity with IcsB. Putative IcsB-like invasion protein from *Sodalis praecaptivus* (H5) that scored 27% identity with IcsB. A hypothetical protein from *Achromobacter sp. Rta* (H6) that showed 26% sequence identity with IcsB. A hypothetical protein from *Endozoicomonas montiporae* (H7) that showed 25% identity with IcsB. A hypothetical protein from *Robbsia andropogonis* (H8) which scored 22% sequence identity with IcsB. A hypothetical protein from *Achromobacter denitrificans* (H9) showed 25% identity with IcsB. A hypothetical protein from *Desulfonatronospira thiodismutans* (H10) scored 28% identity with IcsB. Finally, a hypothetical protein from *Ideonella sakaiensis* (H11) that showed 24% sequence identity when queried against IcsB.

Table 2.1. Information about bacterial species harboring IcsB homologs.

Identity demonstrates the homology percentage between IcsB and each homolog protein separately. Phylum section identify the class of proteobacteria. Isolation describes the source in which the bacteria was isolated from.

Protein	Gene Id	Species or genus	Identity	Phylum	Isolation
H1	53722545	*Burkholderia pseudomallei*	31%	Beta	Soil
H2	740535360	*Escherichia marmotae*	53%	Gamma	Faecal samples of Marmota himalayana in China
H3	860401851	*Chromobacterium sp. LK11*	30%	Beta	Plant
H4	359363722	*Achromobacter arsenitoxydans*	27%	Beta	Soil in pig farm "arsenic-contaminated"
H5	573023760	*Sodalis praecaptivus*	27%	Gamma	Human wound
H6	759857519	*Achromobacter sp. Rta*	26%	Beta	Termite gut
H7	736879821	*Endozoicomonas montiporae*	25%	Gamma	Sea water of Taiwan coast
H8	640442335	*Robbsia andropogonis*	22%	Beta	Soil bacteria "plant"
H9	1209943465	*Achromobacter denitrificans*	25%	Beta	Soil and water
H10	496145519	*Desulfonatronospira thiodismutans*	28%	Delta	Sediment lake in Russia
H11	928997844	*Ideonella sakaiensis*	24%	Beta	Sediment "plastic recycling facility"

The 11 proteins found in the database seem to be interesting in the sense that their sequence identity with IcsB varies, and they originate from different species. Therefore, after obtaining the name of the bacterial species implicated in expressing the homologous proteins, I mined the databases to discover more about their source of isolation, their environment, and evaluate their pathogenicity. In addition to that, we used *in silico* method to mine their genome for evidence of functional type three secretion system components. Hence, if T3SS is present, it will add to data found in the literature.

TXSScan is a computational tool used to identify secretion system in bacteria using sequence similarity using Hidden Markov Model (HMM) protein profile. For type three secretion system, the search was done against each bacterial genome for homology with SctC, SctF, SctJ, SctN, SctS, StcR, StcQ, StcV, StcU, and StcT genes. These genes are from different parts of type three secretion core structure. SctC is part of the basal body and function as secretin and outer membrane ring and SctF is the needle. StcV, StcU, StcR, StcS, and StcT are in the export apparatus and act respectively as the autoportease, the second as an export gate, and inner membrane component for the latter three. StcQ is part of the cytoplasmic ring and is the T3SS ATPase (Deng et al., 2017). The exciting feature of TXSScan is that it searches for all these gene at once and it gives a full report when the search ends. All protein sequences of the full genome were obtained from NCBI database or in case the full genome sequence was not found or incomplete, genome sequences from ENSEMBL bacteria database was used (table 2.2) and the results are shown in (table 2.3).

Furthermore, NCBI database was used to look for the taxonomy for each bacterium. Species expressing homologs H1, H4, H6, H8, H9, and H11 are betaproteobacteria belonging to the order *Burkholderiales*. Specie expressing H3 is also a betaproteobacteria, but it belongs to the

order *Neisseriales*. Species expressing homologs H2, and H5 are gammaproteobacteria belonging to the order *Enterobacterales*. Specie expressing H7 is gammaproteobacteria, but it belongs to the order *Oceanospirillales*. Finally, the specie expressing H10 is a deltaproteobacteria belonging to the order *Desulfovibrionales*.

Nevertheless, many of these bacterial species are not well-studied. A search concerning evidences about the possibilities of being capable to infect or act as commensal bacteria. Search revealed that *Burkholderia pseudomallei* (H1) was isolated from soil, and it is responsible of causing melioidosis, fatal in some cases, in human and animals (Gong et al., 2011). *B. pseudomallei* was known previously to express a protein named BopA and known to be IcsB homolog. When the *B. pseudomallei* is inside the host cell, it uses BopA protein to escape from autophagy and spread to the neighboring cells in a similar manner to IcsB (Cullinane et al., 2008). *B. pseudomallei* is the only bacteria that have been studied the most in the group of homologs. Also, using chromosomal protein sequence, the software was able to predict three clusters for type three secretion and score nine out of nine for the presence of mandatory genes with no forbidden genes identified.

Achromobacter arsenitoxydans SY8 (H4) was isolated from arsenic-contaminated soil in a pig farm (Li et al., 2012). Thirty genes of the type three secretion system were found in the genome. This may indicate the possibility that this specie may be an opportunistic pathogen of pigs and maybe other animals (Li, et al., 2013). Unlike other *Achromobacter* species, *A. arsenitoxydans* was never reported in clinical isolates from patients with respiratory disease. Besides, the full genome sequence was not found in the NCBI database. The shotgun sequences found in NCBI were not complete and not enough for the software to be able to detect any T3SS

homology. Furthermore, A genome sequence obtained from ENSEMBL bacteria database used to search for homology using TXSScan. One cluster of T3SS was found with nine mandatory genes identified and no forbidden genes found.

Achromobacter sp. Rta (H6) was isolated from Termite gut according to the data found in NCBI database (gi| 687281538). The search about *A. sp. Rta* revealed no studies nor information to be found. In addition, the full genome sequence was not found in NCBI database. Instead, a shotgun sequence was used for the TXSScan search. One cluster for type III secretion was identified having nine of the mandatory genes and no forbidden genes identified.

Robbsia andropogonis (H8), which was previously named *Burkholderia andropogonis* (Lopes-Santos et al., 2017), was isolated from plant soil. It was found to infect jojoba plants causing necrotic spots on leaves (Cother, et al., 2004). Another case reported on *R. andropogonis* was on *Loropetalum chinense* plant. The infected shrub found to show symptoms as dark spots on leaf immersed with water. When the infection spread, the infected leaf become heavy and fall off (Conner, et al., 2019). However, information on how does *R. andropogonis* cause the disease has not been found. Besides, the full genome sequence was not found in NCBI database. Instead, a shotgun sequence was used. One cluster of T3SS was identified having seven of the nine mandatory genes. None of the forbidden genes were found in the cluster.

Achromobacter denitrificans (H9) was originally isolated from soil and water. Although the genus of *Achromobacter* have many species considered as non-pathogenic, *A. denitrificans* has been clinically isolated from patients having pneumonia, cystic fibrosis and some other non-

respiratory diseases. The paper suggested that these isolates of *A. denitrificans* might be associated to cause more diseases in patients who have other respiratory and non-respiratory illnesses (Swenson & Sadikot, 2015). Also, the full genome sequence was not found in NCBI database. Instead, a shotgun sequence was used. One cluster of T3SS was identified with eight of the mandatory genes found. Also, none of the forbidden genes were identified.

Desulfonatronospira thiodismutans (H10) was isolated from Sediment of a soda lake in Russia. These soda lakes are characterized by a high level of carbonate alkalinity along with high sulfate concentration, which makes it an appropriate habitat for strains like *D. thiodismutans* that utilizes organic and inorganic compounds for metabolic processes. The optimum growth condition for this strain is at pH 10 and 1.0- 2.0 M Na^+ (Sorokin et al., 2008). This strain of bacteria was found to be able to reduce the sulfur compounds found in such lakes (Sorokin et al., 2011). Nonetheless, there was no evidence for *D. thiodismutans* of becoming a pathogenic bacterium and its requirement for growth makes it unlikely.

Chromobacterium sp. LK11 (H3) was isolated from plant according to the information found in NCBI database (gi| 857313621). There was not much information found about *C.* sp. LK11. The fact that it has not been attributed a systematic species name is a testament to this observation. However, another specie from *Chromobacterium* genus dubbed *Chromobacterium violaceum* was found to be pathogenic and possess a type three secretion system. It is known to cause a fatal infection called septicaemia that infects both humans and animals (Miki et al., 2010). Furthermore, *C. violaceum* is known to be resistant to a wide range of antibiotic drugs which makes the infection difficult to treat (Fantinatti-Garboggini et al., 2004). We do not know the level of similarities between *C. violaceum* and *C.* sp. LK11.

Escherichia marmotae (H2) was isolated from faecal samples of *Marmota himalayana* in China. It was recently proven that *E. marmotae* shared homology with type three secretion system and its effectors. It also carries the identical O antigen gene cluster found in *Shigella dysenteriae* type 8. The strain also shares some virulence genes found in pathogenic *E. coli*. Additionally, the strain showed the capability of becoming invasive and it was able to invade human epithelial cells-2 (HEp-2) (Liu et al., 2019).

Sodalis praecaptivus (H5) was isolated from human wound of a man who pierced himself while manipulating tree branches (Yong, 2017). Genus of *Sodalis* are known to share symbiotic relationships with insects, such as weevils and hemipteran (Santos-Garcia, et al., 2017). It releases insecticidal toxins to mediate host infection. It was also proven that *S. praecaptivus* uses quorum sensing system to maintain and persistent infection period in weevil (Enomoto, et Al., 2017). Furthermore, another *Sodalis* specie dubbed *Sodalis glossinidius* was found to act as a commensal in the vector of sleeping sickness, the tsetse fly (Maltz, et al., 2012). It was found to use type three secretion components to help in the invasion process (Dale et al., 2001).

Endozoicomonas montiporae (H7) was isolated from sea water in Taiwan. Genus of *Endozoicomonas* is normally found in corals, and also it acts as endosymbiont. According to Ding, more than 200 proteins were expressed by *E. montiporae* are predicted to be type three secretion effectors. It is also suggested that the bacteria use type three secretion system for their interaction with their hosts, in addition to delivering effector proteins to modulate infection (Ding, et al., 2016).

Ideonella sakaiensis (H11) was isolated from sediment from plastic recycling facility. The strain was found to produce enzymes to degrade Polyethylene terephthalate (PET), the polymer used in making plastics, into smaller compartments and use it as a carbon source for

growth. (Yoshida et al., 2016). Also, the full genome was not found in NCBI database. A shotgun sequence was used. As for the results obtained from TXSScan, no type III secretion clusters were found. In addition, a full genome sequence was found in ENSEMBL bacteria database, and it used to search for the presence of any of the T3SS, T4SS, and T6SS. The results show that it does not have Type III nor type VI secretion systems. As for type IV search, the results showed two out of three of the mandatory genes, as well as 11 out of 20 accessory genes were found in *I. sakaiensis*.

For *Chromobacterium sp. LK11* (H3), the whole genome sequence was not found. Instead, a shotgun sequence was used. One cluster of T3SS was found and nine of the mandatory genes are present with no forbidden genes identified.

For *E. marmotae* (H2), using the plasmid pEM148 sequence, one cluster was identified for type three secretion system and scored nine out of nine for mandatory gene presence with no forbidden gene identified.

For *S. praecaptivus* (H5), using the chromosomal protein sequence, the software identified three clusters of type III secretion system. Two of these three clusters have nine of the mandatory genes, and the other cluster scored eight out of the nine mandatory genes found with no forbidden genes identified in all cases.

For *E. montiporae* (H7), a protein chromosomal sequence used. The result revealed one cluster of the type III secretion system identified having 8 of the 9 mandatory genes. None of the forbidden genes were identified.

For *D. thiodismutans* (H10), a full genome sequence was obtained from ENSEMBL bacteria database and used to search for homology using TXSScan. None of the genes of type III,

IV, VI were found. Moreover, when looking for type IX secretion system, one out of eight mandatory genes was found.

Table 2.2. Reference sequence used for TXSScan.

Protein	Species or genus	Sequence type	NCBI Reference Sequence	Ensembl Genome assembly number
H1	*Burkholderia pseudomallei*	Chromosome	NC_006351.1*	BP_3921g
H2	*Escherichia marmotae*	Plasmid	NZ_CP025980.1*	NA
H3	*Chromobacterium sp. LK11*	Shotgun	NZ_LDUR01000009.1*	ASM104370v1
H4	*Achromobacter arsenitoxydans SY8*	Chromosome	NZ_AGUF01000055.1	ASM23678v2*
H5	*Sodalis praecaptivus*	Chromosome	NZ_CP006569.1*	ASM51742v1
H6	*Achromobacter sp. Rta*	Shotgun	NZ_JPYO01000036.1*	ACRTA
H7	*Endozoicomonas montiporae*	Chromosome	NZ_CP013251.1*	ASM72256v1
H8	*Robbsia andropogonis*	Shotgun	NZ_LAQU01000001.1*	ASM97034v1
H9	*Achromobacter denitrificans*	Shotgun	NZ_CP020917.1*	NA
H10	*Desulfonatronospira thiodismutans*	Chromosome	NZ_ACJN02000001.1	ASM17443v1*
H11	*Ideonella sakaiensis*	Chromosome	NZ_BBYR01000069.1	ASM129352v1*

* This symbol indicates the sequence used for the prediction method. Generally, we used NCBI, unless it yielded a blank result. The latter stemming probably from annotation issue was resolved by using the data from ENSEMBL bacteria.

Table 2.3. TXSScan search results.

Reference numbers used for these predictions are listed in the previous table (2.2)

Protein	Number of T3SS	Validation of homology number of mandatory and forbidden genes	
H1	3	9/9	0/3
H2	1	9/9	0/3
H3	1	8/9	0/3
H4	1	8/9	0/3
H5	3	9/9	0/3
H6	1	9/9	0/3
H7	1	8/9	0/3
H8	1	7/9	0/3
H9	1	8/9	0/3
H10	-	0/9	0/3
H11	-	0/9	0/3

Although the search with TXSScan successfully identified most of the core components of the T3SS in most bacterial species of interest, it was not able identify similar features within the genome of bacteria expressing the proteins H10 and H11. There was some evidence of the presence of a T4SS in the specie expression H10. Consequently, we decided to perform a manual search for similarities with T3SS components. These genes are MxiH, MxiG, MxiD, MxiA, Spa47, Spa33, and IpaB. From the export apparatus MxiA, from the basal bodies' components MxiD, MxiH, and MxiG, from the cytoplasmic compartment Spa33, from ATPase complex components Spa47, and from translocator proteins IpaB (Deng et al., 2017). Therefore, results of homology identity and the coverage percentages were collected in (table 2.4).

When queried against *D. thiodismutans* (H10) genome, the results showed homology of 34% and 97% coverage with MxiA, homology of 19% and 52% coverage with MxiD, homology of 41% and 98% coverage with Spa47, homology of 41% and 26% coverage with MxiH, homology of 26% and 12% coverage with IpaB, and no homology with Spa33 and MxiG proteins found in *D. thiodismutans.* In addition, the BLAST search against *I. sakaiensis* (H11) genome, the results showed homology of 32% and 97% coverage with MxiA, homology of 47% and 90% coverage with MxiD, homology of 39% and 92% coverage with Spa47, and no homology with Spa33, MxiH, MxiG, and IpaB. Nevertheless, the results of the manual search should be interepreted with care. First, the coverage between the query and the subject was sometimes low. Secondly, the homology with the flagellum genes was not addressed. Finally, H10 and H11 are the only homologs devoid of an N-terminal domain, which normally includes the secretion signal and chaperon binding region of T3SS effectors. Taken together, these data and observations suggest that H10 and H11 are not T3SS effectors and thus are the exception across the IcsB-like family that I unveiled.

Table 2.4. Manual search for T3SS components homology in the bacterial species of interest.

Bacteria harboring homologs H10-11 that scored negative with TXSScan are shaded grey. Bacteria harboring homologs H1-9, for which T3SS was detected by TXSScan are appended as internal controls.

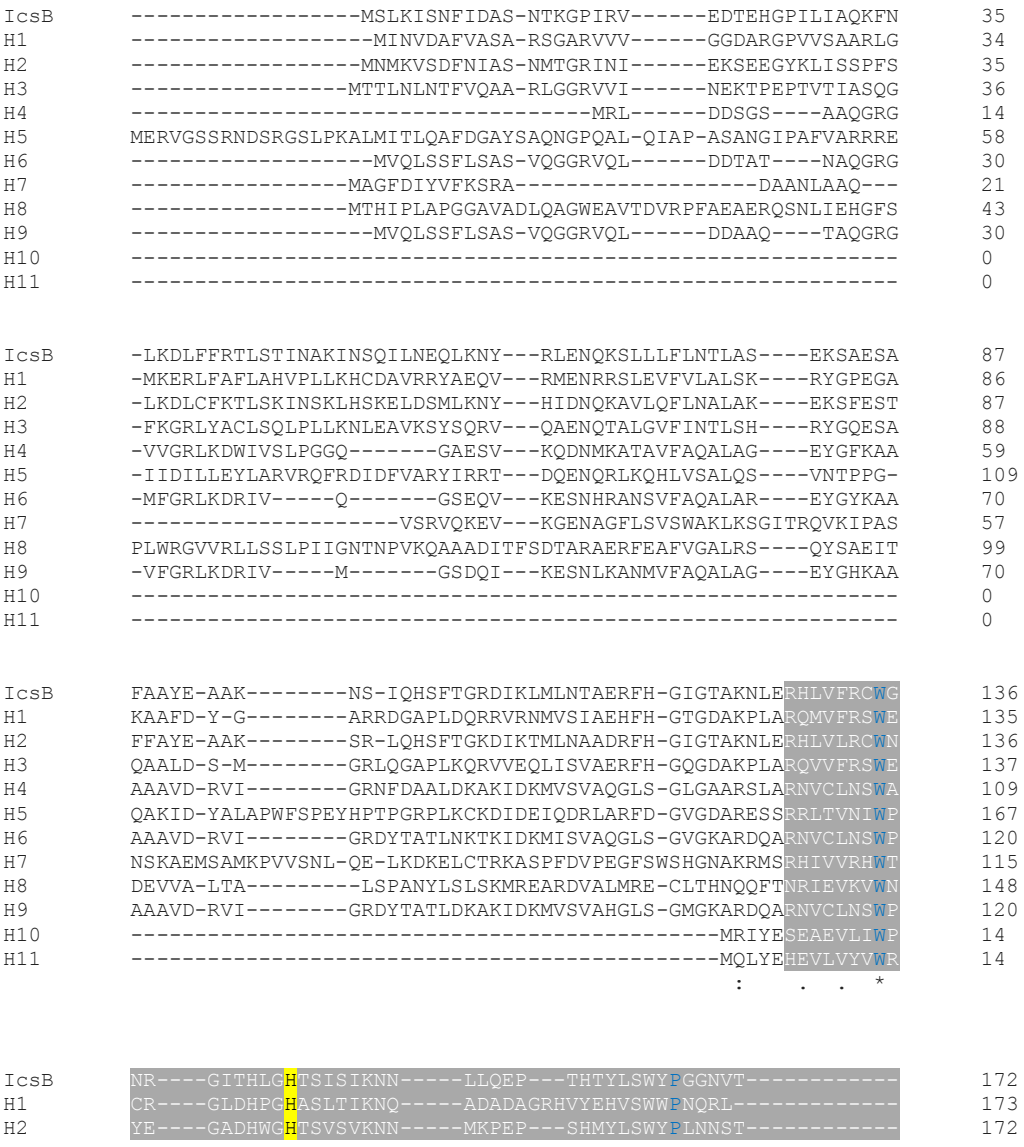
		Identity / Query coverage						
Homolog	**Bacteria Specie**	**MxiA**	**MxiD**	**Spa33**	**Spa47**	**MxiH**	**MxiG**	**IpaB**
H1	***Burkholderia pseudomallei***	55/100	38/87	NA	57/65	51/92	24/52	35/65
H2	***Escherichia marmotae***	85/100	74/99	NA	38/92	83/100	54/99	57/100
H3	***Chromobacterium sp. LK11***	66/97	52/86	25/56	58/99	69/100	23/52	38/97
H4	***Achromobacter arsenitoxydans SY8***	41/98	28/85	32/23	45/92	42/22	NA	NA
H5	***Sodalis praecaptivus***	57/100	47/90	15/24	54/96	58/91	27/52	36/55
H6	***Achromobacter sp. Rta***	41/98	30/79	34/25	45/92	NA	NA	NA
H7	***Endozoicomonas montiporae***	41/99	29/84	27/46	44/96	NA	NA	NA
H8	***Robbsia andropogonis***	46/99	36/96	NA	44/95	44/93	24/56	32/55
H9	***Achromobacter denitrificans***	41/98	28/85	34/24	45/92	56/19	NA	NA
H10	***Desulfonatronospira thiodismutans***	34/97	19/52	NA	41/98	41/26	NA	26/12
H11	***Ideonella sakaiensis***	32/97	22/54	NA	39/92	NA	NA	NA

