STRUCTURAL STUDIES OF THE RECEIVER DOMAIN OF LYTR FROM *STAPHYLOCOCCUS AUREUS* AND INTERACTION STUDIES OF TRAW AND TRBC FROM THE F PLASMID OF *ESCHERICHIA COLI*

AGNESA SHALA

A DISSERTATION SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN CHEMISTRY

YORK UNIVERSITY

TORONTO, ONTARIO

July 2014

ABSTRACT

The emergence of the multi-drug resistant bacteria and the evolutionary spread of the antibiotic resistance genes have become a great concern to human health. Bacteria use various systems in combination or alone in response to environmental stressors and stimuli to adapt to changing conditions. Such systems include the two-component regulatory system (TCS) and the type IV secretion system (T4SS). Microbes use TCSs to regulate important functions such as sporulation, chemotaxis, as well as autolysis that leads to biofilm formation. In addition, microbes utilize T4SSs to deliver DNA, protein substrates, toxin, and virulence factors, from a donor to a recipient cell. These systems are complex, versatile and have a significant impact to human health as they are a major driving force for infection and the spread of antibiotic resistance. The LytSR TCS from *Staphylococcus aureus* has been found to regulate murein hydrolase activity and autolysis by controlling the expression of the *lrgAB* and *cidABC* genes. LytS is predicted to autophosphorylate upon a membrane potential change, consequently transferring the phosphoryl group to the conserved Asp53 residue on the N-terminal domain of the response regulator LytR, leading to the transcription of *lrgAB* genes. In this study we present the X-ray crystal structure of the N-terminal domain of LytR in apo form and in complex with beryllium fluoride. The identification of dimerization interface residues can be utilized to predict modes of dimerization and therefore activation of the protein. A second project focuses on a representative of the conjugative T4SSs, is the F plasmid from *Escherichia coli*, which is responsible for the transfer of virulence and antibiotic resistance genes. Two proteins encoded by the F plasmid, TrbC and TraW, are unique to

the F-like plasmids and are essential for the transfer of DNA from a donor cell to a recipient cell but their mechanism and function is not known. Interaction studies of TrbC and TraW, as well as crystallization experiments are reported. These studies suggest that the roles of TrbC and TraW are related, and that their interaction is important for the functional transfer of DNA in the conjugative process.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my supervisor Dr. Gerald F. Audette. Thank you for giving me the opportunity to work, learn, fail and succeed, all while growing in the midst of intelligent, hard-working and honest individuals. You have taught and encouraged me to express my thoughts and ideas. Your generosity has given me the amazing opportunity to travel and interact with the pioneers of the field. Your honesty, knowledge, and caring personality make you an ideal mentor, and I am lucky to have had the opportunity to work with you.

I wish to thank the members of my Advisory Committee. Your helpful suggestions, continuous support, and encouragement have been greatly appreciated. Dr. Derek Wilson, thank you for the generous opportunity to attend the Gordon Conference, it was an amazing experience both, professionally and socially. Dr. Dasantila Golemi-Kotra, I am truly honored to have worked with you. Thank you so much for trusting me and letting me work with you on your projects. Because of our collaborative work I have had the ability to solve a protein crystal structure. I can never repay you for that.

I wish to thank my extended committee members Dr. David Rose and Dr. Vivian Saridakis. Dr. S. thank you for sharing your knowledge with me, inviting me to your group meetings, and for all the useful crystallography discussions. It was always a pleasure.

I wish to thank my family. Your support, caring, and unconditional love are what inspire me and make me a better person. My parents, Nanë and Babi your love, wisdom, and encouragement drive my ambitions. I love you so much and am proud to be your daughter. My grandmothers Nona and Gjyshja, I will always love you. My wonderful future in-laws: John, Betty-Jean, Grandma Charlotte, Pops who will always be remembered, and the rest of the family. My sisters: Albina, Arbnesha, and Fitore you are my best friends and I could not imagine growing up and getting old without you. My nieces and nephew: Odesa, Juna, baby Charlotte, and baby Miles you make me so happy and look forward to the future while appreciating the past. As well as my extended family, especially my uncle Beli.

I also wish to thank all the people that I have met through the years. My dear friends: Stephanie, Michele, Diana, Julia, Christina, Shaolong, Kevin, Arta, Sara, Ira, Nihareka, Roland, Leila, Ana, Rahima and all other friends from CB and LSB. Working and having fun with you was always a pleasure. I am glad to have met you.

Finally, I want to thank my soon-to-be husband Eric P. Lawrence. Your love, patience and support are the foundations of our success. I would not have been able to do this without you. *You are my everything*.

I dedicate this thesis to Eric P. Lawrence

STATEMENT OF COLLABORATION

A portion of the research described in this Dissertation, regarding LytR studies is part of an on-going collaborative efforts with Prof. Dasantila Golemi-Kotra and Kevin H. Patel from the Golemi-Kotra Group in the Department of Chemistry at York University.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A	Absorbance
Amp	Ampicillin
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BeF-LytR^{N}	N-terminal domain of LytR in complex with beryllium fluoride
CV	column volume
DTT	dithiolthreitol
GlcNAc	N-acetyl glucosamine
GSH	glutathione
HGT	horizontal gene transfer
HK	histidine kinase
HTH	helix-turn-helix
IPTG	isopropyl-β-D-thiogalactopyranoside
Kan	kanamycin
kDa	kilodaltons
LB	Luria-Bertani

$LytR^N$	N-terminal domain of LytR
MALDI-TOF	matrix-assisted laser desorption/ionization 2time of flight
MBP	maltose binding protein
MGE	mobile genetic elements
Mpf	mating pair formation
MR	molecular replacement
MRSA	methicillin-resistant *Staphylococcus aureus*
MurNAc	N-acetyl muramic acid
NBT	nitroblue tetrazolium
NCS	non-crystallographic symmetry
OD_{600}	optical density at λ=600
OG	n-octyl-β-D-glucopyranoside
PAGE	polyacrylamide gel electrophoresis
PBPs	penicillin binding proteins
RMSD	root mean square deviation
RRs	response regulators

SDS	sodium dodecyl sulphate
SEC	size exclusion chromatography
T#SS	type # secretion system
TCSs	Two-component signaling systems
Ve	elution volume
VGT	vertical gene transfer
Vo	void volume
Vt	total volume

Chapter 1

Introduction to Bacterial Evolution and Antibiotic Resistance

1.1 Bacterial Evolution and Antibiotic Resistance

The number of deaths due to the drug-resistant bacterial infections confirms the profoundly disturbing problem of bacterial evolution and antibiotic resistance. While healthcare communities around the world are meant to help humans heal, more than 200,000 patients visiting a hospital get bacterial infections every year with more than 8,000 deaths as a result in Canada alone (Zoutman *et al*., 2003). The adaptation of microorganisms, while advantageous for their survival under adverse conditions (Ochman *et al.,* 2000; Wilkins and Frost, 2001) poses significant challenges to human health causing infectious diseases and make up ~25% of all annual deaths worldwide (Morens *et al*., 2004). Intriguing bacterial evolution examples include *Clostridium difficile*, which has been known since 1930s and was recognized as a threat to human health in 1978. It has only been in the last two decades, since 1997, mortality rates associated with the pathogenic *C. difficile* infections have more than tripled (PHAC, 2013). In 2011 an outbreak of *C. difficile*, which cause diarrhea and intestinal colitis, in a small community in southern Ontario, Canada, caused 20 deaths within 6 weeks of the reported outbreak (CTV, 2011). In addition, *Escherichia coli* is the most common pathogen affecting patients in hospitals causing urinary tract infections and enterocolitis,

in addition to pneumonia and meningitis (Yagi *et al*., 1997). A recent outbreak of Haemolytic Uraemic Syndrome (HUS) caused by enterohaemorrhagic *E. coli* (EHEC) in Germany resulted in over 3000 cases of infection leading to 52 reported deaths (WHO, 2011). In *E. coli* it has been shown that conjugative recombination, increases rates of adaptation by 3-fold under conditions of environmental stress (Davies, 1994). More serious infections are associated with *Staphylococcus aureus*, which cause the most dangerous hospital-acquired infections, most commonly acquired through skin-to-skin contact (Boucher *et al*., 2012). In the early 1940s penicillin was very effectively used to treat illness due to *S. aureus*, but because of its mass production and its wide use for various human infections, the microbes quickly adapted, and today over 90% of human *S. aureus* strains are penicillin-resistant (Olsen *et al*., 2006, Malachowa & DeLeo, 2010). Subsequently, *S. aureus* is believed to have obtained vancomycin resistance from enterococci (facultative anaerobic organisms) leading to the emergence of vancomycin-resistant *S. aureus* (VRSA) (Weigel *et al*, 2003). Moreover, soon after the introduction of the antibiotic methicillin, the emergence of methicillin-resistant *S. aureus* (MRSA) was confirmed (Knight *et al*., 2012; PHAC, 2013). The sequence of events leading to this bacterial resistance, due to the course of their evolution, makes it easy to predict their future, as they will combat whatever means of protection we use against them. The disturbing facts include the fundamental property of *S. aureus* to asymptomatically colonize close to 30% of healthy humans' skin, making them carriers of infection. These persons are more prone to serious infections in case of failure of the immune system or breach of the protective skin layer (Chambers & DeLeo, 2009). In attempting to keep up

with antibiotic resistance in human pathogens, the healthcare community may be contributing to its rapid development by prescribing the overuse and misuse of antibiotics. Statistical data show that healthcare associated MRSA infection and colonization has been continually increasing, where the infection rate has gone up by more than 1,000% from 1995 to 2009 (PHAC, 2013).