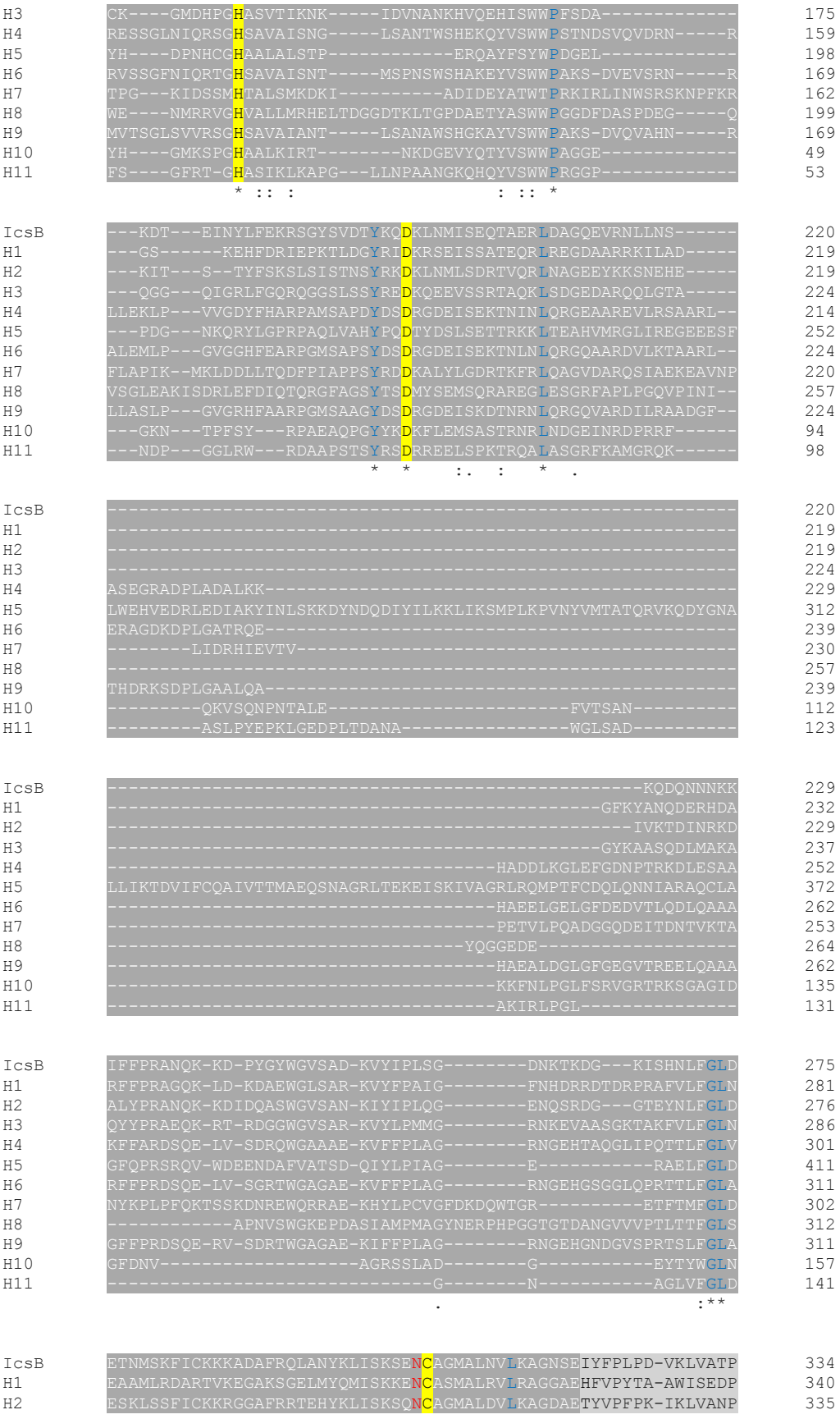
2.2.2 Sequence alignment construction:

The pairwise sequence identity concerning the catalytic domain of IcsB and the homologous proteins was ranging from 22 to 53%. Therefore, these homologous proteins were picked for study, and we had to look further into their amino acids' conservation among their sequences. Subsequently, the full protein sequences of the 11 homologs were submitted through Clustal Omega for the sequences to be aligned. The result of sequence alignment is shown in (figure 2.1). Interestingly, the alignment revealed that the catalytic triad, which are histidine 145, aspartic acid 195, and cysteine 306, were conserved among all homologs. The catalytic domain is shaded in grey, and the catalytic residues are shaded in yellow. Two other polar residues, namely N305 and N352, were also fully conserved across the alignment and are highlighted in red. There were also non-polar amino acids found to be fully conserved among all homologs and are highlighted in blue. These amino acids are tryptophan 135, proline 167, tyrosine 192, leucine 208, glycine 273, leucine 274, and leucine 314. There were also semi conservations found in the catalytic domain represented in lysine 124, serine 147, isoleucine 148, isoleucine 150, tyrosine 162, serine 164, tryptophan 165, isoleucine 200, threonine 204, phenylalanine 272, alanine 307, methionine 309, serine 319, tyrosine 322, alanine 341, valine 344, isoleucine 348, and isoleucine 358.

Figure 2.1. Multiple sequence alignments of IcsB catalytic domain homologs.

The catalytic residues are shaded in yellow. (*) indicates full conservation and (:) indicates semi-conservation. Proteins are identified by their given name abbreviation. IcsB is from *Shigella flexneri*; H1, *Burkholderia pseudomallei*; H2, *Escherichia marmotae*; H3, *Chromobacterium sp. LK11*; H4, *Achromobacter arsenitoxydans*; H5, *Sodalis praecaptivus*; H6, *Achromobacter sp. Rta*; H7, *Endozoicomonas montiporae*; H8, *Robbsia andropogonis*; H9, *Achromobacter denitrificans*; H10, *Desulfonatronospira thiodismutans*; H11, *Ideonella sakaiensis*.





```
H3        EKAILHEARHVKQEAAEGRIGYTLASTEPNCASMALRMLRAGGAENFVPFSA-SWVTEDP      345
H4        EHDMLADARQLKDDAAHGRIGYSMFSTTQNCAAIAARSLQAGGSNIYVPFEA-SWITEDP      360
H5        EQALEEFWLSVKSEIATGKARFRLVSTRHNCSAMALRSLQAAGAGRYVPLPA-RWLLYTP      470
H6        EHDMLADANQLKADAEQGLIGYSKLSTTQNCAAVAARSLQAGGSNIYVPFEA-AWITEDP      370
H7        LEKMRNKWIAVKNP-EHPNHYYKQFSTEQNCSGMCLSLLKEGGAGLFYNFSP--SLVTTQ      359
H8        VDAIKTYWETEA--IGNDAAQFKLVSTDRNCSGMAARLLKAGGAHAFLPVPE-AVLFLDP      369
H9        EHDMLADANQLKSDAAQGRIGYSMFSTTQNCAAVAARSLQAGGSNIYVPFEA-SWISEDP      370
H10       LPGMVKWWMDFSD----TPKGYSFVSTKYNCAAVVAQCLLEGCADAYIDIKQ-PKMYYTP      212
H11       IGAMVRWWNLFTH----SG-SNEFNILKLNCSRVAAMCLLAGGAARYCKPPTRAFNLWTP      196
                 :                        **: :    *  . :  :

IcsB      NDVYAYANKVRQRIESLNQSYNEIMKYIESDFDLSR------------------------      370
H1        NHAHAYALAVQARIDALNQRRADVERRCERLRDSASVRQA-------------------W      381
H2        NDAYKYAYKVMARIEELNKIHKGIIKNVDEKIDIMT------------------------      371
H3        NKAHAYAERLQARMDSLNRQVADIGERCSGMLQQPGVKTE-------------------W      386
H4        NKTYDYARQLQGAIDGLNGKADGIKKFCDTAWQSAPADAD-------------------A      401
H5        RDVEHYTLRLQKALDRLNRQSSAIQAFYEQE---------------TASRPNDLAIRQSD      515
H6        NKTYAYALQLQAAIDQLNDQAGSVKAHCAAAWSKAPAEAD-------------------A      411
H7        SDVEKYSAKLVDKLDRLNKHVDDLDEKIRLYKVPNEPELP------LSSIPEK-------      406
H8        NTMRRYTTDLMTEVTALNAKADTVDASFGRLFTEDEGADE---SGVTPSYSVRLARAQEA      426
H9        NKTYDYARELQGAIDLLNEQEGSVKAHCASAWSKAPAEAD-------------------T      411
H10       NDISALADKLVNRLDDLNIESRRLINNSRQSLDRPG--IWTSEEFYKASYAGRFARRKGQ      270
H11       NSIEAWALKLDSALKARNEAAKEIELERLPSNDQPPSYLLDVGTWQKATKPGSFSVRRGQ      256
                :  :   :   *     :

IcsB      LTQLRRSYLKSFNKINLIH-TPKTFKP----------LS--ISLYKHPTENVS--SEDFD      415
H1        R-----AFSEAGGASASPL------AE---DAGRGR---ASAHMRQARLDEHAREV---E      421
H2        MIKLRRSYLNSFNKVTSSH-KINKFNT----------LL--VIRNKDNIENAS--LDDYN      416
H3        D-----AFLNQGRMGELKR-EISQLKQQINKERRPE---AQAVQRQALKERNAQQV---F      434
H4        IRQLRDALHETANDQAIFA-QAKQLALGIAADQRIDSLRAEVADRSGKMDQARSLLQPLD      460
H5        LTGLRTQ----FNLLFSHL-QTPQQKR-LRRLARAVNTLPDA-----------FLSQQSD      558
H6        ISILRATLLVRADDLDAYD-QTRQQAQGVAADQRILGLSADVAHAKGRVQEMEQELAGLE      470
H7        ----LVFLLSTYSMDESWKAKIQE-----------------VASIIHEIENAPSTLKGLT      445
H8        IASLPRATQASMQALNNAL------------------------------NECKTEAGDYE      456
H9        IAQLRVALHARAHDLDAYD-QTRQQAQGTAADQRIRGLGADVGHARDNLQKAERELKGLE      470
H10       VASIDDLL-KEYHKAGPWD-ENNFPQK-FKALGRMMRNVLDH-----------RQKK--A      314
H11       VLKIDKLL-HEYGVVAGRTDAEGQREA-LNVLEQILGQVHSH-----------IVQKG-D      302

IcsB      AVIN-------------------ACHSYLVKSAPSNM--TRVL--N--ELKTEATDKKEE      450
H1        RIGAYFAELSAGRSGKHRDRADAALADAMKRCAPS------------------ARDDVAA      463
H2        KIIN-------------------SCFESLKKLAPHNS--HSRI--S--EFKPTMTDTKEQ      451
H3        RIGAHFEQLAVGRAD------ALALAKAMREGAPS------------------IDDDAAV      470
H4        RQIRATEKDVGDRQKAVD-NKQKAVDGMSRSAKSSSP--HAAA--RLADSKAELAERQDE      515
H5        RIIPAAKALVE-------------------ALSPLLT--DAIS--QQQDYRT--------      587
H6        TKIVATQKEVNGRSEKVR-LAADSLAARAARLQPDDA--KAAA--KLNADRKSHQKDLAA      525
H7        PIAIRLTTSLD-RLFKIT----SDNPYLTDRLEPALHAFKILKNRMEDAYREQV------      494
H8        GMMRQMKAMVGGLHNHVNGPFGSDINKALRSAYELYSAIRGAA--HEAGYLREA------      508
H9        NTIPAMEKEVAARSDRIN-RTAASLEARAENLNPENT--KEVV--KLKADRDSHDKNVNA      525
H10       --GSDRGGAVE-------------------TLGLQIL--AE--------LSS---------      335
H11       --GSRNREVVL-------------------QLGQHVL--KL--------YRE---------      323

IcsB      IIEK----------------SI------------------------------KIIDYYNS      464
H1        ------------------------------------------------------------      463
H2        LIEK----------------SI------------------------------KIIDFYQS      465
H3        ------------------------------------------------------------      470
H4        LKKGQAELNDLRGQRSRLDTRILAMKDERRALKDLIEERQANPLEAARRKLGGEDDYYAK      575
H5        ------TLSLAAA-------------------MRTWLQQRQQRGV--------PSS-ALYSA      615
H6        LELRRGELATQLALKPRLEQRILALDEDHEALSDLLEDRKAFPVEAAQLRPAGEEGYYAA      585
H7        ----------------------------------EMF-----------------EDPYFE-      503
H8        ----------------SDSDI---------PSDFFDN----------------------      520
H9        LNRRREELSAQVGRKARLEQRIPALRDEHEALSGLLNDRTAFPVEAAQLKLPGEEGYYAA      585
H10       ------E----EL---------------------LIAQNN---N-------AGG-SLAAL      353
H11       ------R----SV-------------------RLALLKQSQAKNL-------VGA-SKYQL      347


IcsB      LKSPDLGTKLYI-----------------------------------------HDLLQINKL      485
H1        ------------------------------------------------------LTRKASV      470
H2        LKNKNQELLFYT-----------------------------------------HDLLLVNKT      486
```

```
H3        ----------------------------------------------------LTRKTAG       477
H4        LRGGLAEGVTGLAHAFGARASGLDTQQQRLLAPLGQAVQDLERGVRNGSDMKTLLSHAKQ       635
H5        AS---ATSQRDHH-----------------------------------------------       625
H6        QRRGLQSCVEGLADQFDAHIQGLDARQQRQLAPLGRAVSELRDAVNAGGGMQALLSKAKP       645
H7        ----------------------------------------------------------       503
H8        ----------------------------------------------------------       520
H9        QRRGLQSCVDGLAQQFDARLQGLDARQQRQLAPLSGAVADLRAAVDAGAGMQALMSKAKP       645
H10       LS----LCEEGNN-------RT-------------------------------------       364
H11       LDEYNALMAEGLK-------LG-------------------------------------       362


IcsB      LLNNSHSNI-----------------------------------   494
H1        LVETLGRHLDAPPPSDSSALRRLAAHAMIGRIEAFMAAAIAA--   512
H2        LLDKSR--------------------------------------   492
H3        LVEALHHYQESHPQG--AEQLSLAAHAMIKRSEELMGLAIQ---   516
H4        LVQTLHDAINGTPAPDRVAA--LAAGALVQACEFLVVNDQDRHS   677
H5        ---RAHYARKY-AVPDSPER-----------------------   641
H6        LVETLHALTSGAPAADQIAI--LAAGSLVEACEILITLDQDRRA   687
H7        ----------------------------------------------   503
H8        ----------------------------------------------   520
H9        LVETLHALKDAAPAADLVAM--LAAGTLVEACEVLVTLDQDQHA   687
H10       LDQSFKKAKKF---------------------------------   375
H11       -TESWGQ-------------------------------------   368
```

Color codes:

Catalytic domain

Catalytic residues

Fully conserved residues (Polar and Non-polar aa)

Conservation beyond the catalytic domain

Furthermore, the full protein sequence is divided into three major domains. The N-terminal domain, which happened to be found in 9 out of the 11 homologs and IcsB, might act as a secretion signal and chaperon binding domain. The central domain, which is the catalytic domain, is originally mapped to residues 128 to 319 of IcsB (Pei & Grishin, 2009). The conservation level observed in the catalytic domain actually extends to a stretch of 41 residues (tyrosine 361) located to its C-terminus. Finally, the carboxy-terminal domain of unknown function and undetectable conservation between IcsB and its homologs.

A qualitative representation of the homologs sequence similarities is shown in dendrogram (figure 2.2) along with (table 2.5) which is representing the pairwise sequence identity percentage of the homologs' catalytic domain with one another. Some homologous proteins showed higher identity percentage with one another than with IcsB. For instance, H4, H6, and H9 showed sequence identities higher than 65% (ranged between 69% and 74%), while their sequences identities with IcsB were lower than 30%. Accordingly, these three proteins H4, H6 and H9 were clustered together in the phylogenetic tree.

Figure 2.2. Dendrogram of IcsB and its homologs.

The figure was generated based on the catalytic domain using Phylogeny FR software. Numbers represent branch length/ branch support values (in red). *Numbers were rounded to two decimal points.

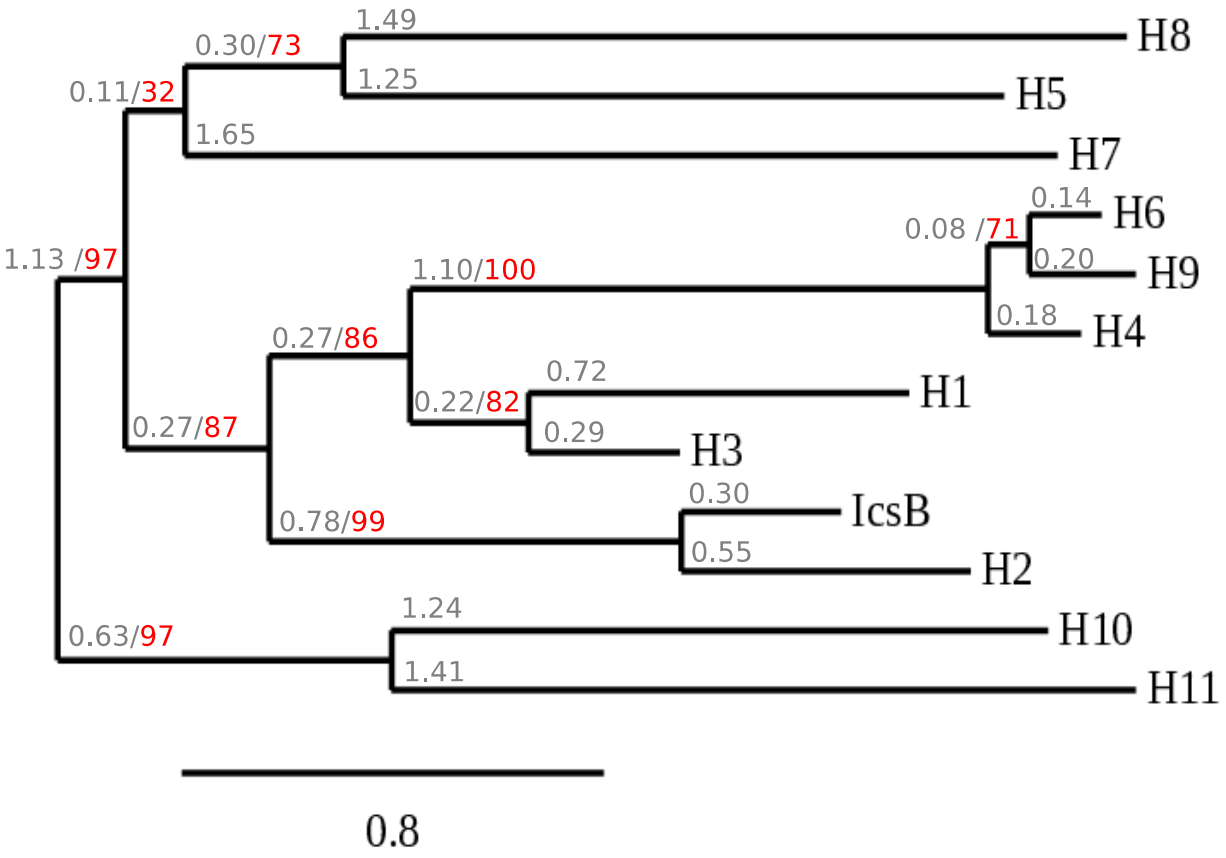


Table 2.5. Sequence identity of IcsB homologs with one another.

Number XX/YY means that there were two regions with homology separated by an insertion with undetectable sequence identity.

	IcsB	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11
IcsB		37	54	35	28	24/26	28	24	21	26	27	25
H1	37		35	50	32	27/32	30	28	23	32	28	29
H2	54	35		34	24	25	24	28	22	24	25	22
H3	35	50	34		32	23/34	33	25	24	34	26	27
H4	28	32	24	32		27/33	71	22	22	69	26	33
H5	24/26	27/32	25	23/34	27/33		22/31	25	23/46	25/36	28/35	29/24
H6	28	30	24	33	71	22/31		22	30	74	26	26/28
H7	24	28	28	25	22	25	22		42	27	NS	30
H8	21	23	22	24	22	23/46	30	42		30	23	28
H9	26	32	24	34	69	25/36	74	27	30		27	32
H10	27	28	25	26	26	28/35	26	NS	23	27		28
H11	25	29	22	27	33	29/24	26/28	30	28	32	28	

2.2.3 Type Three secretion effector protein predictions algorithm results:

Another *in silico* method used to understand the nature of these homologous proteins is the algorithm prediction tool. Three algorithms used to predict the possibilities of the homolog proteins to be type three secretion effectors are pEffect, EffectiveDB, and BEAN. They search for similarities between query sequences and known type III secretion effectors. However, RecA, IcsB and IpaB are used as positive and negative controls to benchmark the algorithms. All three algorithms were successfully able to recognize IcsB and IpaB, but not RecA, as type three secretion substrates. Therefore, a summary of the results is presented (table 2.6).

Furthermore, pEffect, EffectiveDB, and BEAN were able to predict H2, H6, H9 as type three secretion substrates. Additionally, pEffect and BEAN were able to predict H1, H3, H4, H5, H7, H8, H10, and H11 as type III secretion substrates but they were not identified by EffectiveDB tool.

As discussed above, the fact that H10 and H11 lacks the secretion signal and chaperone binding domain decreases the likelihood that these two proteins are T3SS substrate, unlike the rest of proteins in the group. Furthermore, the TXSScan could not identify any homology with the T3SS in these two bacteria. By extension, the eventuality that the bacteria encoding H10 and H11 adopt a pathogenic lifestyle is low.

Table 2.6. Type three secretion effector prediction of IcsB homologs using algorithms.

Shaded in grey: IpaB, which is a translocator protein secreted by *Shigella*'s T3SS used as positive control. RecA used as negative control.