1.1.1 Bacterial Adaptation Through Gene Transfer

The imposed pressures such as the presence of antibiotics has enabled the microbes to find adaptive mutations to provide growth advantages through natural and adaptive means. This adaptation is very easily facilitated via the exchange of genetic material between cells occurring due to horizontal gene transfer (HGT) and vertical gene transfer (VGT) (Figure 1.1). HGT occurs between eukaryotic cells, prokaryotic or a combination of the two from the same generation, where the resistance-determining genetic information is transferred to the more susceptible bacterial cells. In addition, these cells when introduced to environmental stressors such as antibiotics, over time, lead to the survival of the fittest and proliferation of the new generation of resistant bacteria, through lateral or VGT (Malachowa & DeLeo, 2010).

Trafficking of the genetic material (DNA) as a result of adaptation, induced by environmental stressors, is considered to have a major role during HGT in microorganisms (Lindsay, 2010). Most bacteria secrete or intake macromolecules such

as proteins, DNA, or protein-DNA complexes across their cell envelope in order to ensure their metabolism and adaptation in different environments (Baron & Coombes, 2007). DNA fragments that encode various virulence, resistance, as well as enzymes that mediate their own transfer and integration into the host DNA are also known as Mobile Genetic Elements (MGE; Frost *et al*., 2005). The transfer of the MGEs in the form of bacteriophages, pathogenicity islands, plasmids, transposons and staphylococcal cassette chromosomes, account for the 15-20% of the genome of *S. aureus* isolates (Lindsay, 2010). They are the means of transferring genetic information among and within bacterial species. When comparing the genome sequencing of *S. aureus* populations their distinguished MGE profiles suggest the frequent transfer or loss of whole elements, making MGEs a key for identification and keeping track of the genes acquired as a result of resistance (Lindsay, 2010).

The physical process for the intercellular transfer of genomic material between microbes occurs through three different pathways: specific transfer mediated by (1) bacteriophage transduction, (2) direct contact between cells through conjugation, and (3) uptake of free DNA by competent microbes from their environment as transformation (Figure 1.2) (Frost *et al*, 2005; Lindsay 2014). Transduction is defined as DNA transfer mediated by independently replicating bacterial viruses known as bacteriophages. Conjugation is a specific process that requires independently replicating genetic elements known as conjugative plasmids (such as transposons) to encode proteins that facilitate their own transfer and the transfer of other cellular DNA from the host cell to a recipient lacking such plasmid DNA through a direct contact via the pilus (Frost *et al*, 2005).

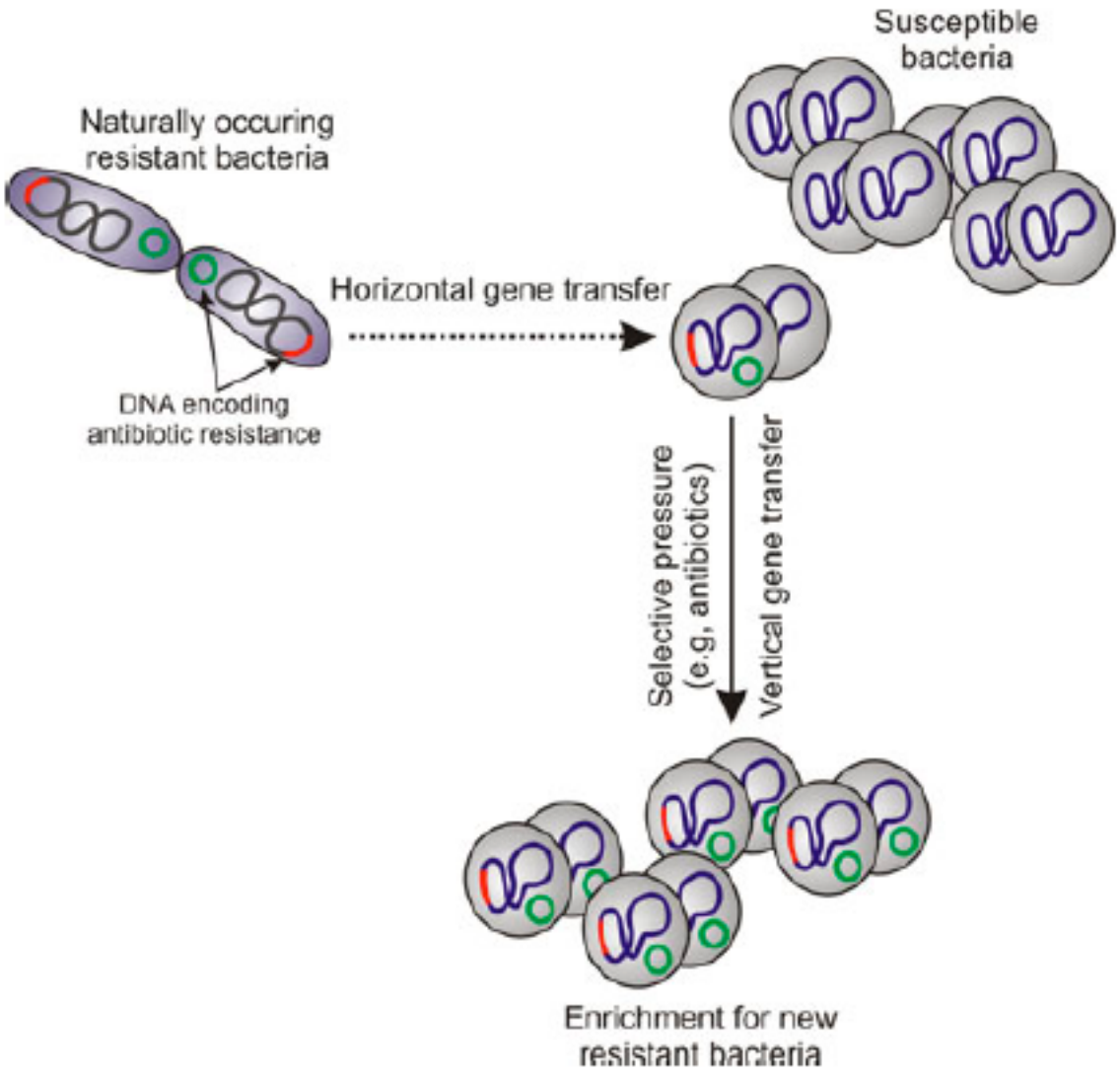


Figure 1.1. Horizontal and Vertical gene transfer between bacterial cells. HGT occurs within the same generation of cells, due to the transfer of antibiotic resistance genes to more vulnerable bacteria. VGT occurs as a result of pressure induced to the microbes by antibiotics destroying the susceptible cells, and allowing the resistant microbes to proliferate, passed from parent to progeny cells. Copied from Malachowa and DeLeo, 2010.

Transformation was the first discovered HGT mechanism involving the non-specific transfer of DNA into competent cells that can easily uptake foreign materials via the permeable membrane.