Protein	pEffect	EffectiveDB	BEAN
IcsB	Yes	Yes	Yes
IpaB	Yes	Yes	Yes
RecA	No	No	No
H1	Yes	No	Yes
H2	Yes	Yes	Yes
H3	Yes	No	Yes
H4	Yes	No	Yes
H5	Yes	No	Yes
H6	Yes	Yes	Yes
H7	Yes	No	Yes
H8	Yes	No	Yes
H9	Yes	Yes	Yes
H10	Yes	No	Yes
H11	Yes	No	Yes

2.3 Conclusion:

IcsB is an effector protein that is released by the type three secretion system of *Shigella flexneri*, which was found to help the bacteria to escape the host autophagy and spreading to neighboring cells. Moreover, IcsB was shown to possess acyltransferase activity to disrupt the actin cytoskeleton of the host cells, which is critical for the invasion and escaping processes. Rho-GTPases were found to be targeted by IcsB, as well as CHMP5 that is implicated in ESCRT-III complex (Liu et al., 2018). Using bioinformatic approach, the catalytic domain of IcsB was used to initiate PSI-BLAST search, and it resulted in variety of homolog proteins expressed from different bacteria species. Some of these bacteria were proven to have Type three secretion system. Others, where shown to possess opportunistic or commensalism activities.

However, when the sequences of the homologs were aligned using Clustal Omega, the catalytic triad, which are His-145, Asp-195, and Cys-306 (Liu et al., 2018), were found to be fully conserved among all homologs. In addition, two other amino acids Asn-305 and Asn-352 were fully conserved. The homology in the catalytic domain was found to be extended beyond amino acid 319, which revealed the relation between IcsB and the homologs at the structural and functional level, suggesting that these previously unstudied proteins might be endowed with IcsB-like acyltransferase activity.

Chapter III Material and method

3.1 Plasmid generation and cloning

3.1.1 Construction of the yeast expression plasmids:

pRS313 and pRS316 yeast shuttle vectors (Genbank number: U03439, and U03442.1) (Sikorski & Hieter, 1989) were obtained from Dr. Cammie Lesser (Slagowski et al., 2008). Both plasmids possess features including the origins of replication for bacteria and yeast, LacZ operon, a multicloning site, T7 promoter and T3 promoter universal primers, bacterial antibiotic resistance gene and yeast auxotrophic marker for selection purposes. The major difference between pRS313 and pRS316 is the yeast auxotrophic marker, where it is HIS3 in pRS313 and URA3 in pRS316. Vectors were used for cloning and gene expression. Plasmids are presented in (figure 3.1) and (appendix table A1).

First, GAL1p DNA insert was gene synthesized (IDT). It included the GAL1 promoter that is induced by galactose, the cycE terminator, a Myc- tag, sequence corresponding to M13 forward and M13 reverse primers, and a short multi cloning site. Subsequently, GAL1p was sup-cloned into pRS313 plasmid. The sub-cloning of galactose promoter was done using XhoI and XbaI restriction enzymes. Second, GAL1p1 was generated by the addition of SP6 and BGH reverse primers sequences. Then, it was sub-cloned into pRS313 and pRS316 plasmids using XhoI and XbaI restriction enzymes. Third, GAL1p2 was generated by replacing the sequence of Myc-tag to 3X Flag-tag. Then, it was sub-cloned into pRS313 and pRS316 plasmids using XhoI and XbaI restriction enzymes. See (appendix table A2) for antibodies reference number.

Forth, IpgA, the chaperone of IcsB, was cloned from IpgA-TEM1 taken from HMI stocks. The PCR cloning for IpgA was done using primers found in the (appendix table A3). Then, it was sup-cloned into pRS316 GAL1p1. The sup-cloning was done using BamHI and EcoRI restriction enzymes.

Fifth, the coding sequence of IcsB was cloned from pGBKT8-IcsB taken from from the laboratory stock using primers found in the (appendix Table A3). IcsB homologs protein sequences, H1 (gi| 53722545), H2 (gi| 740535360), H3 (gi| 860401851), H4 (gi| 359363722), H5 (gi| 573023760), H6 (gi| 759857519), H7 (gi| 736879821), H8 (gi| 640442335), H9 (gi| 1209943465), H10 (gi| 496145519), and H11 (gi| 928997844), were obtained from NCBI database. Subsequently, homologs were synthesized using the gene synthesis service from Bio Basic except for H1 (BopA), which was previously provided by Biomatik. EcoRI and BamHI restriction sites were added to the 5' end and 3' end of the synthesized fragments to facilitate the cloning process. Homologs first came cloned in pUC57, a 2710 bp length vector that is derived from pUC19 used mainly for cloning purposes. The synthesized proteins arrived in a dry form with a total of 4 μg DNA and resuspended using 10 mM TRIS HCL pH fixed to 8.5. Plasmids were transformed in DH10Beta *E. coli* strain and glycerol stock was made. After that, one colony was picked for each homolog, and plasmids purification was performed prior to the cloning step. Afterwards, IcsB and its homologous proteins were sup-cloned into pRS313 GAL1p2 using BamHI and EcoRI restriction enzymes. Both genes and plasmids were digested, run on gel, band cut and purified, and then transformed using *E. coli*.

To do the sub-cloning, 1 μg of DNA was digested using 2 μL of 10X Fast Digest buffer, 1 μL of each of the restriction enzymes, and complete the volume to 20 μL with distilled water.

Mixture was mixed and incubated at 37°C for 10 minutes. Followed by loading digested DNA into 1% agarose gel wells and migrated for 20 minutes. After that, gels were visualized using ultraviolet transilluminator, bands correspond to the right molecular weight were cut and placed into clean microfuge tubes. Gels were purified using PCR and agarose gel purification kit (Bio Basic). The kit comes with all buffers, solutions, and columns needed in the procedure. After the purification step, ligation was performed using the ratio of 2:1 insert: vector for most cases. Then, the ligated plasmid was transformed in either DH10Beta or DH5alpha *E. coli* strain. Colonies were screened using colony screening PCR method and particular primers. The products of PCR were run on electrophoresis gel to check for their size and accuracy. After that, liquid culture was prepared with appropriate antibiotic for positive clones and incubated overnight at 37°C for DNA purification as a next step. After the extraction was done, DNA was then sent to be sequenced at Génome Québec, and, a glycerol stock was made for the confirmed positive clones.

3.1.2 Agarose gel electrophoresis:

To prepare 1% agarose gel, 0.5 g of agarose powder (Invitrogen, USA) was mixed with 50 mL 1X TAE buffer diluted from 50X TAE buffer (Bio-Rad, USA), and heated using a microwave oven for 30 seconds with stirring. The heating and stirring was repeated two- three times to ensure proper agarose dissolution. After proper cooling down at room temperature, the molten gel was next poured in the cast where 5 µL of 1% Ethidium Bromide solution (EtBr) (Invitrogen, CA) was added before adding the combs to the cast tray. Gel was allowed to solidify for 30 minutes, combs removed, 1X TAE buffer was added to the chamber until the gel was completely immersed. DNA was migrated on gel to check the concentration of DNA before

proceeding to the next step. Then, the samples mixture and DNA ladder were migrated on gel before visualizing the gel using alpha imager.

3.2 Generation of point mutations in IcsB:

3.2.1 Polymerase chain reaction PCR

Polymerase chain reaction was performed in order to generate point mutations in the coding sequences of the protein. These mutations affect the protein in which it might lose the phenotypic toxicity seen in its wild type, which means it is part of the catalytic domain. The toxicity of the protein is demonstrated through the cytotoxicity assay after transforming the gene in *saccharomyces cerevisiae* strain. In addition, Mutations made for IcsB are H145A, D195A, C306A, C306S, N305A, and N352A. For IcsB homologous proteins, we generated mutations at residues equivalent to histidine 145 and aspartic acid 195 of IcsB. All primers used to generate previously mentioned mutations are listed in the (appendix table A3) table.

To perform PCR mutagenesis, 50 µL reaction needs to be prepared for each mutation using the corresponded primer pairs. A mixture of 29.5 µL of nuclease free water (Wisent, CA), 10 µL of 5X GC Phusion buffer (ThermoFisher, Europe), 1 µL of 10 mM dNTP prepared from100 mM dATP, dCTP, dGTP, and dTTP stock (ThermoFisher, Europe), 2.5 µL of 10 mM forward primer, 2.5 µL of 10 mM reverse primer , 2.5 µL of 50 mM MgCl2 (ThermoFisher, Europe), 1.5 µL of 10 ng template DNA, and finally 0.5 µL of 2U/µL Phusion DNA polymerase enzyme (ThermoFisher, Europe). Then, the PCR tubes were placed into the thermocycler machine (T100™ Thermal Cycler from Bio-Rad), and the following program was used: 1 cycle of 30s initial denaturation at 98°C, then amplification for only 22 cycle in order to avoid the number of possible errors resulted in undesirable DNA. The 22 cycles are composed of

denaturation at 98°C for 5 seconds, annealing for 30s (temperature used as suggested from ThermoFisher software for each primer pair), followed by extension at 72°C for 2 minutes, and a final cycle for the final extension at 72°C for 10 minutes. Then, the temperature was hold at 4°C.

3.2.2 Colony screening PCR:

PCR colony screening was performed in order to confirm the insertion of the gene into the plasmid prior sequencing step. 6-12 colonies were picked per plate, resuspended in 100 μL supplemented with appropriate antibiotic in microcentrifuge tubes, and let grow for 40- 60 minutes at 37°C with shaking. After the incubation was done, the reaction was prepared by mixing 0.4 μL of cell culture with 2.5 μL of 10X reaction buffer, 1 μL of 10mM dNTP mix, 1.25 μL forward primer, 1.25 μL of reverse primer, 0.12 μL of Taq polymerase enzyme, and complete with water to 25 μL. The mixture was mixed and spun down before placing the tubes into the thermocycler. The thermocycler was programmed as follow: 1 cycle of 30s initial denaturation at 95°C, then amplified for 25- 40 cycles that consist of denaturation at 95°C for 30 seconds, annealing for 30s (annealing temperature used as suggested using ThermoFisher software and it differs for each primer pair), followed by extension at 72°C for 2 minutes, and a final cycle for the final extension at 72°C for 5 minutes. Then, the temperature was hold at 4°C. After the cycles were done, 7.5 μL of 5X loading buffer in each tube and let run on agarose gel for 15- 25 minutes before analyzing the results to confirm positive clones. The gel picture was taken to document the results. Afterward, liquid culture was prepared for positive clones to purify DNA the next morning.

After the PCR cycle is done, the PCR product were migrated on gel to check the concentration of DNA before proceeding to the next step. Then, the samples mixture and DNA

ladder were migrated on gel for 6 minutes at 120 voltage. Moreover, alpha imager device was used to visualize the gel and an image was taken to document the gel results.

3.2.3 DpnI enzyme digestion:

This step is important to cleave methylated DNA in order to get rid of template DNA and keep PCR product untouched. To prepare for the digestion, 47 µL of PCR reaction was mixed with 47 µL of 1X fast digest buffer (ThermoFisher, Europe), and 1 µL DpnI restriction enzyme (ThermoFisher, Europe). After that, the mixture was incubated at 37°C for two hours.

3.2.4 DNA purification from PCR:

The DNA was purified using PCR product purification kit following the protocol provided by the supplier (Bio Basic).

3.2.5 Phosphorylation of 5' end:

To introduce phosphate at the 5' end of the PCR product. To prepare phosphorylation reaction, 15 µL of purified PCR product was mixed with 2 µL of 10 mM ATP, 2 µL of forward reaction buffer (Buffer A), and 1 µL of T4 polynucleotide enzyme. Mixture then incubated at 37°C for 30 minutes followed by incubation at 75°C for 10 minutes to inactivate the enzyme.

3.2.6 Intramolecular ligation:

This step was made to start ligating the phosphorylated DNA in order to proceed to the transformation step. A mixture of 5 µL of phosphorylated DNA, 39.5 µL sterile deionized water, 4.5 µL of 10X ligase buffer (ThermoFisher, Europe), and 1 µL of T4 Ligase enzyme

(ThermoFisher, Europe) was prepared and incubated at room temperature for 15 minutes to overnight prior transformation.

3.2.7 Transformation in bacteria:

To perform the chemo-competent transformation, 100 µL of chemo competent DH10B or DH5alpa Escherichia coli strain is needed for each transformation. Cells were mixed with 10 µL of ligated product and incubated with DNA on ice for 15- 45 minutes, followed by heat shock for 45- 60 seconds. After the heat shock, cells were placed back on ice for 5 minutes. Then, 1.1 mL of Luria-Bertani (LB) broth was added to each tube, which were incubated for 30 to 45 minutes at 37°C with shaking at 225 rpm. After the recovery step, cells were pelleted via centrifugation and plated on LB agar plate supplemented with ampicillin antibiotic. Plates were incubated overnight at 37°C. The same protocol applies on transformation with electro-competent cells, except no incubation step on ice for a long time (first step) and glass cuvettes were used for the electric shock. Electro-competent transformation protocol "High-Efficiency Transformation by Electroporation" was taken from (Seidman et al., 2001).

3.2.8 DNA purification:

After confirming the positive clones with either PCR colony screening, 5 µL of cell culture used to make 5 mL liquid culture and incubated overnight at 37°C with shacking. The next morning, cells were pelleted to be ready for DNA extraction. The DNA extraction kit from Bio Basic was used for the purification step. All buffers were provided by supplier along with the protocol used.

3.2.9 Confirmation digestion:

This step can be done as validation before sequencing. A reaction was prepared by mixing 2.5 µL of 10X fast digest buffer, 0.5- 1 µg of desired DNA, 1 µL of each of the required restriction enzymes, and complete with water to 20 µL. reaction was incubated for 10-15 minutes at 37°C. After that, reaction was migrated on gel for 15- 20 minutes to confirm the presence or absence of the right insert in the vector.

3.2.10 DNA sequencing:

Clones were sent for sequencing with either SP6, BGH, or specifically designed primers to confirm the presence of mutations or gene insertion. Sanger sequencing was performed by Genome Quebec facility in McGill University, Montréal, Québec. After sequencing results obtained, they were analyzed using serial cloner, Blast, and chromatogram. Sequences were checked by comparing them to the theoretical map and blast was used to confirm the coverage of the query to the original sequence. 4Peaks software was used to visualize chromatogram when needed.

3.2.11 Bacterial glycerol stock:

Liquid culture of positive clones was prepared. Tubes were incubated overnight at 37°C with shacking in LB-medium with the appropriate antibiotics. The next morning, 900 µL of bacterial culture mixed with 600 µL of 50% sterile glycerol in cryovial tube and stored at -80 °C freezer. Plasmid, strain type, date and strain number were written on the tube as well as in the laboratory stock file.

3.3 Yeast and the toxicity assay:

3.3.1 Yeast transformation

For these experiments, we used BY4741 mat a, an auxotrophic derivative of the *Saccharomyces cerevisiae* S288C strain. One colony was picked from a freshly made plate and transferred to 4 mL of YPD broth. Liquid culture was incubated overnight at 30 °C with shaking. The next morning, OD600 was measured and inoculated the equivalent of OD 0.2 in 50 mL of YPD, followed by growing the culture for 3 hours at 30 °C with shaking. Liquid culture was then transferred to 50 mL falcon tubes and centrifuged at 4500 rpm for 5 minutes. Cells were resuspended in water and re-pelleted, followed by resuspension in 0.1 M LiAc, pellet for 10 seconds at 10,000 rpm, and resuspended in 500 µL of 0.1 M LiAc. Then, 50 µL of cells was aliquoted in a clean microcentrifuge tube for each transformation including negative control. Transformation mixture was added with the following order: 240 µL of PEG 50%, 36 µL of 1 M LiAc, 50 µL of ssDNA, 33 µL of sterile water, and 1 µL of 250- 500 ng DNA. After adding the transformation mixture, cells were vortexed until they are completely resuspended, and incubated for heat shock at 42 °C for 20 minutes. After incubation, cells were centrifuged for 30 seconds at 10,000 rpm and a quick wash with water was performed without resuspending the pellet. Finally, cells were resuspended in 200 µL of sterile water and plated on agar plate containing histidine drop-out for 3- 5 days.

After that, a glycerol stock had to be made to store the desired yeast strain for a longer period of time in order to make the strain reproducible. To prepare that, one yeast colony was picked from the plate and transferred to cryovial tube with 0.5 mL of 25% glycerol/ YPD broth.

Tubes were stored at -80C freezer. All information about strain, plasmid, date was written on the tube.

3.3.2 Yeast cytotoxicity assay:

Yeast strains were steaked on plates from glycerol stock, and one colony was picked for each strain. Cells were incubated overnight at 30°C with shaking in SD-his broth supplemented with 2% raffinose sugar to prime cells for the induction. The next morning, cells were pelleted, resuspended in 100 µL NaCl 0.9% and three serial dilutions were made (1/10, 1/100, and 1/1000) beside the undiluted cells by taking 10 µL of cells into 90 µL of 0.9% NaCl; the same apply for each dilution. After that, 3-5 µL of cells spotted on the two histidine plates; one supplemented with 2% glucose and the other one with 2% galactose. Cells then incubated at 30°C for 3 days. After the incubation period, a picture of the plates was taken to document the results.

3.3.3 Image J software:

Image J is an image analysis software that can calculate the pixels values in a given image (Schneider, et al., 2012). It allows to make semi-quantitative analyses of certain images.

3.3.4 KaleidaGraph software:

KaleidaGraph 4.5.4 is software used for graphing and analyzing data. It was used to plot histogram image and calculate standard deviation and *p*-value for toxicity assay experiment.

3.4 Protein expression confirmation by Western Blot

3.4.1 Yeast protein extraction:

Protocol was taken from (Kushnirov, 2000). The same liquid cultures used to perform the dot assay were subjected for induction in preparation for protein extraction and western blot later. To do that, cells were diluted to have OD600 of 1.0 and incubated for two hours at 30°C. Followed by adding 4% galactose sugar to each culture and incubated at 30°C for 4 hours. After the induction is done, 500 µL of cells were centrifuged, concentrated five times, resuspended in 100 µL of deionized water. Followed by adding 100 µL of 0.2 M NaOH and incubated for five minutes at room temperature. Following centrifugation at 10,000 rpm, the cellular pellet was resuspended in 50 µL of 1X Laemmli contains 1mM of PMSF and placed in a boiling water bath for five minutes. Finally, tubes were centrifuged, and supernatant was transferred to new clean tubes to use them for western blot.

3.4.2 Western Blot:

For most western blot performed, 12% SDS resolving gel was used. It was prepared by mixing 2.15 mL of water with 1.5 mL acrylamide, 1.25 of 1.5 M TRIS HCL PH 8.8, 50 µL of 10% SDS, 50 µL of 10% APS, and 2 µL of TEMED. Mixture transferred to the glass plate and isopropanol was added to remove bubbles. Isopropanol is removed when the gel solidifies. The following layer is the 5% stacking gel, which can be prepared with mixing 1.46 mL of water with 250 µL of acrylamide, 250 µL of 1M TRIS HCL PH 6.8, 20 µL of 10% SDS, 20 µL of 10% APS, and 2 µL of TEMED. Mixture was added between the glass plate carefully to avoid bubble, and the comb was placed on top of it until it is solidified.

Extracted proteins along with the protein ladder were preheated at 95°C for five minutes and migrated on SDS gel at 100 V until the protein samples reached the resolving gel. Then, the voltage was increased to 180 V and let run until it reaches the end of the gel.

After the electrophoresis, semi-dry transfer was performed. The gel was transferred on a PVDF membrane placed on top of a blotting paper, and then covered with another blotting paper to make a sandwich. Blotting papers and membrane were presoaked in transfer buffer. After that, the sandwich transferred to the Trans-Blot transfer system machine and run for 45 minutes.

After the transfer step is completed, the membrane was blocked with 5% TBST-milk for 20 minutes. Followed by two washes in TBST buffer for 5 minutes each on shaker. Afterwards, membrane was incubated with either anti-flag or anti-myc as primary antibody with dilutions of 1/1000 or 1/5000 for one hour on rolling shaker at room temperature. Then, the membrane was washed three times with TBST for 5 minutes each, followed by incubating the membrane with the secondary antibody (anti-mouse); a dilution of 1/20000 in TBST-milk for 30-60 minutes at room temperature. Subsequently, the membrane was washed three times for 10 minutes each in TBST on a rolling shaker. Next, the membrane is ready for visualization and was transferred to 1X TBS buffer solution.

After the blotting step, the membrane was visualized using ChemiDoc imaging device from Bio-Rad. The ECL clarity or enhanced luminol-based chemiluminescent substrate from Bio-Rad was used to detect HRP on the membrane. A mixture of 1:1 enhancer solution and peroxide solution was prepared and spread on the membrane for taking the image. Results were produced using either Chemi or High sensitivity method with default exposure time.

3.5 Sequence alignment of toxic and non-toxic homologs:

After the toxicity assay was performed and the toxic and non-toxic homologs were identified, Clustal Omega software was used to generate sequence alignment for the killing and non-killing proteins separately. Catalytic domain was colored in grey, and the fully conserved residues were colored in red and shaded in grey.

Chapter III Results

3.6 Generation of the yeast expression plasmid:

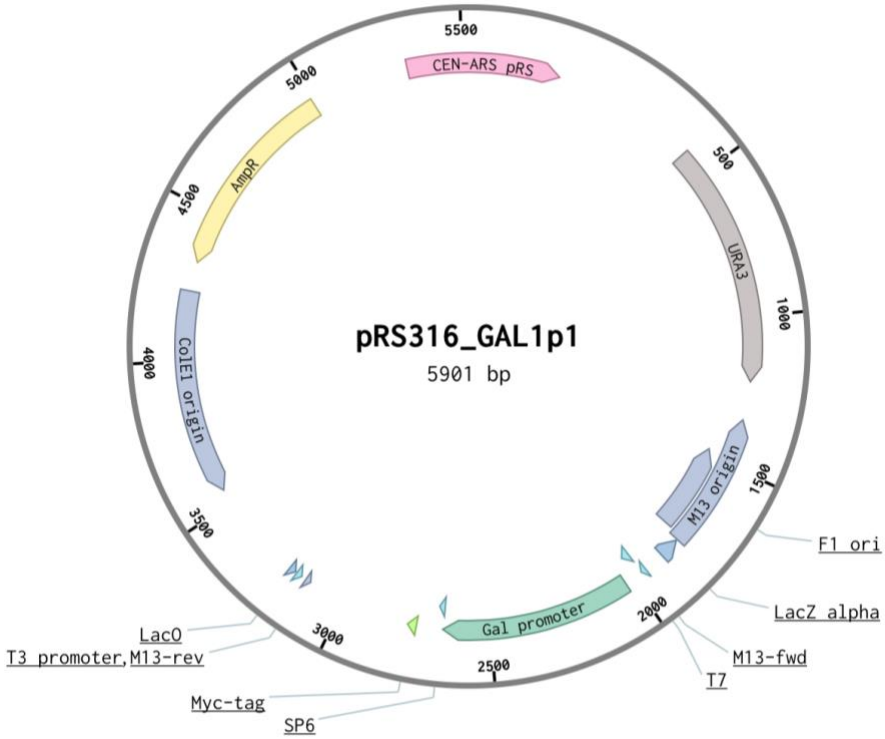
The purpose of having the galactose promoter is to control the protein expression when galactose is present or absent from the media. Therefore, the effect of the protein on cells can be visible using certain assays and can also be detectable though western blot. GAL1p was synthesized by (IDT). After receiving GAL1p, we realized that the sequences of M13 forward and reverse were already present in pRS313 and pRS316. Since it rendered the use of these primers for sequencing purposes impractical, they were exchanged for the SP6 and BGH primers; the resulting plasmid was named GAL1p1. Unfortunately, the sensitivity of our Myc-tag antibody appeared insufficient to detect IcsB (data not shown), probably due to low expression level induced by its cytotoxicity. For technicality purposes, we inserted the sequence of 3X Flag-tag replacing Myc-tag which yielded the GAL1p2.

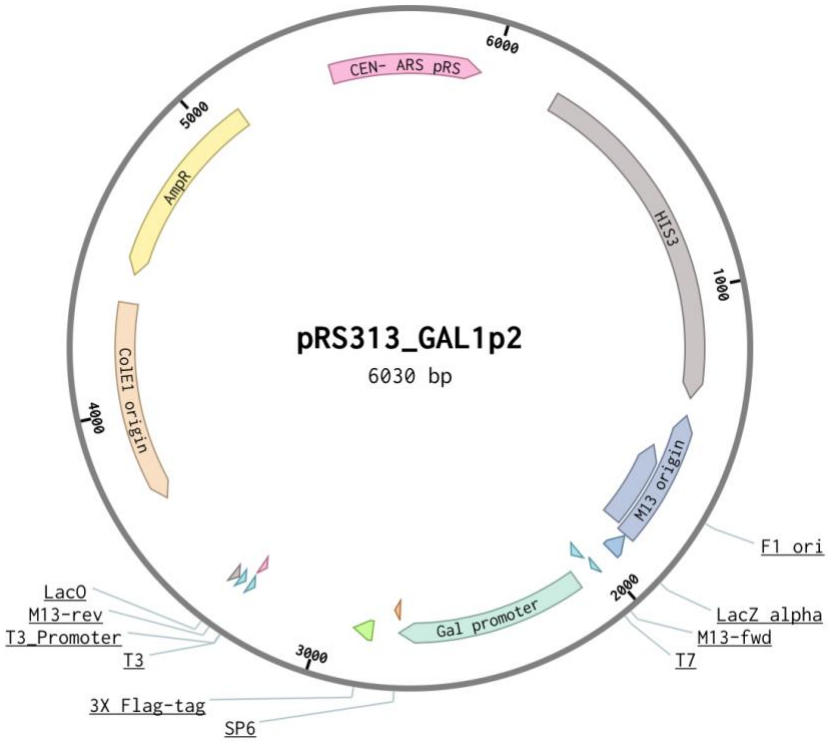
In addition, the empty plasmid, IcsB, and homologs were transformed into *S. cerevisiae* yeast strain in order to have them ready to perform the cytotoxicity assay later.

3.7 Generating IcsB homologs and Cytotoxicity assay results:

IcsB and its homologs were cloned into pRS313 GAL1p2. The chaperone IpgA was cloned into pRS316 GAL1p1 (figure 3.1).

Figure 3.1. Maps of plasmids used for cloning and genes expressions





Prior to testing the toxicity of IcsB homologs, control experiments using the empty vector, IcsB and some of its mutations were performed to establish the toxicity assay by reproducing prior studies done by Slagowski et al. where they demonstrated IcsB toxicity upon expression in yeast (Slagowski et al., 2008). IcsB C306A, IcsB Δ80 that was generated by deleting 80 amino acids from the N-terminus of the protein, and the double mutant, IcsB Δ80 C306A that was generated by removing the first 80 amino acids in the C306A background. It is noteworthy to mention that both mutations of IcsB, which are IcsB Δ80 and IcsB C306A, were already tested by a former colleague in the lab. Their results indicated that IcsB Δ80 toxicity was completely rescued while C306A showed partial rescue. However, in this project we made the expression of proteins controlled by the presence of galactose in the media and the glucose was used as control. Therefore, as shown in (figure 3.2), cells harboring the empty vector (negative control) grew normally in both plates, as expected. Furthermore, the expression of WT IcsB blocked yeast cells growth, while cells grew normally when IcsB expression was inhibited. Both of the mutated version with the truncated N-terminus (Δ80), relieved the toxicity and cells grew healthily. By contrast, the point mutation IcsB C306A displayed the full toxicity phenotype, as it was undistinguishable from the WT. This raised a question about which part of the IcsB coding sequence is responsible for its toxicity since removing amino acids from the N-terminal or the catalytic mutation C306A ablated the effect of IcsB expression on yeast growth. Moreover, the result of C306A mutation conflicts with the results from (Liu et al., 2018). The only difference in their experimental setup was that they co-expressed the chaperone IpgA with all mutants without providing an explanation. Therefore, the toxicity experiment was repeated by co-expressing IpgA. The result presented in (figure 3.3) revealed that the toxicity of C306A was rescued when co-expressed with IpgA. Hence, the chaperone IpgA plays an important role in rescuing the

killing phenotype seen in C306A mutation. As yet, the mechanism behind this phenomenon remains unknown.

Figure 3.2. Establishing the toxicity assay.

Galactose plate indicates the inducing condition while glucose plate indicates the non-inducing condition used as control.

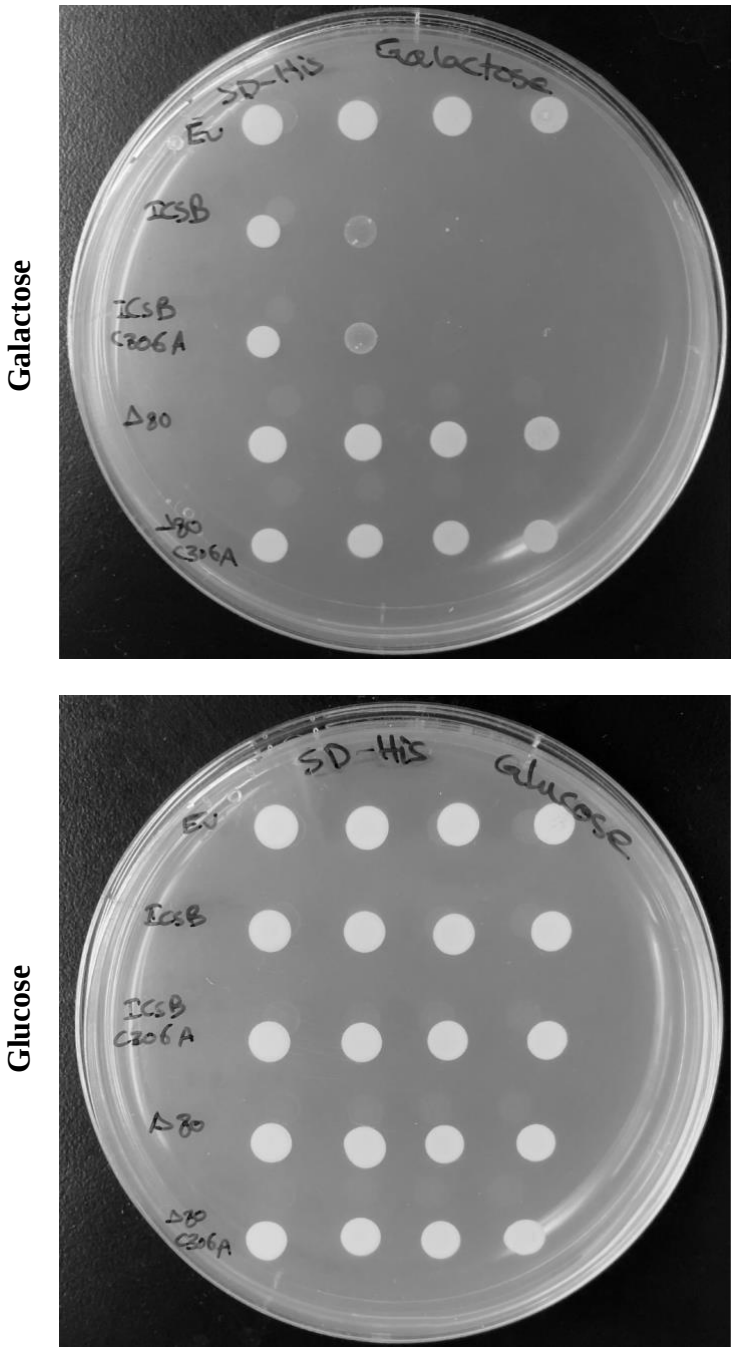
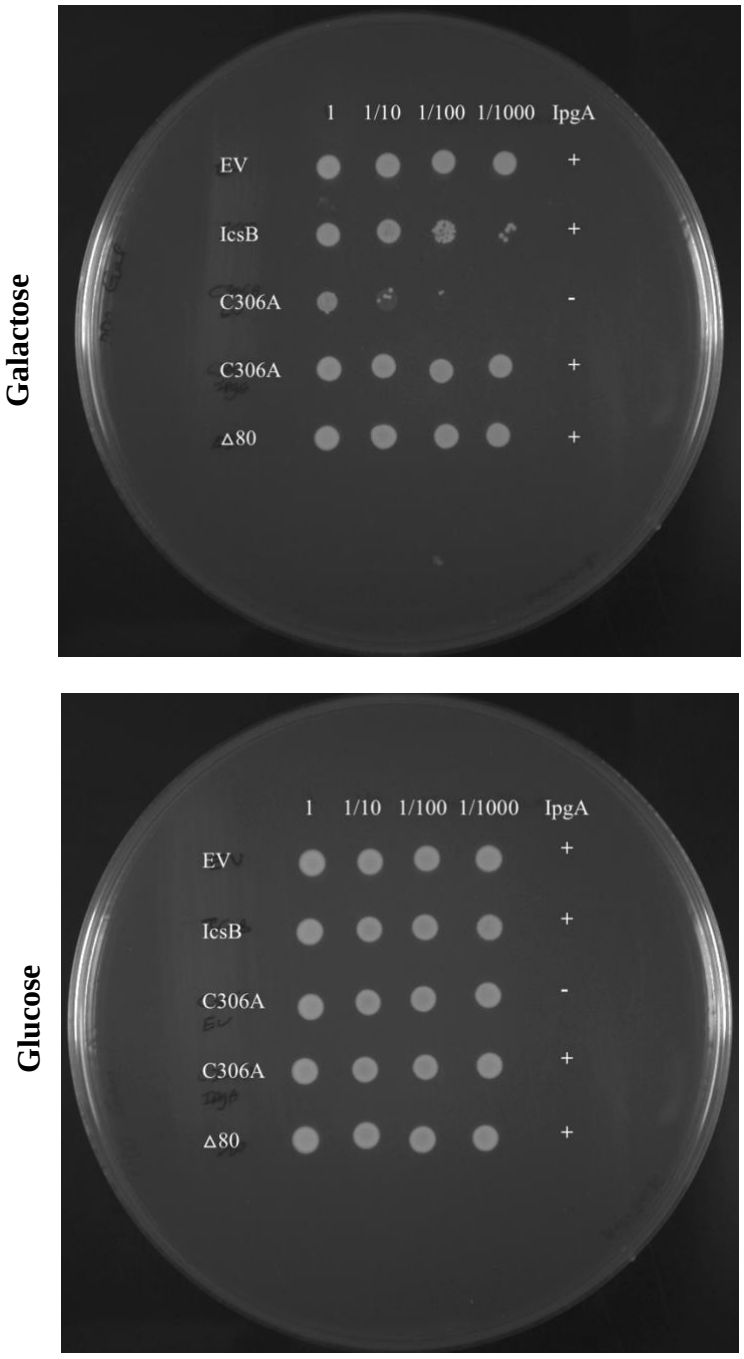


Figure 3.3. Yeast cytotoxicity assay of IpgA co-expression.

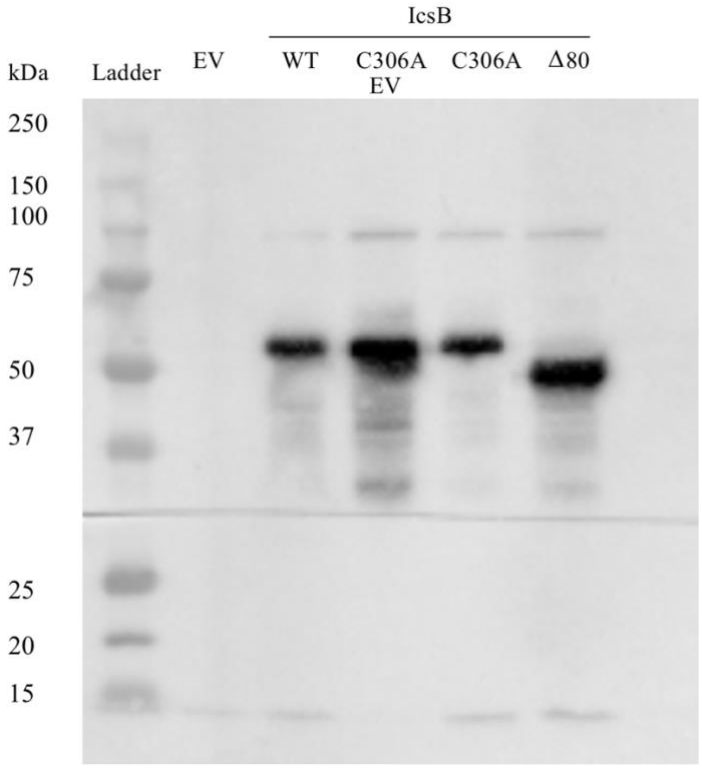
The effect of IpgA co-expression on IcsB C306A mutant's toxicity. (+) indicates the presence of IpgA and (-) indicates its absence. The expression of proteins is shown in the galactose plate. The glucose plate was used as control.



Furthermore, to relate the phenotypic toxicity to the protein expression, western blot was performed. Yeast cells were first incubated overnight in non-inducing condition, followed by four hours induction, and protein extraction before running on a gel for western blot. As shown in (figure 3.4), the expression of both IpgA, IcsB, and its mutants were detectable

Figure 3.4. Western Blot of IpgA candidates' protein expression

The protein size of IcsB, IcsB Δ80 and IpgA are 56 kDa, 47 kDa and 15 kDa respectively.

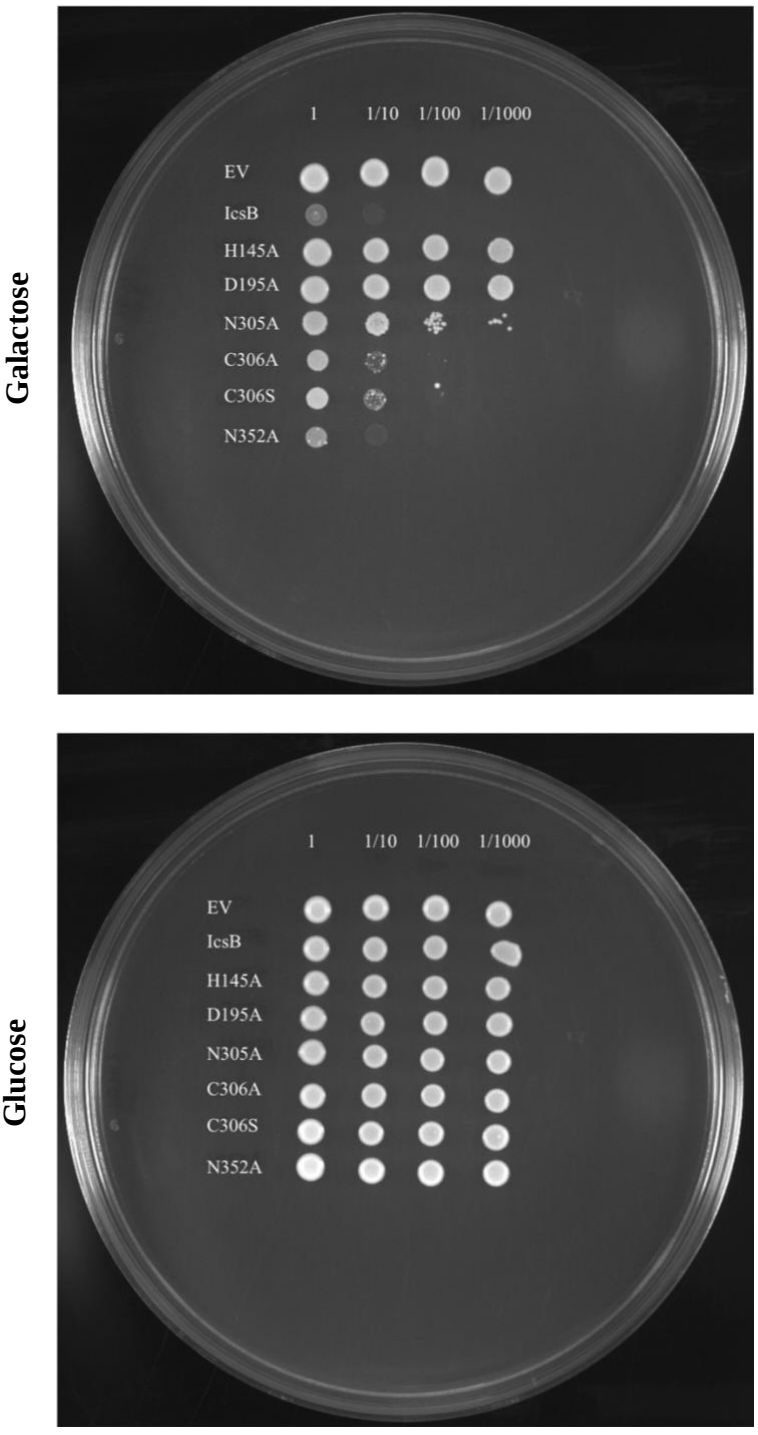


However, since the catalytic cysteine showed partial toxicity and it was able to be removed when the chaperone IpgA was co-expressed, we decided to substitute Cys 306 with serine to test whether or not the mutation will help removing IcsB toxicity without the help of the chaperone. A mutation of IcsB C306S was generated using PCR. After confirming the positive clones, plasmids were transformed into yeast *S. cerevisiae*. Consequently, the toxicity assay was performed as described before. Therefore, with the help of empty plasmid and IcsB as controls, the result is shown in (figure 3.5) were as follow, mutations of His-145-Alan and Asp-195-Ala appeared to relieved IcsB toxicity, while Asn-305-Ala showed partial toxicity when expressed on galactose. As for, Cys-306-Ala, and Cys-306-Ser, and Asn-352-Ala yeast cells were killed upon their expressions, similarly to the WT.

In addition, when the sequences of IcsB homologs were aligned (shown previously in chapter II), full and semi-conserved positions appeared among the sequences. Especially, two amino acids were happened to be fully conserved and polar. These amino acids are asparagine 305 and asparagine 352. We hypothesized that the catalytic domain might expand beyond amino acid serine 319. To test whether they have impact on the toxicity phenotype as they are being part of the catalytic domain, mutations on the wild type IcsB was performed to generate IcsB N305A, and IcsB N352A to be tested later.

Figure 3.5. Yeast cytotoxicity assay of IcsB catalytic mutations.

The expression of proteins is shown in galactose plate while glucose plate was used as control.



Furthermore, we wanted to test the effect of chaperone IpgA on IcsB new mutations that showed complete or partial toxicity. To do that, yeast transformation of pRS316 GAL1p1 IpgA into strains have pRS313 GAL1p2, IcsB, IcsB N305A, IcsB C306A, IcsB C306S, and IcsB N352A and the empty vector pRS316 GA1p1was also transformed into same strain to use as control. After the spotting assay was completed, the results were collected after three days. As shown in (figure 3.6), IcsB H145A and D195A remove toxicity without co-expressing IpgA. The cysteine mutations C306A and C306S completely rescued the toxicity when IpgA was co-expressed, unlike strains devoid of IpgA expression which were not able to grow. As for the N352A, it remains toxic with and without IpgA co-expression. In addition, the protein expressions were detected through western blot and the results are shown in (figure 3.7).

Figure 3.6. Co-expressing IpgA with IcsB catalytic mutations.

Galactose indicates the inducing condition while glucose indicates the non-inducing condition used as control. *EV316 indicates the empty vector pRS316GAL1p1 used to show the wild type phenotype and the effect of IpgA co-expression on the same plate.