1.1.2 Modes of Antibiotic Resistance in Bacteria

Antibiotics can be naturally occurring or synthetic therapeutic agents that act against invasive microorganisms at low concentrations with selective toxicity not harming the affected host. It has been over 80 years since Sir Alexander Fleming discovered penicillin from the antibiotic producing fungi (Fleming, 1929); during the subsequent twenty years, roughly half of the antibacterial drugs used today would be supplied to modern medicine (Davies, 1994). Since the introduction of quinolones in 1962, such as nalidixic acid used for treatment of genitourinary infections, and more recently in the year 2000 the approval of oxazolidinones, used as the last resort for treatment of bacterial infections, no new antibiotics made their way to the pharmacy shelves (Walsh, 2003). Overall the rate of traditional antibiotic discovery is in decline (Nathan, 2004), and as such more targeted, structure-based methods have emerged to develop novel antibiotics (Kuhn *et al.,* 2002; Kan *et al*., 2005). However, it is becoming a trend that antibiotics, given enough time, will fail to provide the necessary means of combating microbes. This is not so much a result of the properties of antibiotics themselves, but rather the ability of microbes to obtain the genes necessary for antibiotic inhibition or digestion as well as

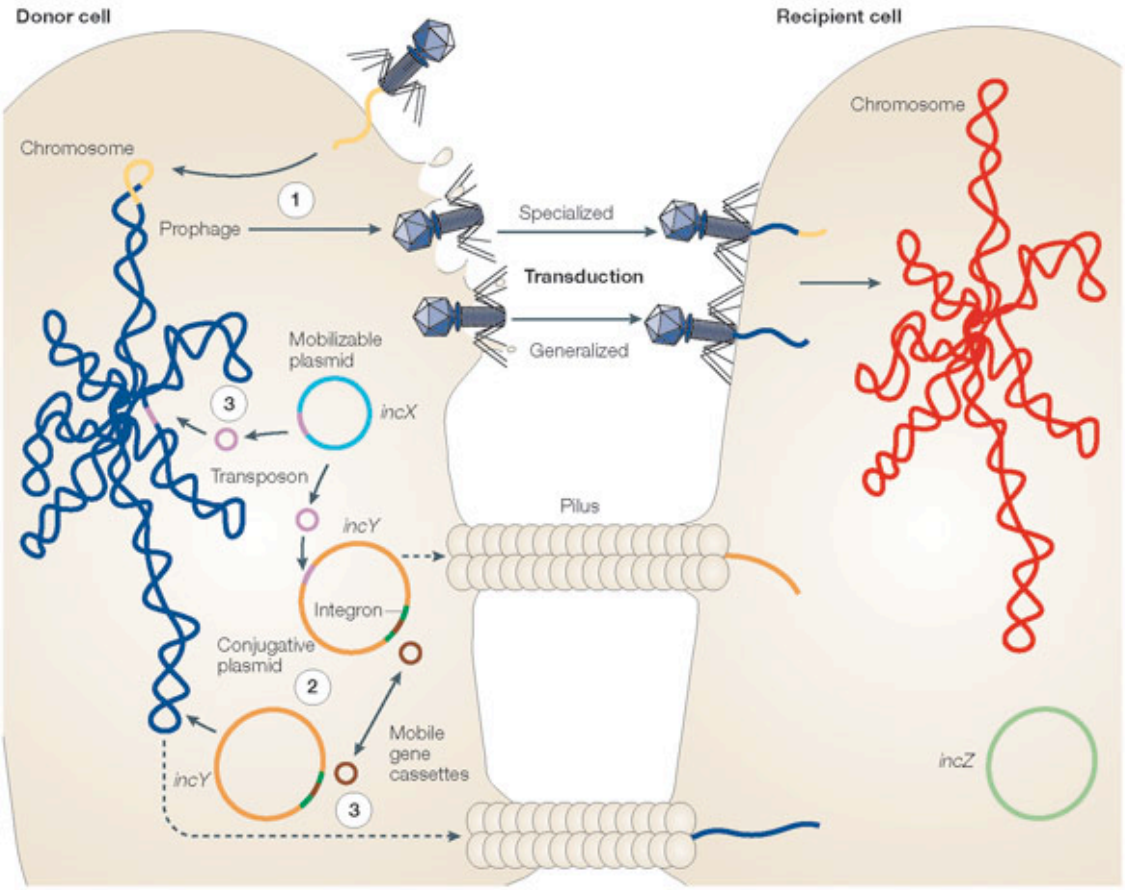


Figure 1.2. DNA and MGE transfer between bacterial cells via the bacteriophages transduction (1), plasmid transfer through pilus formation and conjugation (2) and uptake of free DNA by competent cells known as transformation (3). Copied from Frost et al., 2005.

trafficking the drugs out of the cell. So to a bacterium, development of antibiotic resistance is merely a particular adaptation for a given environment. In addition to acquiring antibiotic resistance genes and molecules involved in metabolism, bacterial plasmids have acquired resistance to various highly toxic organic and inorganic ions, including cadmium, mercury, and arsenate, among others (Jensen & Lyon, 2009).

Antibiotic resistance in prokaryotes is achieved through five different strategies: reduction of membrane permeability to metals and antibiotics; drug and metal inactivation and modification; rapid efflux of the metal and antibiotics; mutation of cellular targets to lower their sensitivity to the toxic metals and antibiotics; and drug and metal sequestration (Table 1.1) (Baker-Austin *et al*., 2006).

Prokaryotes employ different systems for survival, adaptation, and normal cell function. While there are numerous systems for various purposes both Two-Component Regulatory Systems (TCSs), and bacterial Secretion Systems (SSs) are found to contribute indirectly and directly to bacterial antibiotic resistance, respectively. TCSs that may lead to cell autolysis and the release of the extracellular DNA, they make it available for other microbes to pick up (Galperin, 2010). On the other hand, some secretions systems are able to directly inject DNA from one bacterial cell to another lacking such genetic information (Frost *et al*., 2005).

1.2 Two-Component Signal Transduction Systems

Bacteria have developed the means to adapt to constantly changing environmental conditions by responding to their environmental stimuli. Such stimuli include nutrients such as oxygen, nitrogen, and carbon, in addition to the alterations in temperature, osmolarity and acidity. Two-component signaling systems (TCSs) that function as phospho-transfer pathways, have enabled pathogenic bacteria to respond and adapt to the changing environmental conditions, which are vital for their survival. In addition to bacteria, TCSs are also found among fungi and plants. These signal transduction systems regulate gene expression through transcription, post-transcription, and post-translation pathways, in addition to regulating protein-protein interactions (Galperin, 2010). The simplest type of TCSs contains a membrane bound sensory histidine kinase (HK), and a two-domain response regulator (RR) while more complicated systems can include multiple phosphotransfer reactions and a combination of systems (Toro-Roman *et al*., 2005). The typical mode of action of TCSs commences with the signal sensing or ligand binding to the HK, initiating signal transduction, which upon autophosphorylation transfers a phosphoryl to its cognate RR at the conserved N–terminal Asp residue, affecting the properties of the DNA-binding domain at the C-terminus, thus regulating gene expression (Stock *et al*., 2000) (Figure 1.3).

Table 1.1. Structural and functional characteristics of prokaryotic antibiotic* and metal resistance strategies.

Resistance Mechanism	Metal Ions	Antibiotics
Reduction of permeability	Au, Cu, Zn, Mn, Co, Ag	Cip, Tet, Chlor, β-lactams
Drug and metal alteration	As, Hg	β-lactams, Chlor
Drug and metal efflux	Cu, Co, Zn, Cd, Ni, As	Tet, Chlor, β-lactams
Alteration of cellular targets	Hg, Zn, Cu	Cip, β-lactams, Trim, Rif
Drug and metal sequestration	Zn, Cd, Cu	CouA

*Antibiotic abbreviations: Cip, ciprofloxacin; Tet, tetracycline; Chlor, chloramphenicol; Trim, trimethoprim; Rif, rifampicin; CouA, coumermycin A. Adapted from Baker-Austin et al., 2006.

1.2.1 Histidine Kinases

The two-component system Histidine Kinases are catalytically similar to Ser/Thr/Tyr protein kinases. The difference exists in their chemistry where the latter create phosphoesters, and the TCS HKs create phosphoramidates. Given the high negative free energy for phosphoramidates (N-P bond), the equilibrium for phosphorylation of HKs prefers the unphosphorylated protein, resulting in phosphoryl transfer (Stock *et al.,* 1990). Despite the high-energy barrier, as a result of environmental stimuli, HKs efficiently get autophosphorylated during the efflux of phosphoryl groups. The ATP-dependent autophosphorylation of HKs involves two homodimers, where one HK monomer catalyzes the phosphorylation of the conserved His residue on the second monomer (Surette *et al.,* 1996; Pan *et al*., 1993; Yang *et al.,* 1993; Wolfe *et al*., 1993). Another difference between the HKs and other kinases is that HKs transfer the phosphoryl group to a conserved Asp residue of its specific RR of TCSs, unlike other types of kinases, which can phosphorylate multiple target RRs. In addition to autophosphorylation and phosphotransfer HK functions include dephosphorylation of their cognate RR (Keener & Kustu, 1988).

HKs are divided into two classes: orthodox and hybrid kinases (Parkinson & Kofoid, 1992; Alex & Simon, 1994). Orthodox HK function as periplasmic membrane receptors, which have the N-terminal signal sensing domain and C-terminal catalytic kinase core, such as EnvZ from the *E. coli* involved in osmoregulation (Figure 1.4a). On the other hand, hybrid kinases that are mostly found in eukaryotic systems contain

multiple phosphodonor and phosphoacceptor sites, examples include the *E. coli* anoxic redox control ArcB HK (Figure 1.4b). Orthodox kinases, in addition to being membrane bound, they also exist in soluble form functioning in the cytoplasm, such as the chemotaxis kinase CheA (Figure 1.4c) (Stock *et al*, 1988) and the nitrogen regulatory kinase NtrB (MacFarlane & Merrick, 1985). In addition to the regular forms of HK, there is also a combination of multiple independent proteins that can convey regulatory functions such as the *B. subtilis* sporulation control (Burbulys *et al*., 1991)(Figure 1.4d).