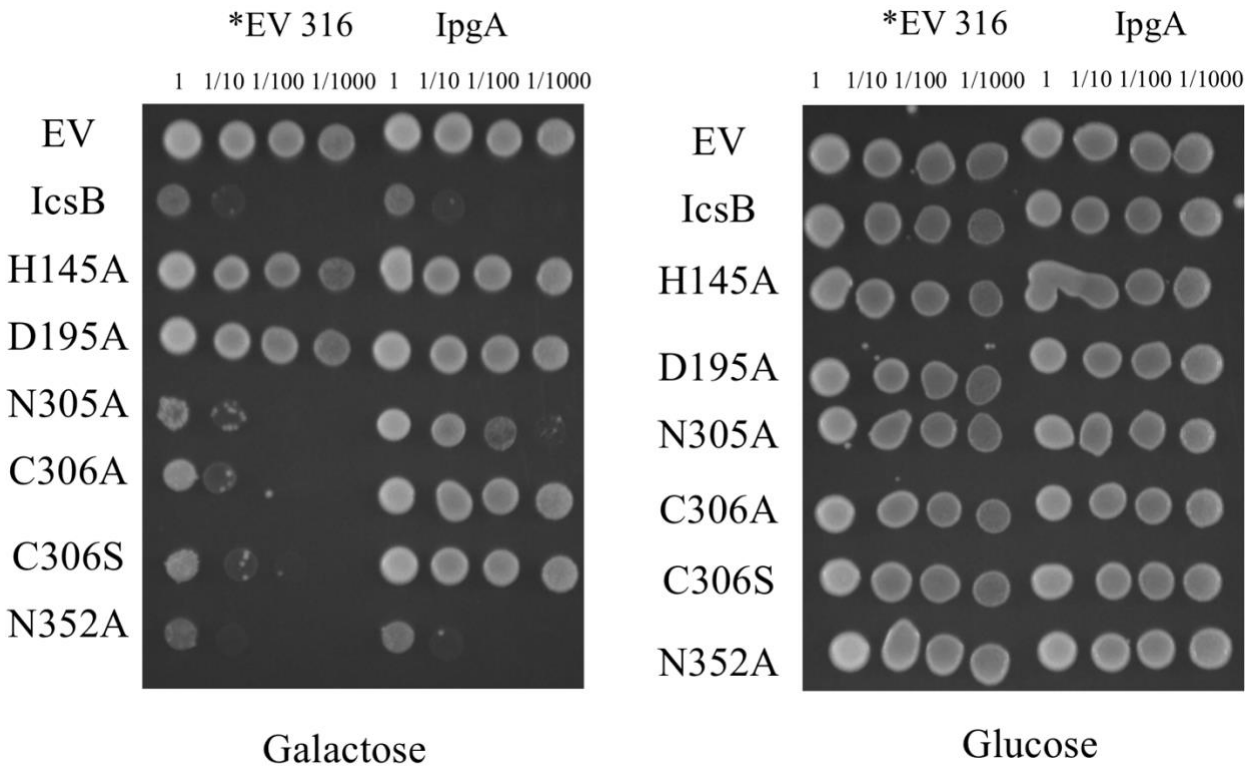
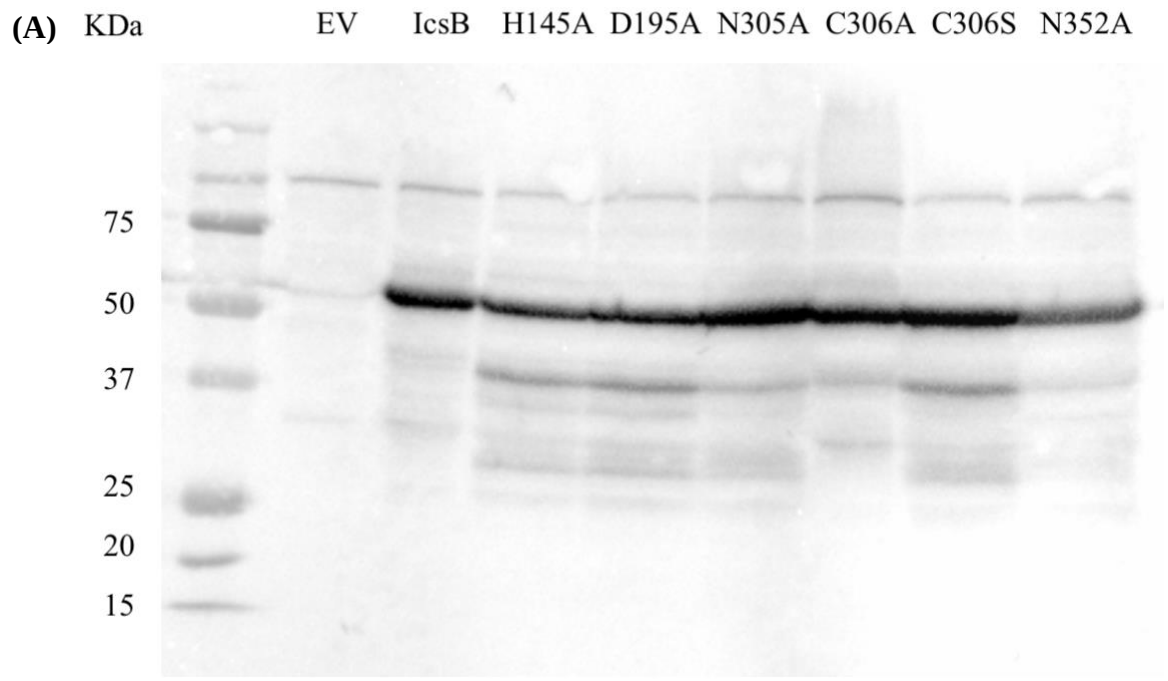
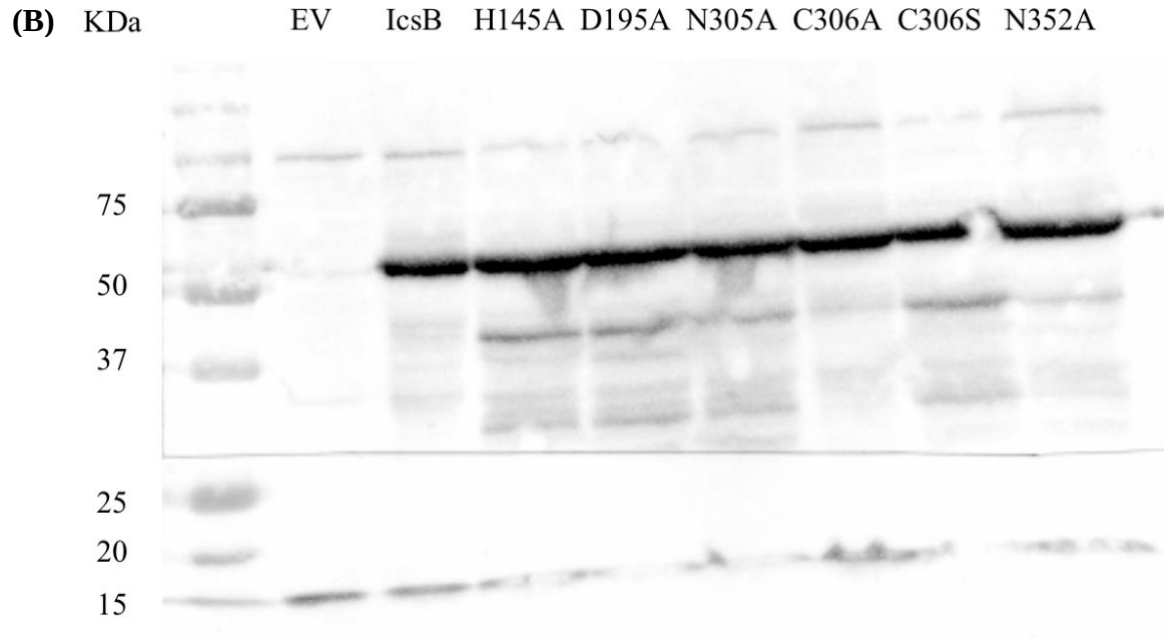


Figure 3.7 Western Blot of IcsB wild type strains and their catalytic mutations.

The protein size of IcsB is 56 kDa. figure **A** shows protein expression of wild type strains, and figure **B** shows IpgA co- expression for the same mutations.





After being successfully able to validate the cytotoxicity phenotype seen in IcsB among expression in yeast, the toxicity of homologous proteins was next to be tested. As mentioned previously, the homologs were synthesized and inserted into pRS313 GAL1p2 vector, the same vector used for IcsB experiment using the standard cloning technique. Constructs were then transformed in the *E. coli* DH10Beta strain. PCR colony screening and confirmation digestions were performed prior sequencing process. Suspected positive clones were picked to make liquid culture in order to extract their DNA. After that, purified plasmids were sent to be sequenced. After confirming the presence of each gene though sanger sequencing, homologs were then transformed in yeast *saccharomyces cerevisiae* strain and the cytotoxicity assay was performed.

In order to perform the cytotoxicity assay for the homologous proteins, the empty vector, pRS313 GAL1p2 and IcsB were used as controls in this experiment. After the transformation of the homologs in yeast, cells were picked for an overnight incubation in order to do the spotting test. Cells were spun down and resuspended in saline-like solution. Serial dilutions were made of 0.1, 0.01, and 0.001 and spotted on glucose and galactose single dropout media. After incubation period of three days, the results were collected. The result is shown in (figure 3.8). In the case of pRS313 GAL1p2 that was used as control, cells were viable, unlike IcsB, where it prevents yeast cells from growing. Moreover, the expression of homologs H1, H2, H3, H4, H5, H8, and H10 showed similar phenotypic effect as IcsB. Yeast cells were not able to grow upon their expression. On the contrary, homologs H6, H7, H9, and H11 did not affect the growth of yeast cells. The experiment was run in triplicate, and we were able to get quantitative measures of their growth using Image J software, and it was used to create a semi-quantitative analysis for the toxicity assay. A histogram of the result was created using KaleidaGraph 4.5.4 and presented in (figure 3.9). The student T-test was calculated for each protein regarding galactose and glucose

growth condition and the overall values for the toxic proteins were <0.01, and the difference is considered to be statistically significant. Therefore, to show the differences in growth between the toxic and non-toxic states. To count the pixels, the lane of 1/10 dilution was used in both galactose and glucose plates. For instance, the average and standard deviation counted for IcsB and the toxic homologs on galactose plate were ranged between 0-29 and 0-23, while for the same strains plated on glucose, they ranged between 232-403 and 0-36 respectively. As for the negative control (empty vector) and the non-toxic homologs, in galactose plates the average and standard deviation counted for their growth ranged between 415-484 and 11-47, while for the same strains plated on glucose, they ranged between 368-487 and 3-24 respectively.

Figure 3.8. Yeast cytotoxicity assay of IcsB homologs.

The effect of proteins expression on yeast cells is displayed in galactose plate. Glucose plate is used as control.

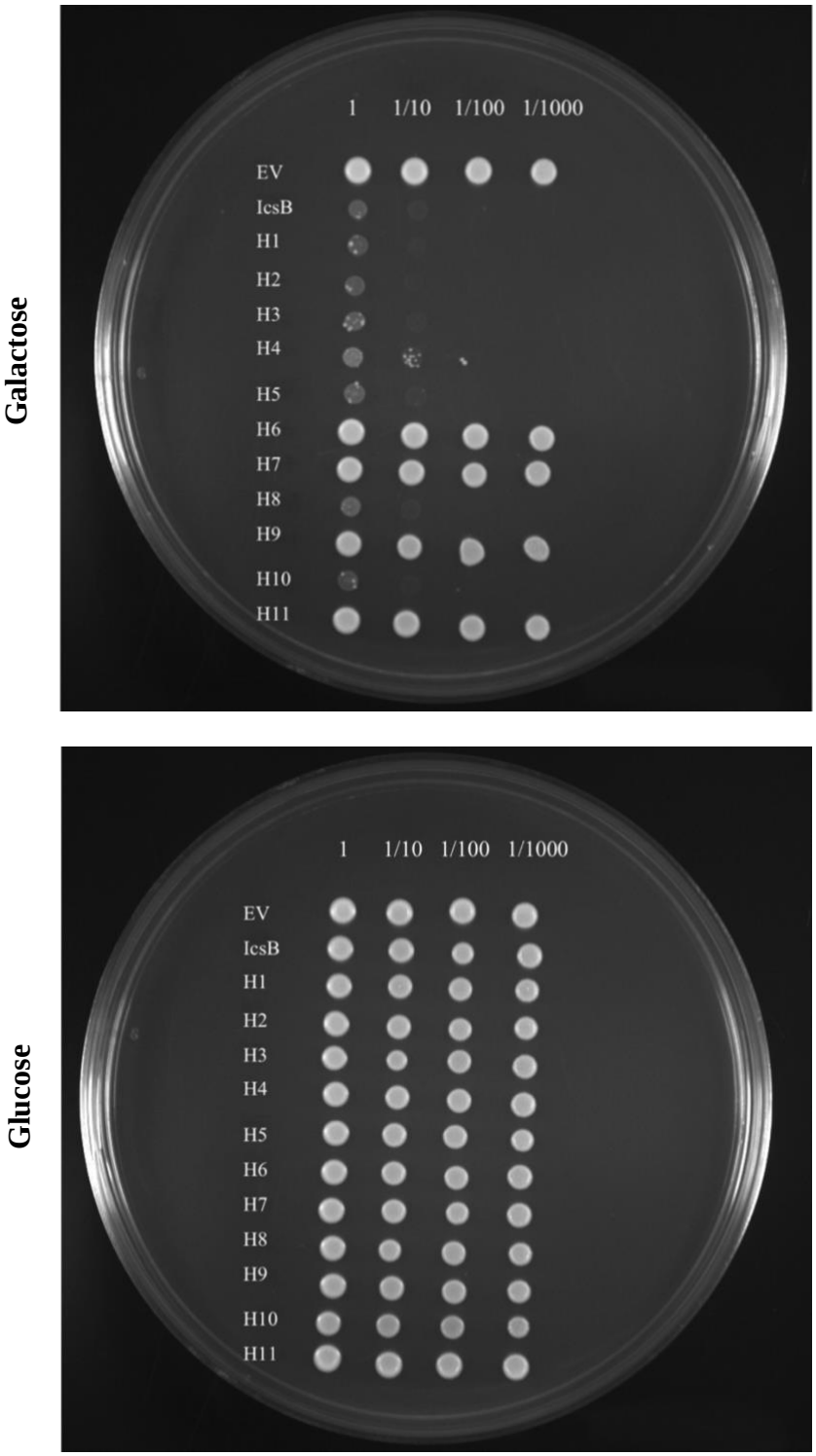
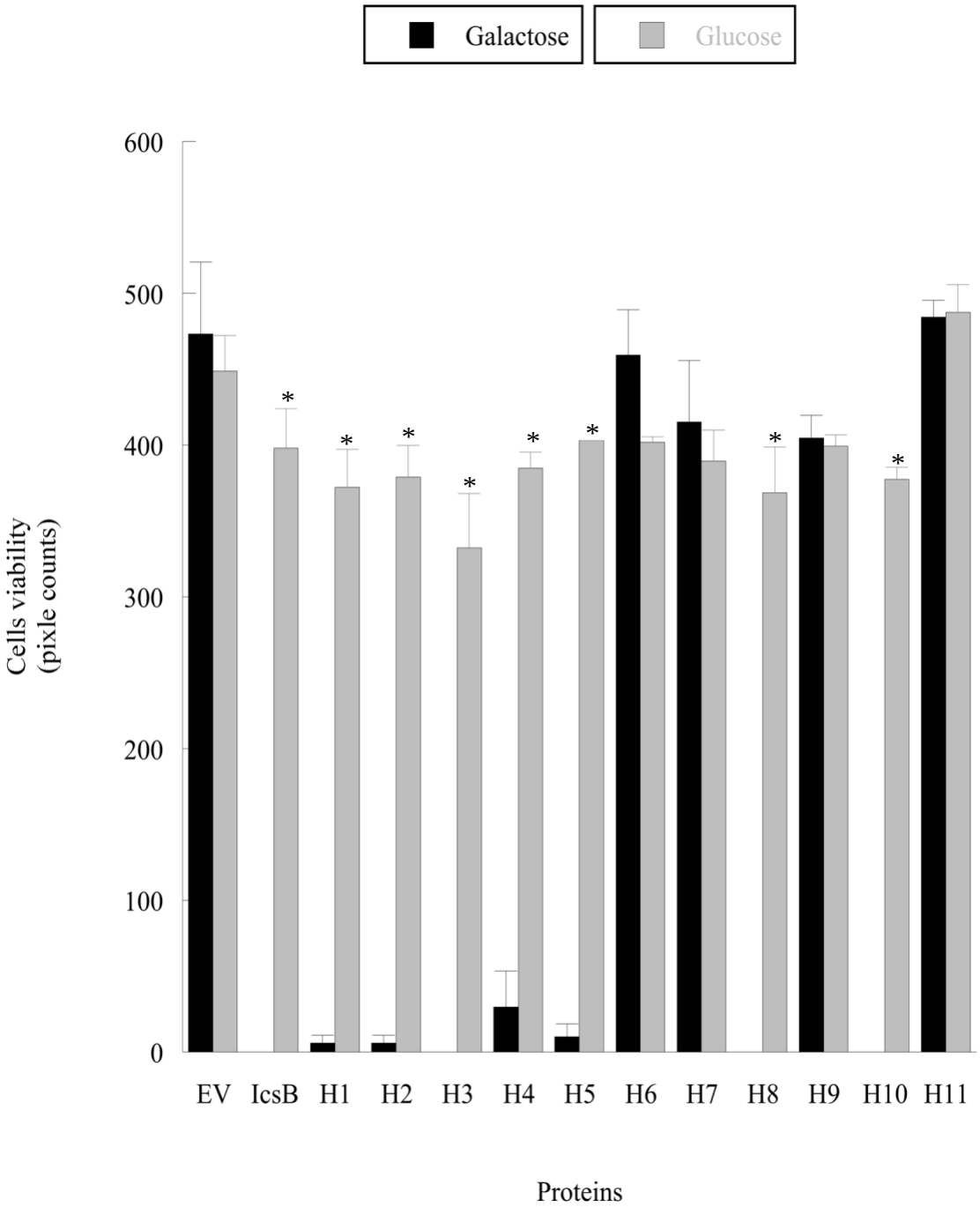


Figure 3.9. Histogram representative of homologous' toxicity assay.

Error bars represent standard deviation. (*) represents the overall value of the student T-test done for each protein regarding galactose and glucose growth conditions, and it was calculated to be <0.01.

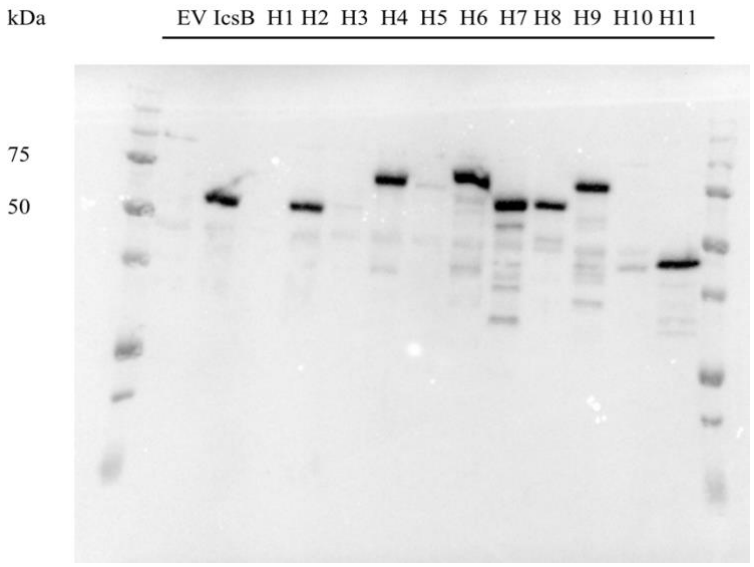


To relate the phenotypic toxicity to the protein expression, western blot experiment was performed. The experiment was done as described in detail in the materials and methods. Results are shown in (figure 3.10). All proteins in addition to IcsB were able to be detected on the membrane except for homolog (H1), BopA, in which expression of the protein could not be detected. As in the case of homolog H1, H3, H5, and H10, the level of expression was very low compared to the other homologs. Moreover, when those four proteins were repeated on a separate membrane, the expression of homolog H3, H5, and H10 were detected unlike homolog H1 that remained undetectable.

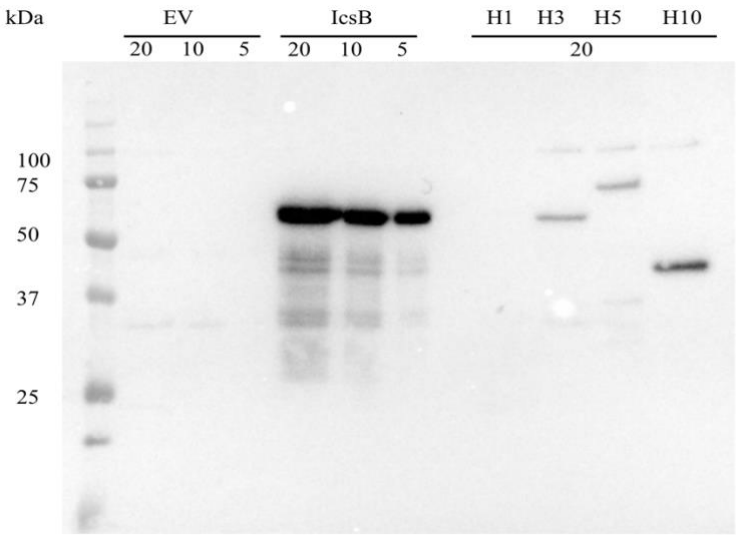
Figure 3.10. Western Blot of IcsB homologs' protein expression.

The membrane in figure **B** shows H1, H3, H5, and H10 that could not be detected in in the first figure **A**. The numbers (20, 10, and 5) indicate the volume loaded. Proteins' sizes are present in figure **C**.