The general structure of transmembrane HKs is composed of the sensing domain, linker domain, and the kinase core (Figure 1.4a), EnvZ histidine kinase; sensing domain in white, linker domain in black, the kinase core consisting of the five conserved residues: H, N, G1, F, G2, in dark grey). The kinase catalytic domain consists of the catalytic ATP-binding domain and the dimerization domain (Stock, 1999). The catalytic ATP-binding domain houses the four conserved residues, N, G1, F and G2 forming the highly flexible ATP binding region (Stock & Ninfa, 1989).

The dimerization domain holds the sensor His residue, which is located on the face of the solvent exposed helix of the dimerization domain of one of the monomers, in close proximity to the ATP pocket of the other monomer (Stock *et al*., 2000). While the kinase core is conserved among different systems, the sensing domain has very low sequence similarities with other such domains because of its selective stimuli detection (Stock *et al.,* 2000). In addition the linker region connecting the sensor and the catalytic

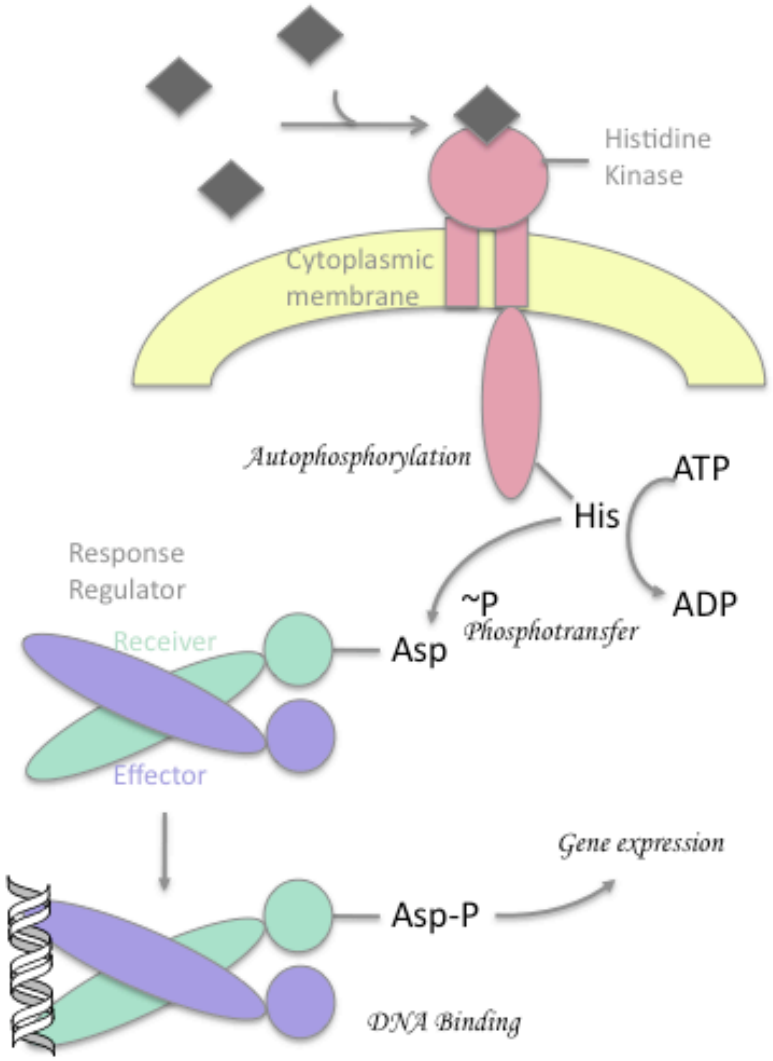


Figure 1.3. Gene expression as regulated by two-component regulatory systems. Sensory information detected by HK is transferred through phosphorylation to the Receiver domain of RR, leading to the DNA binding of the effector domain, and gene output regulation. The grey diamonds represent the stimuli sensed by the HK. Modified from Parkinson, 1993.

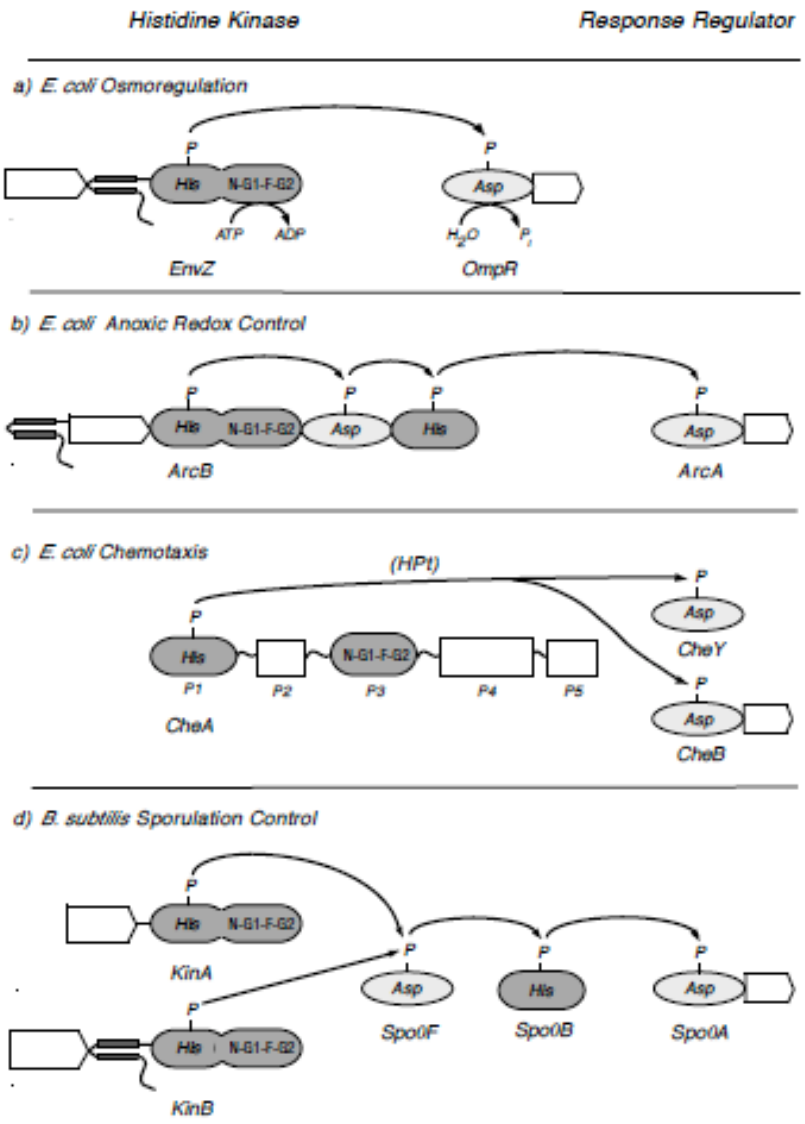


Figure 1.4. Representative types of orthodox and hybrid HKs. His-domains in dark grey, Asp-domains in light grey, variable domains in white. (a) the Orthodox HK example, from the *E. coli* EnvZ/OmpR osmoregulatory TCS. (b) the Hybrid HK example, containing multiple His-containing phosphotransfer sites. (c) An orthodox, soluble HK example which upon receiving signal from the cytoplasm CheA, phosphorylates the single domain CheY, or the methylesterase CheB. (d) Alternate multicomponent independent protein phosphorelay system from *B. subtilis* sporulation, the phosphate is transferred from either HK, to Spo0F, subsequently to Spo0B before reaching the RR Spo0A. Copied from Stock *et al*, 2000.

domains, although not fully evaluated, appears to have a crucial role for proper signal transduction (Atkinson & Ninfa, 1992).

1.2.2 Response Regulators

The bacterial TCSs regulate and respond to environmental changes using phospho-relay pathways, starting with the signal detection by HK and ending with the phospho-signal transfer to the terminal component RR. As the final step for regulating the response to different extracellular and intracellular stimuli RRs have many different targets, such as regulation of chemotaxis, osmoregulation, nitrogen fixation, in addition to antibiotic gene transcription through different types of DNA-biding domains (Galperin, 2006).