(A)



(B)



(C)

Protein	IcsB	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11
kDa	56.3	56.8	56.6	56.9	73	72.8	73.7	57.5	56.4	73.5	42	40.8

Furthermore, to investigate more about the relativity between the homologs and IcsB, using PCR mutagenesis technique, a mutation of the catalytic triad had to be generated. In reference to (Liu et al., 2018) and (Slagowski et al., 2008), the catalytic triad of IcsB are Histidine 145, aspartate 195, and Cysteine 306 which were found to be conserved in the homologs. Due to the partial toxicity of IcsB Cys-306-Ala mutant that requires co-expressing the chaperone IpgA to be fully rescued, homologs catalytic cysteine mutation was excluded from being tested. Moreover, Alanine substitution to His145 was produced in each of the toxic homolog, which are H1, H2, H3, H4, H5, H8, and H10. Several steps were done starting from PCR to transformation in bacteria and DNA extraction in order to introduce the mutation. Also, to confirm its presence though sanger sequencing before proceeding to the next step. Consequently, the plasmids of the toxic homologs, were transformed in yeast and the survival assay was performed as described previously. As for the controls, the empty vector pRS313 GAL1p2, IcsB, and its mutated version IcsB H145A were used as positive and negative controls. Furthermore, as shown in (figure 3.11), the mutated version of the toxic homologs diminished the toxicity phenotype shown in the wild type strains, similar to the effect of the H145A mutation in IcsB.

To validate the expression of the homolog proteins, western blot was performed in order to draw a conclusion that the protein is expressed normally, and the mutation did not affect their stability in any way. The results are shown in (figure 3.12) for both wild type and non-toxic mutant strains. All protein expressions were able to be detected except for H8 that was undetectable. This might be due to a technicality issue since we were able to detect its expression previously in (figure 3.10 A).

Figure 3.11. Yeast cytotoxicity assay of mutated IcsB homologs (Histidine 145).

The effect of protein expression is displayed in galactose plate, while glucose plate was used as control.

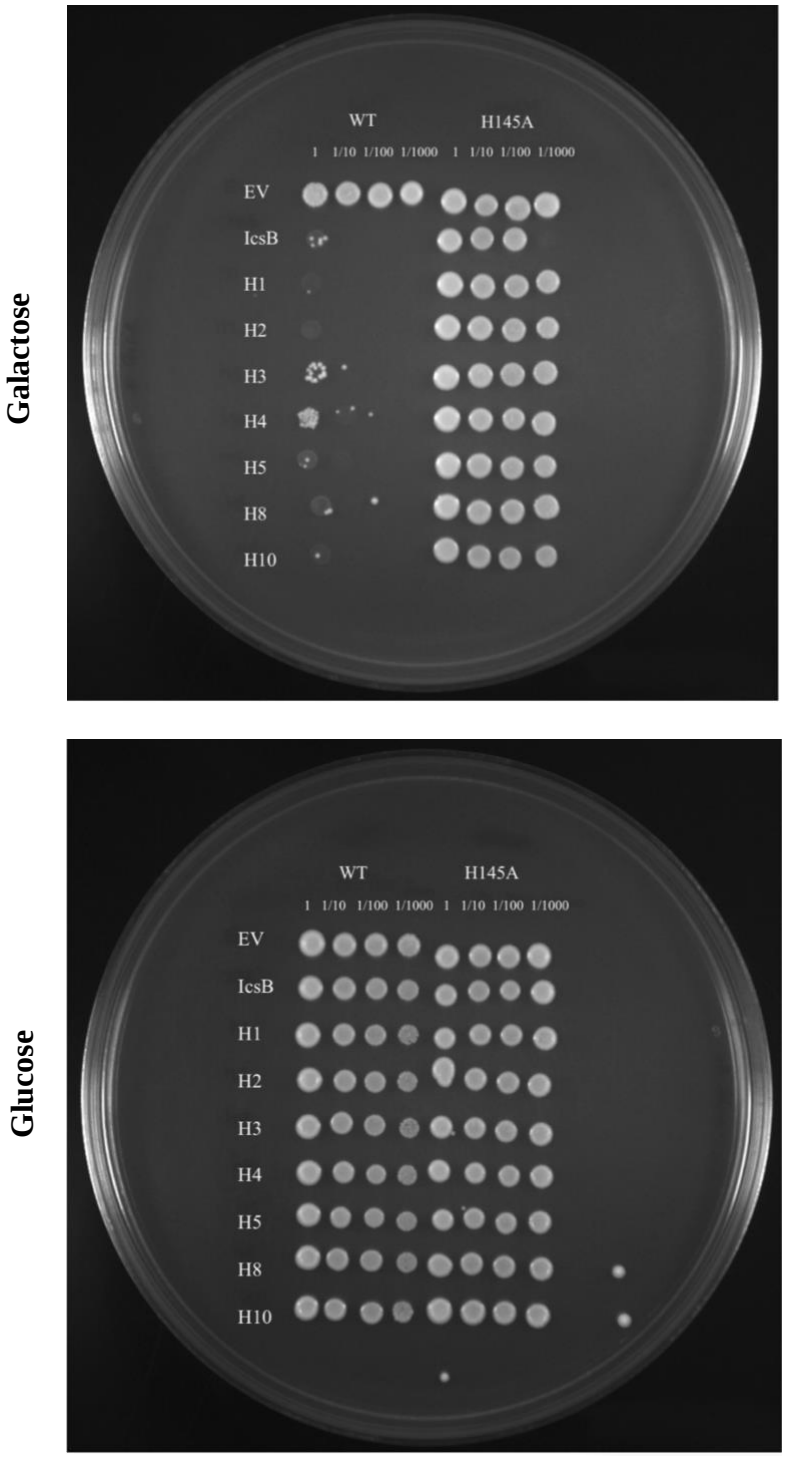
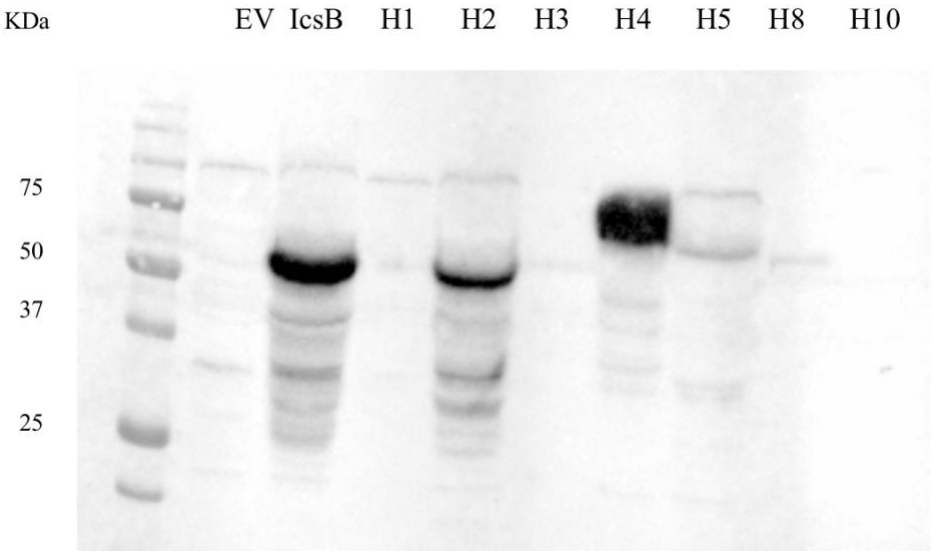


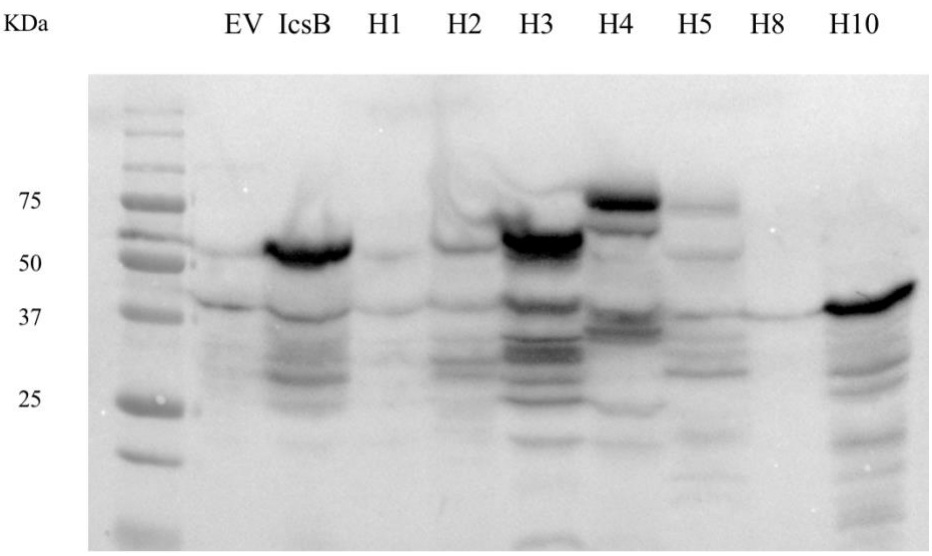
Figure 3.12. Western Blot of histidine mutation of IcsB homologs

Figure **A** represents protein expression of wild type toxic IcsB homologs, while figure **B** represents the protein expression of IcsB homologs non-toxic histidine mutation. **C** represents the proteins' sizes. The induction period for both was 4 hours.

(A)



(B)



(C)

Protein	**IcsB**	**H1**	**H2**	**H3**	**H4**	**H5**	**H8**	**H10**
kDa	56.3	56.8	56.6	56.9	73	72.8	56.4	42

3.8 Results of toxic and non-toxic proteins sequence alignment:

The sequences of the toxic and non-toxic homologs were aligned separately using Clustal Omega. The result of the alignment revealed that the conservations among the killing homologs started from the catalytic domain. The sequence alignment is present in full in the (appendix figure A1). In addition, several semi and full conservations found, especially near the C-terminus of the proteins. Moreover, the last fully conserved amino acid in the sequence was actually the asparagine 352, which found to be conserved amino acid among all homologs aligned together.

As for the non-killing proteins, which are H6, H7, H9, and H11, the conservations are clustered manly in the catalytic domain and the C-terminus of the proteins. The N-terminal part of these four proteins is not conserved. Moreover, the alignment shows two fully conserved amino acids after the asparagine 352 (conserved in all homologs). These residues are asparagine 451 and leucine 494 in homolog 6, as shown in (appendix figure A2), noting that these two amino acids are not conserved in IcsB.

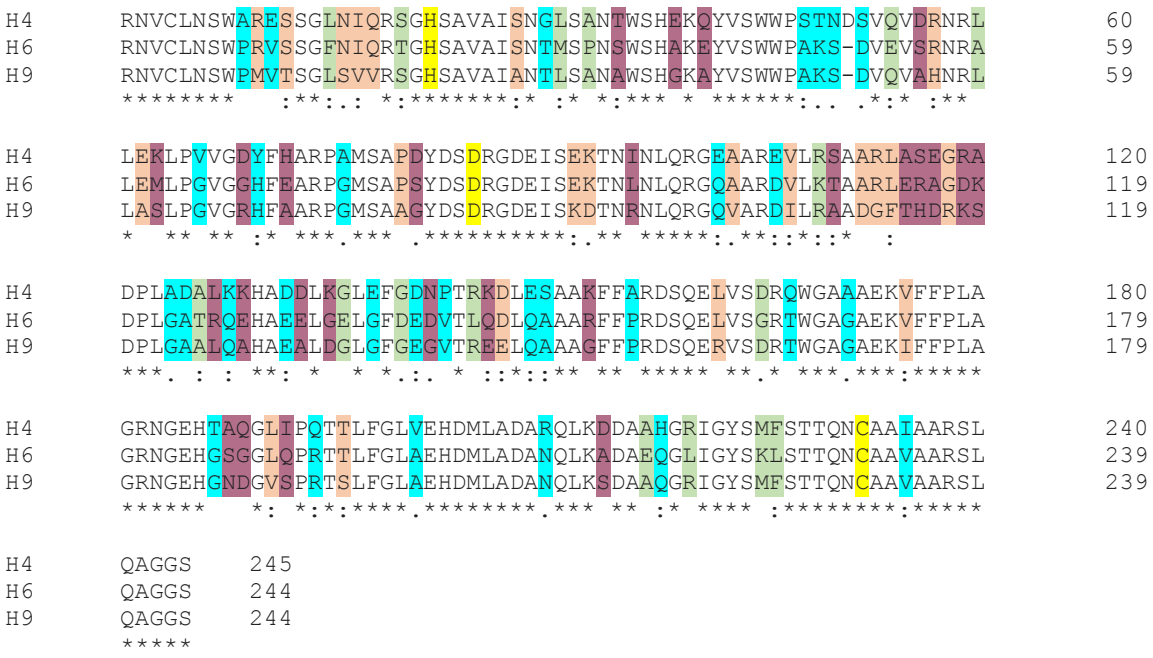
However, the conservations found among the toxic and non-toxic proteins provide evidence about the probability of sharing the same ancestors and being evolutionarily related.

Furthermore, since the three homologs proteins H4, H6, and H9 have more homology than others and H4 is the only toxic protein among the three, we were thinking if their high percentage of homology was because they arrive from the same ancestor. We were wondering if generating mutation on H6 and H9 would have an effect and make them toxic. Therefore, we aligned their sequences together using Clustal Omega looking for possible mutations. Sequence alignment is presented in (figure 3.13). The alignment retrieved more conservation among the three sequences. The highlighted residues represent the possible mutations to be generated in H6 and H9 based on their homology with H4. The catalytic residues equivalent to IcsB catalytic

triad are highlighted in yellow. The multiple colors represent different conservations, such as the highlight in turquoise represent conversations in H6 with H9 not including H4. The conservations in H4 with H6 not including H9 are highlighted in apricot. Conservation in H4 with H9 not including H6 are highlighted in green, and when no conservation occurred, it is highlighted in purple and (*) represents the fully conserved residues.

Figure 3.13. The sequence alignment of H4, H6, and H9 catalytic domain.

The catalytic residues are highlighted in yellow. (*) represents the fully conserved residues.



Color codes:

Catalytic residues.

Conservation in H6 with H9 but not H4.

Conservation in H4 with H6 but not H9.

Conservation in H4 with H9 but not H6.

No conservation.

3.9 Conclusion:

IcsB protein was recently proven to be endowed with an acyltransferase activity, which is crucial for an important step of the intracellular lifecycle of *Shigella* (Liu et al., 2018). PSI-BLAST was used to search for homologous protein in NCBI database using IcsB catalytic domain histidine 145, aspartate 195 and cysteine 306 (Pei et al., 2009 and Liu et al., 2018). As showed in chapter II, search revealed 11 proteins share homology with IcsB. Therefore, these proteins were synthesized and sup-cloned into a vector that express galactose promoter in order to study the effect of the protein expression on yeast cells. The results demonstrated that seven out of the 11 homologs identified in Chapter II are cytotoxic in the budding yeast *S. cerevisiae*. Moreover, when one of the catalytic triad residues (His-145) was mutated to alanine in the toxic homologs and IcsB, the cytotoxicity vanished. Furthermore, three of these homologs, H4, H6 and H9, showed more similarities (>70%) sequence identity although only one was cytotoxic despite similar expression level. They displayed discrete sequence conservation that might be useful to understand the connection between their enzymatic activity and their phenotype.

Chapter IV Discussion

The autophagy mechanism of a host cell relies on the degradation of the cytoplasmic contents in a double membrane compartment. However, xenoautophagy is a type of autophagy that target foreign particles (e.g. bacteria, parasites or viruses) that have invaded host cells. LC3, an autophagosome marker, was found to be recruited around type III secreting *Shigella* (Ogawa et al., 2005 and Campbell-Valois et al., 2015). In the case of *Shigella*'s invasion, the T3SS and the secreted translocators allow the plasma membrane of the epithelial cells to engulf *Shigella* in a phagocytosis-like manner. Other T3SS effectors such as IcsB helps *Shigella* escape its dissemination vacuole, thus facilitating spreading to neighbouring cells (Ogawa et al., 2003, Campbell-Valois et al., 2015, and Liu et al., 2018). It was recently proven that IcsB acts as an acyltransferase, and this is how it contributes to vacuole escape of Shigella (Liu et al., 2018).

Furthermore, the paper published by (Liu et al., 2018) demonstrated that CHMP5, one of the components of the endosomal sorting complex required for transport III (ESCRT-III), is acylated by IcsB, an event which seems crucial for *Shigella* to escape autophagy. They also demonstrated that IcsB acylates lysin residues found in the C-terminus of the polybasic region (PBR) of GTPases proteins, although the role of this event in the pathogenesis of *Shigella* has not been confirmed. Interestingly though, the RID of MARTX toxins, which possess <20% sequence identity with IcsB are also acyltransferase that modify the C-terminal lysine residues of the PBR of Rho-GTPases. The crystal structure of RID was resolved and the catalytic domain is located in the C-terminal lobe and it was found to adopt inverted papain-like fold (Zhou et al., 2017). The MARTX toxins and IcsB were predicted previously in 2009 by Pei and Grishin to adopt inverted papain-like fold and it was suggested that they may act as protease or acyltransferase enzymes.

The sequence alignment revealed that IcsB and MARTX toxins share conserved catalytic histidine and cysteine but not aspartate residues (Pei & Grishin, 2009). However, there was no evidence of other proteins being identified possessing similar activity as IcsB and MARTX toxins. We wondered if there are homolog proteins present in other bacteria that may have related enzymatic activity. Therefore, the work conducted in this project revealed a number of interesting results that are discussed in this Chapter.

First, using bioinformatics approaches, 11 homologs of IcsB have been identified. Their sequence identities with IcsB ranged between 22% to 53%. Furthermore, when the full protein sequences were aligned, the catalytic triad His-145, Asp-195 and Cys-306 (Pei & Grishin, 2009, and Liu et al., 2018) was found to be fully conserved across the alignment. We were successfully able to reproduce equivalent results to what is presented in Liu et al. paper for catalytic mutations, and we were able to test homologous proteins' toxicity using the same assay. As a result, 7 out of the 11 proteins inhibited yeast cell growth. This observation confirmed that the inhibition of yeast growth can be used for identifying the enzymatic activity of putative effectors, as discussed previously (Slagowski et al., 2008). Yeast *Saccharomyces cerevisiae* represent the simplest eukaryotes with many cellular processes similar to human. It is widely used as model to understand the functional characterizations of type III secretion effector proteins in eukaryotic. (Popa et al., 2016).

In agreement with Slagowski's observation, we discovered that in the toxic proteins, the substitution of histidine 145 with alanine revealed the toxicity phenotype observed in their wild type strains. Therefore, the next step would be mutating aspartate 195 with alanine to test

whether or not the mutation will diminish the toxicity phenotype regarding its catalytic activity. We expect the result to yield the same results as for the histidine mutation.

Furthermore, the reason we excluded cysteine 306 from being tested is that cysteine 306 needs the chaperone IpgA to be co-expressed to relieve its toxicity since both substitutions with alanine and serine produced the same results. Also, the uncertainty that IpgA would be able to rescue non-cognate cargos and the complexity in finding equivalent chaperones for each one of the toxic homologs. In addition to His-145 mutation result, and until Asp-195 mutations effect is confirmed, it would mean that the killing phenotype observed in these proteins is linked to their catalytic activity and that they possess the same catalytic triad as IcsB. Thus, these proteins could be forming an IcsB-like acyltransferase family.

Moreover, the protein sequence was divided into three main domains represented in the N-terminal, the catalytic, and the C-terminal domain. In addition to the sequence similarities found in the catalytic domain, the homology happened to be extended to 41 amino acids after the serine 319. As for the lack of conservation in the N-terminus between homologs and IcsB and the fact that these proteins are expressed from different bacteria, it is possible that the N-terminus might play a role in its protein localization in the host cell. Besides, the N-terminus usually serves a secretion signal for type III secretion proteins. The absence of the N-terminal sequence makes the secretion of some effector proteins ineffectual (Wang et al., 2013). However, the two conserved residues N305 and N352 in IcsB were substituted with alanine, and the results for the former were unexpected since it showed partial toxicity with and without co-expression of the chaperone IpgA. The only justification we could think of is that the N305 mutant might be masking the region where the chaperone IpgA binds or disrupting the binding of the chaperone to the protein. Therefore, this prevents its stabilization to some extent since the region where

IpgA binds to IcsB protein was suggested to be between 171 to 247 residues (Ogawa et al., 2003). The binding of a chaperon to the region in type III secretion effector's catalytic domain is atypical though and it contradicts with what has been reported for the binding of the homologous chaperon SicP to its cognate cargo SptP (Stebbins & Galán, 2001). As for N352, the alanine mutant appeared to be toxic with or without co-expression with chaperone IpgA.

Most of these homolog proteins that we have identified originate from a wide diversity of bacteria, many of which are not well-studied. They all are classified as Gram-negative bacteria belonging to different classes of the proteobacteria phylum. Some of these bacteria found to cause infection to animals or plants as the source of isolation revealed, and that was discussed in detail in Chapter II. Despite the information found for these bacteria, we used the *in-silico* method (TXSScan) to explore the possibilities of type III secretion system present in these bacteria. The results revealed that 9 out of the 11 bacteria might have one or more functional T3SS, and the presence of the N-terminal encoded for the secretion signal and chaperone binding, increased their chances of being potential T3SS substrates. Therefore, this observation builds up the probabilities of these homologous proteins originating from pathogenic bacteria. The exception to this observation was the last two homologs H10 and H11 that lacked the N-terminus domain, and there was no evidence for T3SS using the TXSScan prediction method. Thus, their potential of being human pathogenic bacteria decreased, although we cannot discard that they might express another type of secretion system. In fact, the manual search using BLAST against these two bacterial genomes revealed homology with some of the type III secretion components with less than 50% sequence identity even though some of these components were used in the TXSScan method. Given their low homology, this might be the reason why they failed to be predicted by TXSScan.