Response regulators consist of a receiver domain located on the N-terminal end of the protein and an effector domain on the C-terminal end of the protein. While most RRs share a similar phospho-acceptor, receiver domain, they differ and are classified into subfamilies according to their DNA-binding effector domain (Galperin, 2010). The RR catalyzes the phosphoryl group transfer from the Histidine (His) residue of the HK to a conserved Aspartate (Asp) residue on its receiver domain. In addition to the transfer of the phosphate by HK, RRs catalyze the removal of the phosphoryl group independently of HK as suggested by the observation of phosphorylation and dephosphorylation of the Asp by small molecules such as acetyl phosphate and imidazole phosphate (Lukat *et al.,*

1992). Dephosphorylation is an important function of the receiver domains, as they are activated or deactivated by the presence or the removal of the phosphate. Activation is generally triggered by the presence of the phosphoryl group causing major conformation changes leading to protein dimerization (Fiedler & Weiss, 1995; McCleary, 1996), higher order oligomerization (Weiss *et al.,* 1992) and protein-protein (Welch *et al*., 1993) or protein-DNA interactions (Makino *et al*., 1989). Phosphorylation-induced dimerization causes major conformational changes on the protein, leading to the DNA binding on the effector domain, thus controlling gene regulation (Stock *et al*., 2000). In rare cases dephosphorylation of the RR can trigger activation as well (Simms *et al.,* 1985), an example of this is osmoregulation in yeast, which is activated by RR dephosphorylation (Posas & Saito, 1998).

It has been almost thirty years since the first bacterial RR genes were sequenced and their domains identified (Drury & Buxton, 1985). The most abundant transcriptional regulators belong to the OmpR/PhoB family classified by 33% of all RR, characterized with a winged helix-turn-helix (*w*HTH) DNA binding domain (Buckler *et al*., 2002; Galperin, 2006; Galperin, 2010). The second most abundant family with 18.7% is NarL/FixJ, characterized by a typical HTH DNA binding domain (Galperin 2006; Noriega *et al*., 2010; Galperin, 2010). Both of these types of transcriptional regulators are further divided into subfamilies based on the similarities of their output domains covering 95.3% of all the characterized RRs.

A much different family of transcription factors is AlgR/AgrA/LytR characterized by a non-HTH DNA binding domain, named LytTR domain, covering the remaining 4.7% of the DNA-binding response regulator proteins (Nikolskaya & Galperin, 2002; Sidote *et al.,* 2008) involved in regulation of important virulence factors and toxin production in pathogenic bacteria (Rood, 1998; Lizewski *et al*., 2002; Novick, 2003; Dawid *et al*., 2007; Galperin, 2008). While the above-mentioned RRs have two domains, the stand-alone receiver domains make up ~14% of RR (Galperin, 2006). The best example of stand-alone RRs, which is often used to describe the receiver domains of all other RRs, is the chemotaxis protein CheY that regulates the direction of the flagellar movement in *E. coli* (Voltz and Matsamura, 1991; Atkinson & Ninfa, 1992). Although RRs are mainly associated with transcriptional regulators, many of them do not regulate transcription directly. These RRs are enzymatic proteins such as the methylesterase CheB, which upon phosphorylation controls receptor methylation levels; other examples include protein phosphatases, histine kinases etc. While CheY and CheB are different types of RR they are both regulated by the same HK CheA, and CheB activation involves multiple protein-protein interactions (Djordjevic *et al.,* 1998).

While the protein data bank (PDB) contains numerous structures of different RR receiver domains, there are very few available structures for the full-length proteins or for the DNA-binding domains; in many cases insolubility and lack of stability of the DNA-binding domains is to blame (Galperin, 2010). A representative of the receiver domains from different types of RRs is CheY. The structure of the protein consists of an α/β motif, with five-stranded β-sheets surrounded by the five α-helices (Figure 1.5). The

conserved phosphorylation site is on Asp57, located on the solvent exposed loop between β3 and α3. The remaining conserved active site residues include Asp12, Asp13, Thr87 and Lys109 (Sanders *et al*., 1992). While numerous residues are affected by the activation of the protein, the signature Tyr106 rotameric states between the active and inactive forms, as shown in Figure 1.5, are seen in other RRs as well (Zhu *et al*., 1997; Jiang *et al*., 1997). Many similarities exist between different receiver domains but they display major differences when forming the active dimers and the residues affected by the activation at the dimerization interface. The dimerization interface between PhoB (from the *w*HTH family) monomers occurs between α4-β5-α5 (Bachhawat *et al*., 2005). In FixJ (from the HTH family) this dimerization interface occurs between α4-β5 (Birck *et al*., 1999). In ComE (which belongs to the lytTR family) between α4-loop α4-β5 (Boudes *et al.,* 2014), whereas in VraR the interaction between helices α1-α5 form this interface (Leonard *et al*., 2013). While these proteins from different families of RRs have many similar functional and structural aspects, they do regulate the transcription of different genes and so they have different ways of doing so as suggested by the different modes of dimer activation.

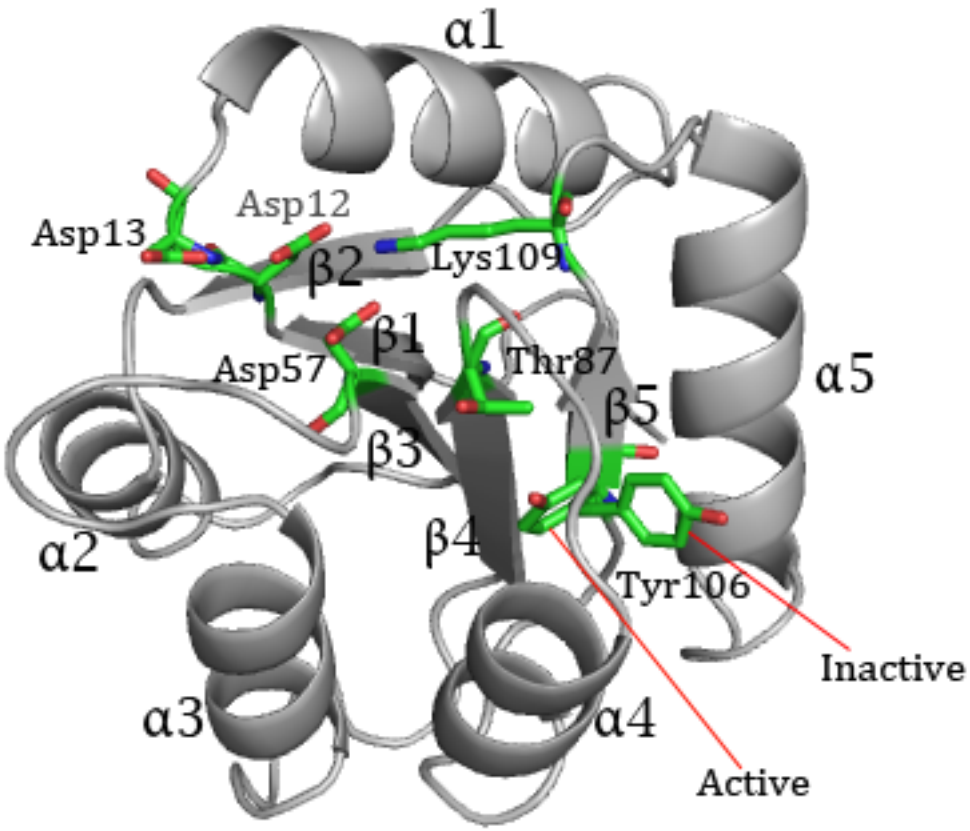


Figure 1.5. Structure of the CheY protein (PDB ID: 3CHY) shown as grey cartoon. The conserved residues in the active site are shown in green as sticks. In addition shown in green are the two conformations of Tyr106, in active or inactive forms of the protein. Oxygens are shown in red, and Nitrogen atoms in blue. This figure and all other structure figures were generated by using PyMol (Schrodinger, 2010).

1.3 Bacterial Secretion Systems

In response to changing environmental conditions, besides transcription based adaptation by TCSs, bacteria also use secretion systems (SS) for adapting into their new milieus (Campbell-Valois & Sansonetti, 2014). Bacteria have evolved specialized SSs to support the transfer of macromolecules, DNA, or toxins across their cellular membranes. Based on their genes and the virulence effectors that these secretion systems transfer different types exist and they are numbered types I through VIII (Chen *et al*., 2005; Frost *et al.,* 2005; Abdallah *et al.,*2007). Types I and V secretion systems are relatively simplistic, containing 1 to 3 component proteins, while Types II, III and IV are more complex (Hazes & Frost, 2008). The type IV secretion system (T4SS) is the most complex, containing between 8 and 20 core proteins; bioinformatics has shown that T4SSs are ancestrally related to bacterial conjugation systems (Lawley *et al*., 2003).

1.3.1 Type I Secretion System (T1SS)

One of the simplest bacterial secretion systems is the type I secretion system (T1SS) used by many gram negative bacteria to translocate proteins, pore-forming toxins, proteases, and lipases, across both inner and outer membranes into their surroundings (Gentschev *et al*., 2002). This secretion systems is independent of periplasmic intermediates, since the secretion signals recognized by the system are located at the carboxy terminus of the secreted proteins (Duong *et al*., 1996). The best example is the

E. coli α-hemolysin HlyA transfer system from *E. coli*, which causes urinary tract infections, due to its cytolytic and cytotoxic activity against a wide range of mammalian cell types (Hueck, 1998).