Taken together, the literature about the bacterial species expressing the homologous protein of IcsB and the data presented in my thesis suggest that these proteins are likely to be structurally related, and that the subset that is cytotoxic might even be endowed with a similar enzymatic activity. Besides, some of them have already be implicated in clinical cases or shown to infect other organism (e.g. *Sodalis praecaptivus*, *Robbsia andropogonis*, respectively) or have such high similarity with Shigella (e.g. *E. marmotae*) that they might have a higher probability of being pathogenic. Moreover, the acylation assay was successfully established in our laboratory, and the results (unpublished) revealed that 4 out of the seven toxic proteins showed acylation patterns comparable with IcsB in an in-gel fluorescence assay using a 16-carbon fatty acid, suggesting that these proteins might possess acyltransferase activity. Furthermore, the three homolog proteins H2, H5 and H8, expressed from *E. marmotae*, *S. praecaptivus* and *R. andropogonis* respectively, have shown toxicity phenotype when they are expressed in yeast *saccharomyces cerevisiae,* in addition to their ability to either invade, act as a commensal, or cause the disease, it would be interesting to test if they have the same substrates as IcsB. Furthermore, it would be interesting to test whether or not these toxic acyltransferase proteins have the ability to complement *IcsB* defect during infection of human cells (Liu et al., 2018). As for the bacteria that have already shown to invade human cells, it would be interesting to test whether LC3 is also recruited in their vicinity and if the corresponding IcsB homologs might counteract this. In the case of the *E. marmotae* strain, where they used epithelial cell invasion assay to infect human epithelial cells with. They successfully were able to show that *E. marmotae* is invasive but lacks some pathogenic factors which make it less invasive than *Shigella* (Liu et al., 2019). This study would thus help conducting further experiments. It is obvious that similar strategies might also be applied to less studied species such *Achromobacter*

arsenitoxydans and *Chromobacterium* sp. LK11 at the condition of obtaining the corresponding samples.

However, looking deeper into the sequence alignment of the toxic and non-toxic proteins revealed more conservation when each category was aligned separately. Also, the conservation of the amino acids observed mainly in the catalytic domain. Apart from that, the three closely related proteins in the group H4, H6, and H9 have shown more conservation among their sequence alignment, noting that H4 is toxic while H6 and H9 are not. In order to link the relationship between the three proteins, one suggestion that might be interesting is to generate a set of mutations in the proteins H6 and H9 that might turn them to be toxic to yeast. The resulted phenotype might be an indication that they have enzymatic activity. Subsequently, if H6 and H9 become toxic, the acylation assay would then be performed. The reason behind conducting this experiment is that we are trying to figure out if the three proteins are divergent from the same ancestor regarding the extremely high conservations among their sequences. If that was the case, these three proteins could also be acyltransferase, and during evolution, the substitutions of their amino acid composition have affected and changed their nature.

Furthermore, for the next step, it would be interesting to complement a strain of *Shigella* lacking IcsB with the homologous proteins that were able to acylate in order to test their ability to replace the role of IcsB in escaping autophagy. Since IcsB targets CHMP5 and Rho-GTPases, it would be interesting to see if the homologous proteins target the same substrate. An example of such experiments would be a pull-down assay to identify any binding between proteins, followed by mass spectrometry analysis that will reveal any modifications made by the homologous proteins on proteins found in host cells. Similar to the experiments conducted for IcsB towards finding its target in mammalian cells (Liu et al., 2018). Another way to test that

would be by performing the plaque assay to investigate the virulence activity of *Shigella* complemented strain with the toxic IcsB homologs. In the case where the homolog proteins are behaving similar to IcsB, there should be plaque formation as a result of invading, replicating intracellularly and spreading from cell-to-cell (Campbell-Valois et al., 2015).

To conclude, there are several more experiments to be performed in order to confirm the capabilities of the homologous proteins in acylating fatty acids molecules that are involved in the structural components of a cell. Also, additional experiments are a prerequisite to confirm the pathogenic potential of bacteria expressing these homologs.

References

Abby, S. S., Néron, B., Ménager, H., Touchon, M., & Rocha, E. P. (2014). MacSyFinder: A Program to Mine Genomes for Molecular Systems with an Application to CRISPR-Cas Systems. *PLoS ONE, 9*(10), e110726. doi: 10.1371/journal.pone.0110726

Arena, E. T., Campbell-Valois, F., Tinevez, J., Nigro, G., Sachse, M., Moya-Nilges, M., Nothelfer, K., Marteyn, B., Shorte, S. L., & Sansonetti, P. J. (2015). Bioimage analysis of *Shigella* infection reveals targeting of colonic crypts. *Proceedings of the National Academy of Sciences*, *112*(25), E3282-E3290. https://doi.org/10.1073/pnas.1509091112

Bhagwat M, Aravind L. PSI-BLAST Tutorial. In: Bergman NH, editor. Comparative Genomics: Volumes 1 and 2. Totowa (NJ): Humana Press; 2007. Chapter 10. Available from: https://www.ncbi.nlm.nih.gov/books/NBK2590/

Boardman, B. K., & Fullner Satchell, K. J. (2004). Vibrio cholerae Strains with Mutations in an Atypical Type I Secretion System Accumulate RTX Toxin Intracellularly. *Journal of Bacteriology*, *186*(23), 8137-8143. doi:10.1128/jb.186.23.8137-8143.2004

Buchrieser, C., Glaser, P., Rusniok, C., Nedjari, H., D'Hauteville, H., Kunst, F., Sansonetti, P., & Parsot, C. (2000). The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella* flexneri. *Molecular Microbiology*, *38*(4), 760-771. https://doi.org/10.1046/j.1365-2958.2000.02179.x

Campbell-Valois, F., & Pontier, S. M. (2016). Implications of spatiotemporal regulation of *Shigella flexneri* type three secretion activity on effector functions: Think globally, act locally. *Frontiers in Cellular and Infection Microbiology*, *6*. https://doi.org/10.3389/fcimb.2016.00028

Campbell-Valois, F., Sachse, M., Sansonetti, P. J., & Parsot, C. (2015). Escape of actively secreting *Shigella flexneri* from ATG8/LC3-positive vacuoles formed during cell-to-cell spread is facilitated by IcsB and VirA. *mBio*, *6*(3). https://doi.org/10.1128/mbio.02567-14

Campbell-Valois, F., Schnupf, P., Nigro, G., Sachse, M., Sansonetti, P., & Parsot, C. (2014). A fluorescent reporter reveals on/Off regulation of the *Shigella* type III secretion apparatus during entry and cell-to-cell spread. *Cell Host & Microbe*, *15*(2), 177-189. https://doi.org/10.1016/j.chom.2014.01.005

CDC. (2019). *Drug-resistant Shigella*. Centers for Disease Control and Prevention. https://www.cdc.gov/drugresistance/pdf/threats-report/shigella-508.pdf

CDC. (2019). *Shigellosis - Chapter 4 - 2020 yellow book | Travelers' health | CDC*. Centers for Disease Control and Prevention. https://wwwnc.cdc.gov/travel/yellowbook/2020/travel-related-infectious-diseases/shigellosis

Deng, W., Marshall, N. C., Rowland, J. L., McCoy, J. M., Worrall, L. J., Santos, A. S., Strynadka, N. C., & Finlay, B. B. (2017). Assembly, structure, function and regulation of type III secretion systems. *Nature Reviews Microbiology*, *15*(6), 323-337. https://doi.org/10.1038/nrmicro.2017.20

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J., Guindon, S., Lefort, V., Lescot, M., Claverie, J., & Gascuel, O. (2008). Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, *36*(Web Server), W465-W469. https://doi.org/10.1093/nar/gkn180

Dong, N., Zhu, Y., Lu, Q., Hu, L., Zheng, Y., & Shao, F. (2012). Structurally distinct bacterial TBC-like GAPs link Arf GTPase to Rab1 inactivation to counteract host defenses. *Cell*, *150*(5), 1029-1041. https://doi.org/10.1016/j.cell.2012.06.050

Durand, J. M., Dagberg, B., Uhlin, B. E., & Bjork, G. R. (2000). Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of *Shigella flexneri*: The expression of the virF gene. *Molecular Microbiology*, *35*(4), 924-935. https://doi.org/10.1046/j.1365-2958.2000.01767.x

Conner, C., Khan, A., Bocsanczy, A. M., Spakes-Richter, B., & Norman, D. J. (2019). First Report of Bacterial Spot Caused by Robbsia andropogonis in Loropetalum chinense. *Plant Disease*, *103*(9), 2467-2467. doi:10.1094/pdis-11-18-1929-pdn

Cother, E. J., Noble, D., Peters, B. J., Albiston, A., & Ash, G. J. (2004). A new bacterial disease of jojoba (Simmondsia chinensis) caused by Burkholderia andropogonis. *Plant Pathology*, *53*(2), 129-135. doi: 10.1111/j.0032-0862.2004.00982.x

Cullinane, M., Gong, L., Li, X., Adler, N., Tra, T., Wolvetang, E., … Adler, B. (2008). Stimulation of autophagy suppresses the intracellular survival ofBurkholderia pseudomalleiin mammalian cell lines. *Autophagy*, *4*(6), 744-753. doi:10.4161/auto.6246

Dale, C., Young, S. A., Haydon, D. T., & Welburn, S. C. (2001). The insect endosymbiont Sodalis glossinidius utilizes a type III secretion system for cell invasion. *Proceedings of the National Academy of Sciences*, *98*(4), 1883-1888. doi:10.1073/pnas.98.4.1883

Deng, W., Marshall, N. C., Rowland, J. L., McCoy, J. M., Worrall, L. J., Santos, A. S., … Finlay, B. B. (2017). Assembly, structure, function and regulation of type III secretion systems. *Nature Reviews Microbiology*, *15*(6), 323-337. doi:10.1038/nrmicro.2017.20

Ding, J., Shiu, J., Chen, W., Chiang, Y., & Tang, S. (2016). Genomic Insight into the Host–Endosymbiont Relationship of Endozoicomonas montiporae CL-33T with its Coral Host. *Frontiers in Microbiology*, *7*. doi:10.3389/fmicb.2016.00251

Dong, X., Lu, X., & Zhang, Z. (2015). BEAN 2.0: an integrated web resource for the identification and functional analysis of type III secreted effectors. Database, 2015, bav064. doi:10.1093/database/bav064

Eichinger, V., Nussbaumer, T., Platzer, A., Jehl, M., Arnold, R., & Rattei, T. (2015). EffectiveDB—updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. *Nucleic Acids Research, 44*(D1), D669-D674. doi:10.1093/nar/gkv1269

Enomoto, S., Chari, A., Clayton, A. L., & Dale, C. (2017). Quorum Sensing Attenuates Virulence in Sodalis praecaptivus. *Cell Host & Microbe*, *21*(5), 629-636.e5. doi: 10.1016/j.chom.2017.04.003

Fantinatti-Garboggini F, Almeida R. d, Portillo V. d A, Barbosa TAP, Trevilato PB, Neto CER, Coêlho RD, Silva DW, Bartoleti LA, Hanna ES, Brocchi M, Manfio GP. 2004. Drug resistance in *Chromobacterium violaceum. Genet Mol Res* 3:134–147.

Fu, Y., & Galán, J. E. (1998). Identification of a specific chaperone for SptP, a substrate of the Centisome 63 type III secretion system ofSalmonella typhimurium. *Journal of Bacteriology*, *180*(13), 3393-3399. https://doi.org/10.1128/jb.180.13.3393-3399.1998

Goldberg, T., Rost, B., & Bromberg, Y. (2016). Computational prediction shines light on type III secretion origins. *Scientific Reports*, 6(1). doi:10.1038/srep34516

Gong, L., Cullinane, M., Treerat, P., Ramm, G., Prescott, M., Adler, B., … Devenish, R. J. (2011). The Burkholderia pseudomallei Type III Secretion System and BopA Are Required for Evasion of LC3-Associated Phagocytosis. *PLoS ONE, 6*(3), e17852. doi: 10.1371/journal.pone.0017852

Hu, B., Morado, D. R., Margolin, W., Rohde, J. R., Arizmendi, O., Picking, W. L., Picking, W. D., & Liu, J. (2015). Visualization of the type III secretion sorting platform of *Shigella flexneri*. *Proceedings of the National Academy of Sciences*, *112*(4), 1047-1052. https://doi.org/10.1073/pnas.1411610112

Kotloff, K. L., Winickoff, J. P., Ivanoff, B., Clemens, J. D., Swerdlow, D. L., Sansonetti, P. J., et al. (1999). Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. *Bull. World Health Organ.* 77, 651–666

Kushnirov, V. V. (2000). Rapid and reliable protein extraction from yeast. *Yeast*, *16*(9), 857-860. doi:10.1002/1097-0061(20000630)16:9<857::aid-yea561>3.0.co;2-b

Li, P., Jiang, W., Yu, Q., Liu, W., Zhou, P., Li, J., Xu, J., Xu, B., Wang, F., & Shao, F. (2017). Ubiquitination and degradation of GBPs by a *Shigella* effector to suppress host defence. *Nature*, *551*(7680), 378-383. https://doi.org/10.1038/nature24467

Li, X., Hu, Y., Gong, J., Lin, Y., Johnstone, L., Rensing, C., & Wang, G. (2012). Genome Sequence of the Highly Efficient Arsenite-Oxidizing Bacterium Achromobacter arsenitoxydans SY8. *Journal of Bacteriology*, *194*(5), 1243-1244. doi:10.1128/jb.06667-11

Li, X., Hu, Y., Gong, J., Zhang, L., & Wang, G. (2013). Comparative genome characterization of Achromobacter members reveals potential genetic determinants facilitating the adaptation to a pathogenic lifestyle. *Applied Microbiology and Biotechnology*, *97*(14), 6413-6425. doi:10.1007/s00253-013-5018-3

Liu, S., Feng, J., Pu, J., Xu, X., Lu, S., Yang, J., … Xu, J. (2019). Genomic and molecular characterisation of Escherichia marmotae from wild rodents in Qinghai-Tibet plateau as a potential pathogen. *Scientific Reports*, *9*(1). doi:10.1038/s41598-019-46831-3

Liu, W., Zhou, Y., Peng, T., Zhou, P., Ding, X., Li, Z.; Zhong, H.; Xu, Y.; Chen, S.; Hang, H. C.; Shao, F. (2018). Nε-fatty acylation of multiple membrane-associated proteins by *Shigella* IcsB effector to modulate host function. *Nature Microbiology, 3(9)*, 996-1009. doi:10.1038/s41564-018-0215-6

Lopes-Santos, L., Castro, D. B. A., Ferreira-Tonin, M., Corrêa, D. B. A., Weir, B. S., Park, D., et al. (2017). Reassessment of the taxonomic position of *Burkholderia andropogonis* and description of *Robbsia andropogonis* gen. nov., comb. nov. *Antonie van Leeuwenhoek* 110, 727–736. doi: 10.1007/s10482-017-0842-6

Ogawa, M., Suzuki, T., Tatsuno, I., Abe, H., & Sasakawa, C. (2003). IcsB, secreted via the type III secretion system, is chaperoned by IpgA and required at the post-invasion stage of *Shigella* pathogenicity. *Molecular Microbiology*, *48*(4), 913-931. https://doi.org/10.1046/j.1365-2958.2003.03489.x

Maltz, M. A., Weiss, B. L., O'Neill, M., Wu, Y., & Aksoy, S. (2012). OmpA-Mediated Biofilm Formation Is Essential for the Commensal Bacterium Sodalis glossinidius To Colonize the Tsetse Fly Gut. *Applied and Environmental Microbiology*, *78*(21), 7760-7768. doi:10.1128/aem.01858-12

Mavris, M., Page, A., Tournebize, R., Demers, B., Sansonetti, P., & Parsot, C. (2002). Regulation of transcription by the activity of the *Shigella flexneri* type III secretion apparatus. *Molecular Microbiology*, *43*(6), 1543-1553. https://doi.org/10.1046/j.1365-2958.2002.02836.x

Maurelli, A. T., Baudry, B., D'Hauteville, H., Hale, T. L., & Sansonetti, P. J. (1985). Cloning of plasmid DNA sequences involved in invasion of Hela cells by *Shigella* flexneri. *Infection and Immunity*, *49*(1), 164-171. https://doi.org/10.1128/iai.49.1.164-171.1985

Maurelli, A. T., & Sansonetti, P. J. (1988). Identification of a chromosomal gene controlling temperature-regulated expression of *Shigella* virulence. *Proceedings of the National Academy of Sciences*, *85*(8), 2820-2824. https://doi.org/10.1073/pnas.85.8.2820

Miki, T., Iguchi, M., Akiba, K., Hosono, M., Sobue, T., Danbara, H., & Okada, N. (2010). Chromobacterium pathogenicity island 1 type III secretion system is a major virulence determinant for Chromobacterium violaceum-induced cell death in hepatocytes. *Molecular Microbiology*, no-no. doi:10.1111/j.1365-2958.2010.07248.x

Pei, J., & Grishin, N. V. (2009). The Rho GTPase inactivation domain inVibrio choleraeMARTX toxin has a circularly permuted papain-like thiol protease fold. *Proteins: Structure, Function, and Bioinformatics*, *77*(2), 413-419. doi:10.1002/prot.22447

Popa, C., Coll, N. S., Valls, M., & Sessa, G. (2016). Yeast as a heterologous model system to uncover type III effector function. *PLOS Pathogens*, *12*(2), e1005360. https://doi.org/10.1371/journal.ppat.1005360

Sansonetti, P. J., Hale, T. L., Dammin, G. J., Kapfer, C., Collins, H. H., & Formal, S. B. (1983). Alterations in the pathogenicity of escherichia coli K-12 after transfer of plasmid and chromosomal genes from *Shigella flexneri*. *Infection and Immunity*, *39*(3), 1392-1402. https://doi.org/10.1128/iai.39.3.1392-1402.1983

Santos-Garcia, D., Silva, F. J., Morin, S., Dettner, K., & Kuechler, S. M. (2017). The All-Rounder Sodalis: A New Bacteriome-Associated Endosymbiont of the Lygaeoid Bug Henestaris halophilus (Heteroptera: Henestarinae) and a Critical Examination of Its Evolution. *Genome Biology and Evolution*, *9*(10), 2893-2910. doi:10.1093/gbe/evx202

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671-675. doi:10.1038/nmeth.2089

Seidman, C. E., Struhl, K., Sheen, J., & Jessen, T. (2001). Introduction of plasmid DNA into cells. In *Current protocols in molecular biology*. DOI: 10.1002/0471142727.mb0108s37

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., … Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(1), 539. doi:10.1038/msb.2011.75

Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics. 1989; 122(1):19–27. PMID: 2659436

Slagowski, N. L., Kramer, R. W., Morrison, M. F., LaBaer, J., & Lesser, C. F. (2008). A functional Genomic yeast screen to identify pathogenic bacterial proteins. *PLoS Pathogens*, *4*(1), e9. https://doi.org/10.1371/journal.ppat.0040009

Sorokin, D. Y., Tourova, T. P., Henstra, A. M., Stams, A. J., Galinski, E. A., & Muyzer, G. (2008). Sulfidogenesis under extremely haloalkaline conditions by Desulfonatronospira thiodismutans gen. nov., sp. nov., and Desulfonatronospira delicata sp. nov. - a novel lineage of Deltaproteobacteria from hypersaline soda lakes. *Microbiology*, *154*(5), 1444-1453. doi:10.1099/mic.0.2007/015628-0

Sorokin, D. Y., Tourova, T. P., Kolganova, T. V., Detkova, E. N., Galinski, E. A., & Muyzer, G. (2011). Culturable diversity of lithotrophic haloalkaliphilic sulfate-reducing bacteria in soda lakes and the description of Desulfonatronum thioautotrophicum sp. nov., Desulfonatronum thiosulfatophilum sp. nov., Desulfonatronovibrio thiodismutans

sp. nov., and Desulfonatronovibrio magnus sp. nov. *Extremophiles, 15*(3), 391-401. doi:10.1007/s00792-011-0370-7

Stebbins, C. E., & Galán, J. E. (2001). Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. *Nature, 414*(6859), 77-81. https://doi.org/10.1038/35102073

Swenson, C. E., & Sadikot, R. T. (2015). AchromobacterRespiratory Infections. *Annals of the American Thoracic Society, 12*(2), 252-258. doi:10.1513/annalsats.201406-288fr

Vander Broek, C. W., & Stevens, J. M. (2017). Type III secretion in the Melioidosis pathogen Burkholderia pseudomallei. *Frontiers in Cellular and Infection Microbiology, 7*. https://doi.org/10.3389/fcimb.2017.00255

Wang, Y., Sun, M., Bao, H., Zhang, Q., & Guo, D. (2013). Effective identification of bacterial type III secretion signals using joint element features. *PLoS ONE, 8*(4), e59754. https://doi.org/10.1371/journal.pone.0059754

Wandel, M. P., Pathe, C., Werner, E. I., Ellison, C. J., Boyle, K. B., Von der Malsburg, A., Rohde, J., & Randow, F. (2017). GBPs inhibit motility of *Shigella flexneri* but are targeted for degradation by the bacterial Ubiquitin ligase IpaH9.8. *Cell Host & Microbe, 22*(4), 507-518.e5. https://doi.org/10.1016/j.chom.2017.09.007

WHO. (2017). *Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. World Health Organization.* https://www.who.int/medicines/publications/global-priority-list-antibiotic-resistant-bacteria/en/

WHO. (2018). *The top 10 causes of death*. WHO | World Health Organization. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