T1SSs consist of three proteins: an outer membrane (OM) pore, an inner membrane bound ATP-Binding Cassette (ABC) used to power protein secretion, and an adapter protein that connects these two components (Henderson *et al*., 2004). The prototypical T1SS in *E. coli* is composed of the HlyB (ABC), TolC (OM pore) and HlyD (adapter) proteins (Figure 1.6) (Gentschev *et al.,* 2002). Crystallographic analysis of TolC has shown that a TolC trimer forms a 12-stranded antiparallel β-barrel that spans the outer membrane, while the periplasm-spanning domain of the protein is composed of α-helices. TolC forms a 140 Å (14 nm) long tunnel 3.5 nm in diameter through which effector molecules can travel (Holland *et al*., 2005). In addition, HlyD has also been shown to be trimeric, enabling its interaction with TolC resulting in a structure that spans both inner and outer membranes including the periplasmic space (Henderson *et al*., 2004).

1.3.2 Type II Secretion System (T2SS)

The type 2 secretion system (T2SS), first described in *Klebsiella oxytoca* in the 1980s (d'Enfert *et al*., 1987) is one of the more complex secretion systems in addition to being one of the more versatile found in different pathogenic strains. Unlike other

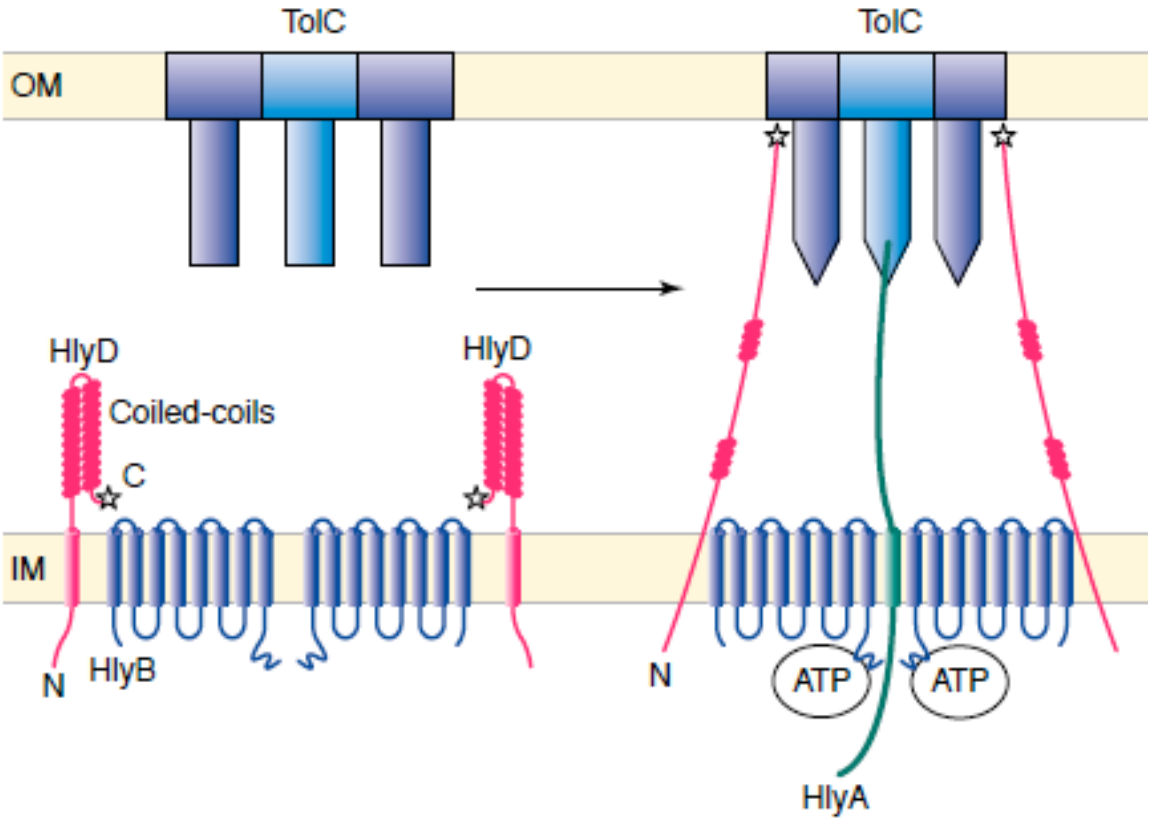


Figure 1.6. Topological model of the T1SS hemolysin secretion. In response to HlyA, the HlyD trimer interacts with the trimeric TolC forming a trans-periplasmic export channel. ATP hydrolysis by HlyB provides the energy for the transport process and secretion of HlyA into the extracellular medium. OM, outer memebrane; IM, inner membrane. Copied from Gentschev *et al.*, 2002.

systems, this does not span both inner and outer membranes but it uses Sec translocon pathways to cross the inner membrane. This system promotes specific transport of folded periplasmic proteins in gram negative bacteria, across a channel in the outer membrane. For instance, *Vibrio cholera* employs a T2SS in the secretion of its cholera toxin (Zhang *et al*., 1995), as well as the enteropathogenic *E. coli* for the secretion of the shiga toxin (Johnson *et al*., 2006). In addition, a T2SS is utilized in the assembly of the type IV pilus (T4P), a fibre-like structure that serves multiple roles in the interaction of a bacteria and it's environment, in *Pseudomonas aeruginosa* (Hazes & Frost, 2008). T2SS proteins have been shown to have homologies to other secretion systems, including the T3SS (Bitter *et al*., 1998) and T4SS (Hazes & Frost, 2008) in addition to archaeal pili and flagella. There are 12-15 genes that encode for a T2SS, the complexity and the various components, some of known and many still unknown structure/function, assembled and involved in T2SS are shown in Figure 1.7 (Mangayarkarsi & Francetic, 2014).

The T2SS is utilized to secrete a range of effectors including proteases, cellulases, lipases and toxins that destroy various tissues and contribute to disease (Johnson *et al*., 2006). Secretion through a T2SS involves a two-step mechanism; the first step involves the ATP-dependant transport into the periplasm across the inner membrane of a protein substrate containing an N-terminal signal sequence via the secretion pathway (Henderson *et al*., 2004; Johnson *et al*., 2006). Once the protein has reached the periplasmic space, the N-terminal signal peptide is cleaved, thereby enabling translocation across the outer membrane through the outer membrane secretion complex. This outer membrane complex is a ring-like assembly of an integral membrane protein (for example PulD and

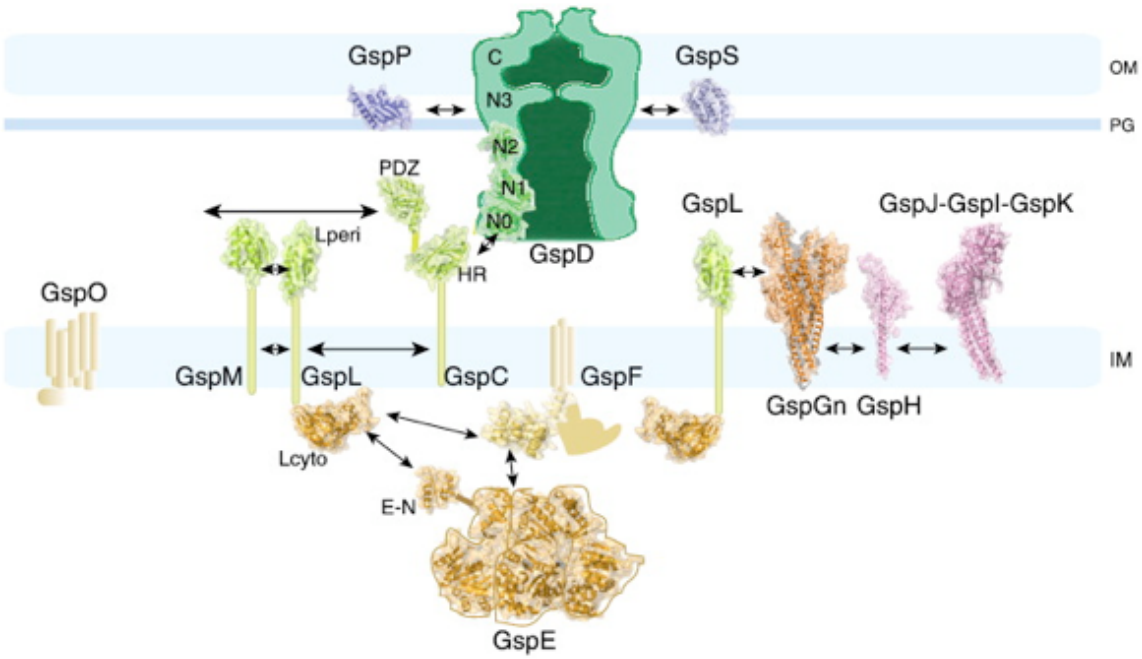


Figure 1.7. T2SS components location and interactions in gram negative bacteria envelope. Components of known structure are shown in cartoon, whereas the unknown structures are shown as cylinders or lines. Known interactions between proteins are shown as double arrows. OM,outer membrane; PG, peptidoglycan; IM, inner membrane. Copied from Mangayarkarasi & Francetic, 2014.