Yong, E. (2017, May 10). The Making of a Friendly Microbe. Retrieved from https://www.theatlantic.com/science/archive/2017/05/a-friendly-microbe/526152/

Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., … Oda, K. (2016). A bacterium that degrades and assimilates poly (ethylene terephthalate). *Science*, *351*(6278), 1196-1199. doi:10.1126/science. aad6359

Zhou, Y., Huang, C., Yin, L., Wan, M., Wang, X., Li, L., … Zhu, Y. (2017). Nε-Fatty acylation of Rho GTPases by a MARTX toxin effector. *Science*, *358*(6362), 528-531. doi:10.1126/science.aam8659

Appendix

Table A1. List of plasmids

Plasmid	Origin
pRS313	(Slagowski et al., 2008)
pRS313 GAL1p	WB-HMI
pRS313 GAL1p1	WB-HMI
pRS313 GAL1p2	WB-HMI
pRS313 GAL1p2 IcsB	WB-HMI
pRS313 GAL1p2 IcsB d80	WB-HMI
pRS313 GAL1p2 IcsB H145A	WB-HMI
pRS313 GAL1p2 IcsB D195A	WB-HMI
pRS313 GAL1p2 IcsB C306A	WB-HMI
pRS313 GAL1p2 IcsB N305A	WB-HMI
pRS313 GAL1p2 IcsB C306S	WB-HMI
pRS313 GAL1p2 IcsB N352A	WB-HMI
pUC57_IcsB_h2	Gene synthesis (Bio Basic)
pUC57_IcsB_h3	Gene synthesis (Bio Basic)
pUC57_IcsB_h4	Gene synthesis (Bio Basic)
pUC57_IcsB_h5	Gene synthesis (Bio Basic)
pUC57_IcsB_h6	Gene synthesis (Bio Basic)
pUC57_IcsB_h7	Gene synthesis (Bio Basic)
pUC57_IcsB_h8	Gene synthesis (Bio Basic)
pUC57_IcsB_h9	Gene synthesis (Bio Basic)
pUC57_IcsB_h10	Gene synthesis (Bio Basic)
pUC57_IcsB_h11	Gene synthesis (Bio Basic)
pRS313 GAL1p2 H1	WB-HMI
pRS313 GAL1p2 H2	WB-HMI

pRS313 GAL1p2 H3	WB-HMI
pRS313 GAL1p2 H4	WB-HMI
pRS313 GAL1p2 H5	WB-HMI
pRS313 GAL1p2 H6	WB-HMI
pRS313 GAL1p2 H7	WB-HMI
pRS313 GAL1p2 H8	WB-HMI
pRS313 GAL1p2 H9	WB-HMI
pRS313 GAL1p2 H10	WB-HMI
pRS313 GAL1p2 H11	WB-HMI
pRS316	(Slagowski et al., 2008)
pRS316 GAL1p1	WB-HMI
pRS316 GAL1p1 IpgA	WB-HMI

Table A2. List of antibodies used

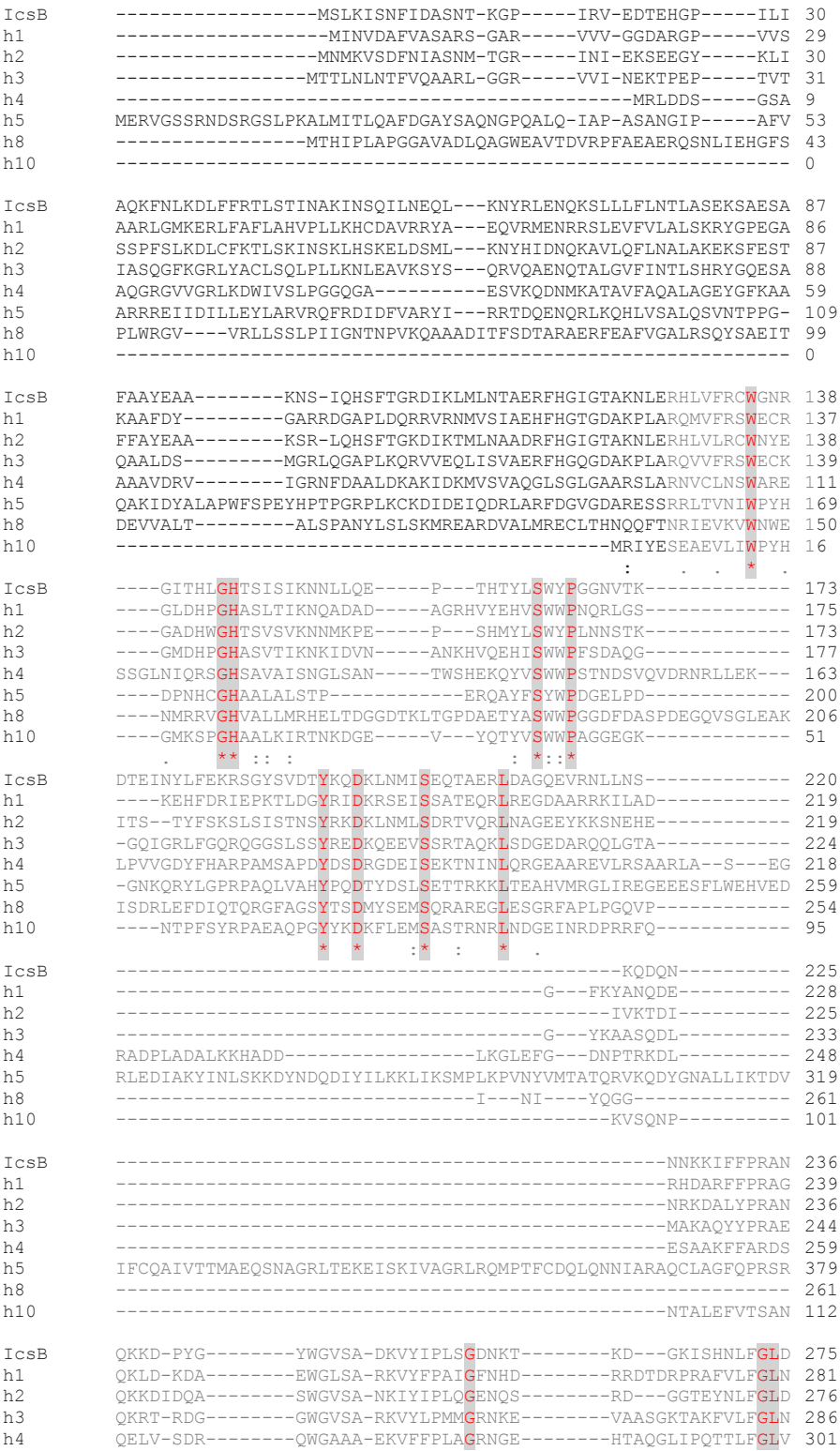
Antibody	Reference number	Supplier
Anti-Flag	F3165	Sigma
Anti-Myc	CLANT211	Cedrlane
Anti-Myc	A00704	Genscript

Table A3. Primers' sequences that used for either mutagenesis PCR, colony screening PCR, or sequencing.

Primer	Sequence 5′- 3′
Gal1p_Sp6_S	ACACTATAGTAAGTGACGCTGACTAAC
Gal1p_Sp6_R	CACCTAAATAGAGGTATATTAACAATTTTT
Gal1p_BGH_S	TGCCTTCTATCATGTAATTAGTTATGTCAC
Gal1p_BGH_R	CAGTCGAGGTCGGCCCGTGGCTCAGCTGC
Gal1p2_3xFlagBGH_F	AAGGAGGCCGAATTCC
Gal1p2_3xFlagBGH_R	GTAAGCGTGACATAACTAATTACA
IpgA_S	AGAGgaattcgctaggagaaattaaccatgtgtcgcaaactatatgataaactttatg
IpgA_R	AGAGAGggatccgttcacttctgaagtgatgtttgc
BopA_S	AGAGgaattcgctaggagaaattaaccATGATTAATGTTGATGCTTTTGTTGCATC
BopA_R	AGAGAGggatccTTAATCAAGCTGCAATTGCAGCTGCC
IcsB80_S	ATGgaattcgctaggagaaattaaccgaaaaatctgcagaaagtgcgtt
IcsB_Bam_R	AgagagggatccttagCTATATATTAGAATGAGAGTTATTC
BopA80_S	ATGgaattcgctaggagaaattaaccTATGGTCCAGAAGGTGCTAAAGC
IcsB_C306A_S	GCCgctggtatggcacttaatgttc
IcsB_305_R	Gttttcagatttactgattaatttata
IcsB_H145A	GCCactagtATCAGTATAAAAAACAATTTACTACA
IcsB_144_R	CCCTAAATGGGTTATGCCTC
IcsB_D195A_S	GCCAAACTAAATATGATTTCAGAGCAAACAGC
IcsB_D194_R	TTGCTTATAGGTATCGACACTGTAAC
IcsB_N305A_S	GCCTGTGCTGGTATGGCACTTAATG
IcsB_304_R	TTCAGATTTACTGATTAATTTataATTTGCC
IcsB_N352A_S	GCCCAGTCTTATAACGAAATAATGAAGTAT
IcsB_351_R	GagACTTTCAATGCGTTGCC
IcsB_C306S_S	TCTgctggtatggcacttaatgttc
H1_144_S	GCCGCTTCATTAACTATTAAAAATCAA
H1_143_R	ACCTGGATGATCCAAAC
H2_H145A_S	GCCACTTCCGTTAGCGTGAAAAACA
H2_144_R	GCCCCAATGGTCCGCCCCC
H3_H146A_S	GCCGCTTCCGTCACAATTAAGAATAA
H3_145_R	GCCCGGATGGTCCATC
H4_H122_S	GCCTCAGCCGTCGCGATTTC
H4_121_R	TCCTGAACGCTGAATGTT
H5_H176A_S	GCCGCCGCACTTGCACTGAGTACC
H5_175_R	TCCACAGTGATTGGGATCGTG
H8_H157A_S	GCCGTAGCATTGCTGATGAGGC
H8_156_R	TCCGACACGTCTCATATT
H10_H23A_S	GCCGCTGCATTGAAGATCAGGACGAA
H10_22_R	CCCCGGGGACTTCATTCCA

Figure A1. Sequence alignment of toxic homologous proteins

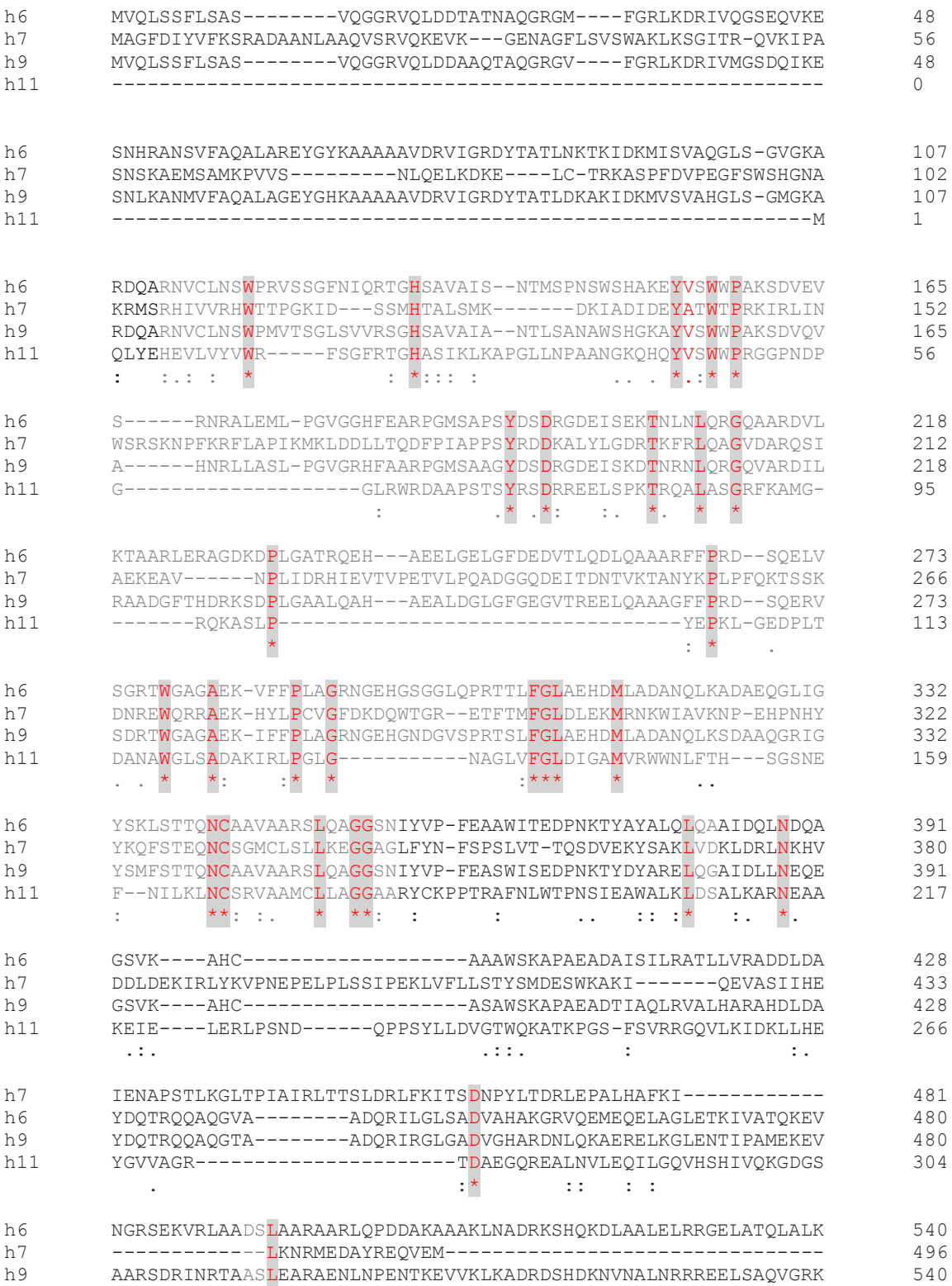
Fully conserved residues



```
h5        QVWDEEND--------AFVATS-DQIYLPIAGER-------------------AELFGLD 411
h8        -EDEAPNV--------SWGKEPDASIAMPMAGYNERPHPGGTGTDANGVVVPTLTTFGLS 312
h10       KKFNLPGLFSRVGRTRKSGAG--IDGFDNVAGRS-------------SLADGEYTYWGLN 157
                                .      *                        :**

IcsB      ETNMSKFICKKKADAFRQLANYKLISKSENCAGMALNVLKAGNSEIYFPLPDVKLVATPN 335
h1        EAAMLRDARTVKEGAKSGELMYQMISKKENCASMALRVLRAGGAEHFVPYTAAWISEDPN 341
h2        ESKLSSFICKKRGGAFRRTEHYKLISKSQNCAGMALDVLKAGDAETYVPFPKIKLVANPN 336
h3        EKAILHEARHVKQEAAEGRIGYTLASTEENCASMALRMLRAGGAENFVPFSASWVTEDPN 346
h4        EHDMLADARQLKDDAAHGRIGYSMFSTTQNCAAIAARSLQAGGSNIYVPFEASWITEDPN 361
h5        EQALEEFWLSVKSEIATGKARFRLVSTRHNCSAMALRSLQAAGAGRYVPLPARWLLYTPR 471
h8        VDAIKTYWETE--AIGNDAAQFKLVSTDRNCSGMAARLLKAGGAHAFLPVPEAVLFLDPN 370
h10       LPGMVKWWMDF----SDTPKGYSFVSTKYNCAAVVAQCLLEGCADAYIDIKQPKMYYTPN 213
             :                 : : *.  **:.:.   *  . :  :.     :   *.

IcsB      DVYAYANKVRQRIESLNQSYNEIMKYIESDFDLSRLT---------QLRRSYLKSFNKIN 386
h1        HAHAYALAVQARIDALNQRRADVERRCERLRDSASVRQ---------AWRAFSEAGGASA 392
h2        DAYKYAYKVMARIEELNKIHKGIIKNVDEKIDIMTMI---------KLRRSYLNSFNKVT 387
h3        KAHAYAERLQARMDSLNRQVADIGERCSGMLQQPGVKT---------EWDAFLNQGRMGE 397
h4        KTYDYARQLQGAIDGLNGKADGIKKFCDTAWQSAPADAD----AIRQLRDALHETANDQA 417
h5        DVEHYTLRLQKALDRLNRQSSAIQAFYEQETASRPNDLAIRQSDLTGLRTQFNLLFSHLQ 531
h8        TMRRYTTDLMTEVTALNAKADTVDASFGRLFTEDEGAD------ESGVTPSYSVR----- 419
h10       DISALADKLVNRLDDLNIESRRLINNSRQSLDRPGIWTS----------EEFYKASYAGR 263
               :  :   :  **     :

IcsB      LIH------TPKTFK---PLSIS--LYKHPT---------------------------- 406
h1        SPL-----------A------E-----DAGR-----GR---ASAHMRQAR--LDEHAREV 420
h2        SSH------KINKFN---TLLVI--RNKDNI---------------------------- 407
h3        LKR------EISQLK------QQ--INKERR-----PE---AQAVQRQAL--KERNAQQV 433
h4        IFA------QAKQLA------LG--IAADQRIDSLRAEVADRSGKMDQARSLLQPLDRQI 463
h5        TPQQKRLRRLARAVNTLPDAFLS--QQSDRII--------------PAAKALVEALSPLL 575
h8        ------LARAQEAIASLPRATQASMQALNNALNECKTEAGDYEGMMRQMKAMVGGLHNHV 473
h10       FAR------RKGQVASIDDLLKE--YHKAGPWDE---------NNFPQK---FKALGRMM 303


IcsB      -----------ENVSSEDFDAV----INACHSYLVKSAPSNMTRVLNELKTEATDKKEEI 451
h1        E----RIGAYFAELSAGRSGKHRDRADAALADAMKRCAPSARD-D----VAALTRKASVL 471
h2        -----------ENASLDDYNKI----INSCFESLKKLAPHNSHSRISEFKPTMTDTKEQL 452
h3        F----RIGAHFEQLAVGRAD------ALALAKAMREGAPSIDD-D----AAVLTRKTAGL 478
h4        R---------ATEKDVGDRQKAVDNKQK-AVDGMSRSAKSSSPHA----AARLADSKAEL 509
h5        TDAISQQQDYRTTLSLAAAMRT------WLQQRQQRGVPSSALYS----AASA------- 618
h8        NGPFGS-----------DINKALRSAY-ELY-SAIRGAA---------------------- 499
h10       R----NVLDHRQKKAGSDRGGAVETL------GLQILAELSS---------------EEL 338
                                          .

IcsB      IEKSIKIIDYYN---SLKSPDL--GT-KLYIHDLLQINKLLLNNS--------HSNI--- 494
h1        VETLGRHL-------DAPPPSDSSALRRLAA----------------------HAMIGRI 502
h2        IEKSIKIIDFYQ---SLKNKNQ--EL-LFYTHDLLLVNKTLLDKS--------R------ 492
h3        VEALHHYQ-------ESHPQG--AEQLSLAA----------------------HAMIKRS 507
h4        AERQDELKKGQAELNDLRGQRSRLDTRILAMKDERRALKDLIEERQANPLEAARRKLGGE 569
h5        TSQRDHHRAHYARKYAVPDSPER----------------------------------- 641
h8        ------HEAGYLREASDSDIPS-----------------DFFDN---------------- 520
h10       LIAQNNNA-----------GGSLAAL------------LSLCEEGNNRTLDQSF---KKA 372

IcsB      ------------------------------------------------------------ 494
h1        EA------------FMAAA-------------------------IAA-------------- 512
h2        ------------------------------------------------------------ 492
h3        EE------------LMGLA-----------------------IQ---------------- 516
h4        DDYYAKLRGGLAEGVTGLAHAFGARASGLDTQQQRLLAPLGQAVQDLERGVRNGSDMKTL 629
h5        ------------------------------------------------------------ 641
h8        ------------------------------------------------------------ 520
h10       KKF--------------------------------------------------------- 375

IcsB      ------------------------------------------------   494
h1        ------------------------------------------------   512
h2        ------------------------------------------------   492
h3        ------------------------------------------------   516
h4        LSHAKQLVQTLHDAINGTPAPDRVAALAAGALVQACEFLVVNDQDRHS   677
h5        ------------------------------------------------   641
h8        ------------------------------------------------   520
h10       ------------------------------------------------   375
```

Figure A2. Sequence alignment of non-toxic homologous proteins.

Fully conserved residues



```
h11      RNREVVLQ-----LGQHVLKLYRERS---VRLALLKQSQAKNLVGA--------------       342
                      *  :      :

h6       PRLEQRILALDEDHEALSDLLEDRKAFPVEAAQLRPAGEEGYYAAQRRGLQSCVEGLADQ       600
h7       --------------------F-----------------EDPYFE----------------       503
h9       ARLEQRIPALRDEHEALSGLLNDRTAFPVEAAQLKLPGEEGYYAAQRRGLQSCVDGLAQQ       600
h11      ---------------SKYQLLDEYNAL------------------MAEGLKLGTESWGQ-       368
                         :

h6       FDAHIQGLDARQQRQLAPLGRAVSELRDAVNAGGGMQALLSKAKPLVETLHALTSGAPAA       660
h7       ------------------------------------------------------------       503
h9       FDARLQGLDARQQRQLAPLSGAVADLRAAVDAGAGMQALMSKAKPLVETLHALKDAAPAA       660
h11      ------------------------------------------------------------       368


h6       DQIAILAAGSLVEACEILITLDQDRRA      687
h7       -------------------------        503
h9       DLVAMLAAGTLVEACEVLVTLDQDQHA      687
h11      -------------------------        368
```