PilQ in *V. cholera* and *P. aeruginosa* T2SSs, respectively) that forms a 5-10 nm pore through which the toxin or T4P emerges (Hazes & Frost, 2008).

1.3.3 Type III Secretion System (T3SS)

Numerous gram negative pathogens, including *Salmonella typhimurium* (Valdez *et al*., 2009) and *Burkholderia pseudomallei* (Sun & Gan, 2010), use a T3SS to deliver effector molecules to target cells. This system uses a dual function needle at the membrane tip for translocation of virulence proteins into a host cell, or as a flagellar hook where polymerization of an extracellular filament occurs dedicated for bacterial locomotion (Abrusci *et al.,* 2014). The prototypical T3SS is the *Yersina pestis* system, which delivers the Yop proteins (Michiels *et al*., 1990). Often termed the injectisome, T3SSs with approximately 30 proteins is one of the more complex secretion systems (Henderson *et al*., 2004). A defining feature of the T3SS is the needle complex, a 120 nm structure that spans both inner and outer membranes of the bacterium (Sun & Gan, 2010). The 8 nm wide injectisome needle extends out towards the target cell approximately 60-80 nm and is analogous to the flagellar hook (Figure 1.8). Similar to T1SS, the T3SS is a Sec-independent secretion system; protein secreted through a T3SS do not contain a cleavable signal peptide, which is a hallmark of proteins secreted via Sec-dependent pathways (Galán, 1999). It is through this needle complex that contact dependent secretion of effector proteins into the target cell occurs; T3SS effectors have the capacity to manipulate the host cell in several ways. For example, some bacteria secrete effector

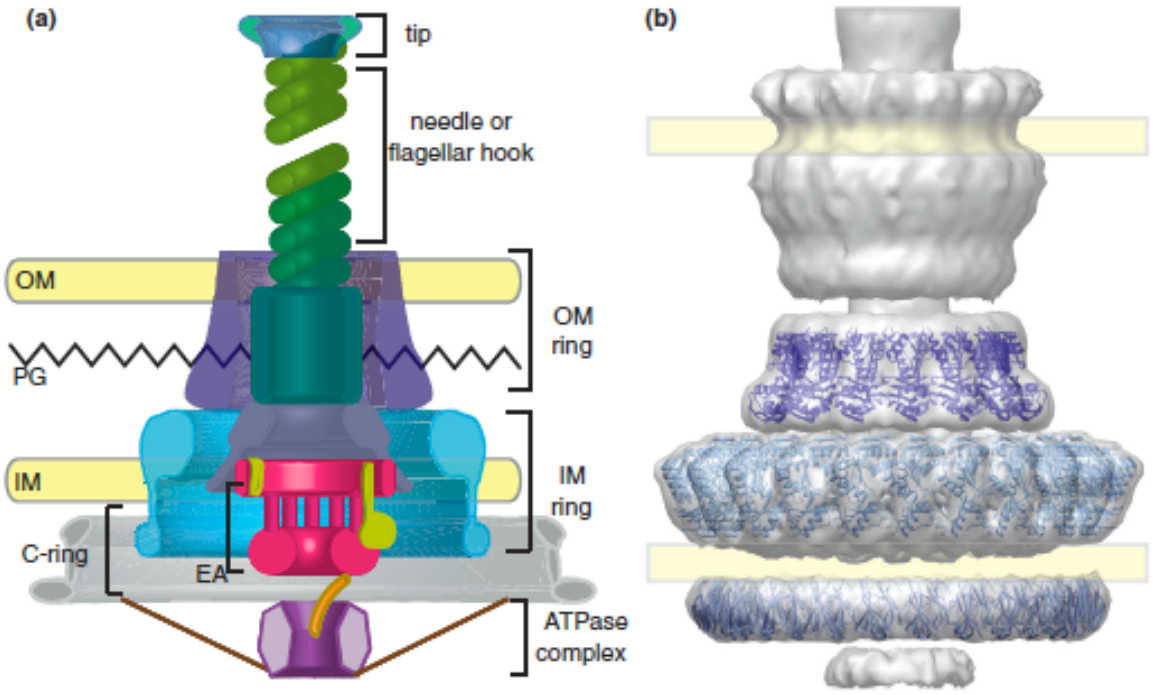


Figure 1.8. General assembly of the T3SS. **A.** A cartoon showing the localization of the different system components. **B.** Surface view of T3SS from *Salmonella typhimurium* with fitted atomic models of known structures. Copied from Abrusci *et al.,* 2014.

proteins that generate a signal for the host cell to engulf the bacterium allowing for quick and efficient access into the host cell machinery and infection of the entire tissue, while others modify the host's cell cycle or induce apoptosis (Gophna, 2003).

1.3.4 Type V Secretion System (T5SS)

The simplest protein secretion system is the type five secretion system (T5SS). Similarly to T2SS, this system does not form a complex through both inner and outer membranes, but it uses Sec translocon pathway for two-step secretion. The transport through the inner membrane is in a Sec-mediated fashion, followed by effector self-translocation through the outer membrane (Henderson *et al.,* 2004). Generally, members of this family contain a signal peptide at the N-terminus mediating transport though the inner membrane with an effector domain that exerts biological activity in the extracellular space. In addition a linker domain connects the effector domain and the C-terminal β-domain that assembles into a transmembrane β-barrel, forming a hydrophilic pore in the outer membrane, enabling the effector domain to cross the outer membrane (Gawarzewski *et al*., 2013) (Figure 1.9). T5SS is one of the most simplistic means of exporting adhesins, enzymes, toxins and other virulence factors of varying sizes and structures in the extracellular milieu (Oomen *et al*., 2004). Although not many components are involved, the complete mechanism of these systems is not fully developed (Gawarzewski *et al*., 2013).

1.3.5 Type VI Secretion System (T6SS)

Since 2006 when the type VI secretion system (T6SS) was first described (Pukatzki *et al.,* 2006), this system is found in 25% of all sequenced gram negative bacteria, making it the most widespread secretion system (Bingle *et al.,* 2008). T6SSs are encoded by 13 conserved components that are required for its function (Zheng & Leung, 2007). Structurally, T6SSs span both inner and outer membranes across the periplasm through a protein complex whose components have not entirely been elucidated, although most proteins identified thus far are employed in the formation of the protein channel that facilitates the translocation of substrates across the peptidoglycan layer (Pukatzki *et al*., 2009). Recent EM and crystallographic studies indicate that the T6SS contractile injection make them analagous to tailed bacteriophages, suggesting specificity and efficiency for these systems (Pukatzki *et al*., 2007). The best studied examples of T6SSs include the secretion of both, the hemolysin co-regulated protein (Hcp) and the valine-glycine repeat protein G (VgrG). Hcp forms a hexameric ring structure that polymerizes to form a channel wide enough to allow for the secretion of globular proteins up to 50 kDa in size that are up to 100 nm long (Ballister *et al*., 2008). The Hcp tube is capped by VgrG, which forms a trimer containing a C-terminal triple-stranded β-helix. The VgrG trimer forms a needle shaped structure with a purpose to puncture the target cell membrane and enable effector molecule delivery (Figure 1.10).

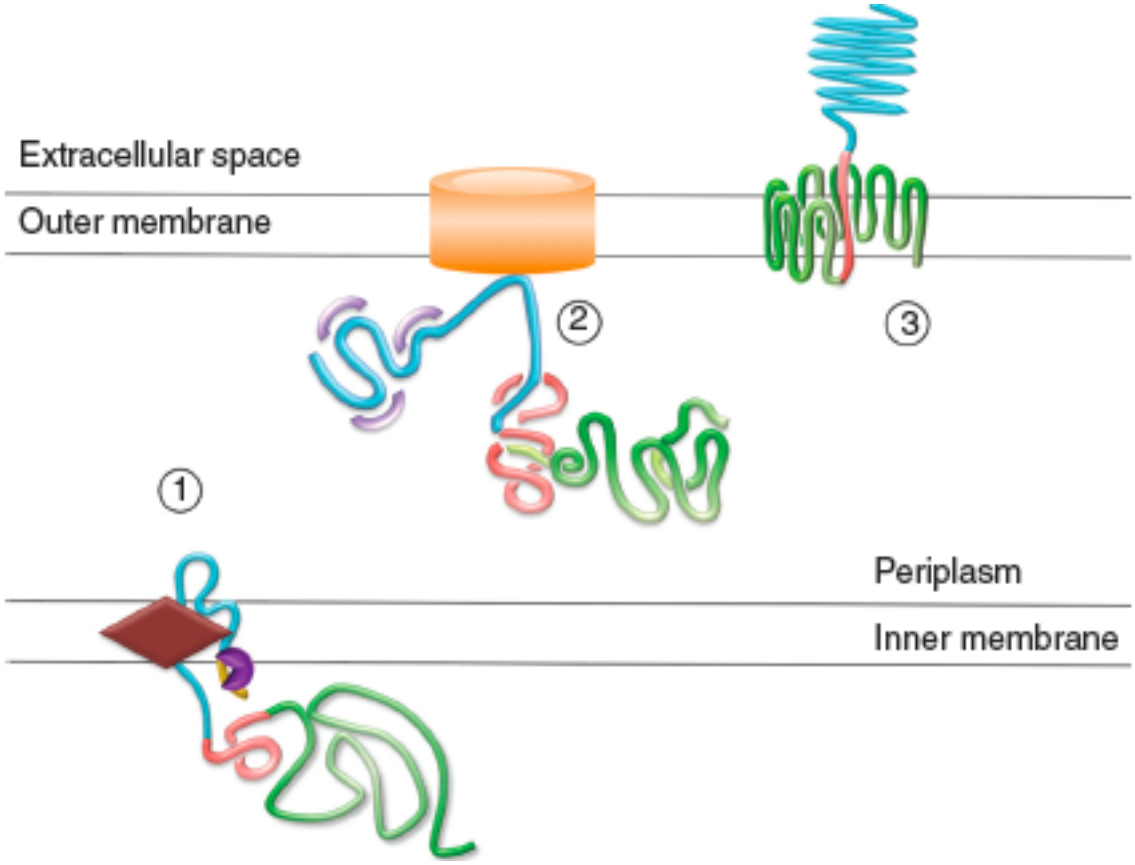


Figure 1.9. Schematic function of the T5SS. Brown, Sec translocon; purple, signal peptidase; yellow, signal peptide; green beta-barrel; blue, effector domain; red, linker domain; violet, periplasmic chaperones; orange, Omp85. 1), peptide chain transported through the inner membrane by the Sec translocon. 2) in the periplasm the intermediate forms a stable conformation while interacting with chaperones and Omp85. 3) the effector domain is translocated through the outer membrane. Copied from Gawarzewski *et al.,* 2013

While the T6SS delivered VgrG1, or Hcp proteins from *Vibrio Cholerae, Agrobacterium tumefaciens*, or *Edwardsiella tarda*, affecting animal, plant or fish, respectively, have been reported to have pathogenic effects; these systems can also have a bactericidal effect by increasing resistance against gram negative bacteria and unicellular eukaryotes (Kapitein & Mogk, 2013). The delivery of Tse1 and Tse3 peptidoglycan degrading effectors, is done by T6SS (Rusell *et al*., 2012). These effectors can increase bacterial fitness allowing the good bacteria, such as the gastrointestinal tract flora, to outcompete pathogenic bacteria. In addition the T6SS of *Pseudomonas augerinosa* has been found to increase the resistance in the presence of *Saccharomyces cerevisiae* by suppressing growth of yeast cells. Due to their dual role T6SS are an attractive target for the discovery of novel antibacterial therapeutics (Kapitein & Mogk, 2013).

1.3.6 Type VII Secretion System (T7SS)

While other secretion systems are used by bacteria to export unfolded polypeptides or monomeric folded proteins, type seven secretion system (T7SS) is used for secretion of a folded multimeric complex (Sysoeva, *et al*., 2014). As a recently identified secretion system, T7SS initially found in the gram positive *Mycobacterium tuberculosis* affecting bacterial processes such as sporulation, conjugation, and bacterial cell wall stability (Sysoeva *et al.,* 2014). The mycobacterial genome can encode up to five of these transport systems, termed ESX-1 to ESX-5; two of these systems, ESX-1 and ESX-5 are involved in virulence (Abdallah *et al*., 2007). In *Staphylococcus aureus*,

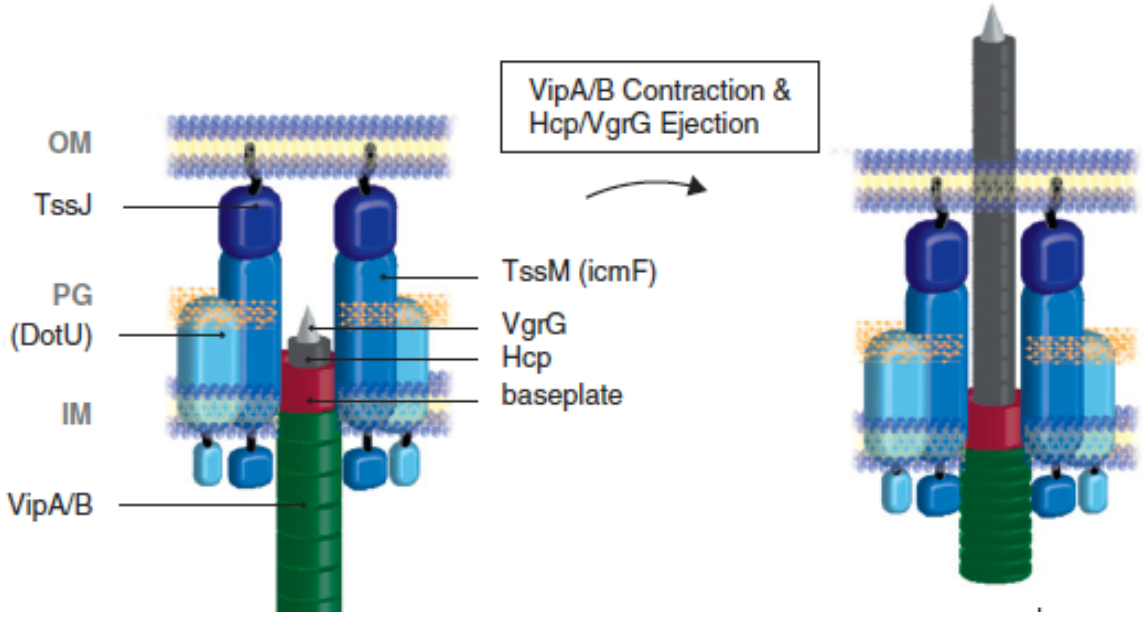


Figure 1.10. Schematic function of the T6SS. Hcp tubes are engulfed by extended VipA/VipB tubules, which act like a viral syringe. The complex is attached to the cell envelope by the membrane associated proteins: TssJ, TssM, DotU, and a baseplate like structure. Contraction of VipA/VipB causes ejection of Hcp and VgrG into the host cell. OM, outer membrane; IM, inner membrane; PG, peptidoglycan. Copied from Kapitein and Mogk., 2013.

two proteins encoded by ESAT-6 secretion system (T7SS), EsxA and EsxB share homology with the ESX-1 genes, and deletion of one gene impairs the secretion of both. In both well known human pathogens, *Mycobacterium turbeculosis* and *Staphylococus aureus*, ESX secretion was found to be crucial for infection (Anderson *et al*., 2013).

Although there is limited structural data of the ESX-1 components, it is assumed that they form a multi-subunit cell envelope spanning complex similar to that of type IV secretion systems (Abdallah *et al*., 2007). It has however been shown through protein-protein interaction studies that the 6 kDa early secreted antigenic target (ESAT-6) and the 10 kDa culture filtrate protein (CFP-10) are functionally dependent on one another as each protein forms a 2-helix hairpin held together by hydrophobic interactions (Renshaw *et al*., 2005). Both proteins have been shown to be important for the virulence of *M. tuberculosis,* as they are important T-cell antigenic targets. On the other hand, the T7SS protein EsxA and EsxB are negatively regulated by the TCS SaeRS which control a large set of secreted virulence factors as targeted by environmental cues (Anderson *et al.,* 2013). Further research in the field will continue to shed light into this unique secretion system in which all of the secreted proteins are co-dependent on each other for secretion.

1.3.7 Type VIII Secretion System (T8SS)

The final known secretion system is type eight secretion system (T8SS), which is recognized in *E. coli* as being responsible for the formation of aggregative fibers known

as curli. These fibers are primarily involved in bacterial biofilm formation and attachment to nonbiotic surfaces (Goyal *et al*., 2013). The key component of these systems is the formation of the oligomeric secretion channel, formed by lipoprotein CsgG, in the outer membrane only similar to the T2SS, and T5SSs. Curli are extracellular fibres of 4 to 7 nm in diameter, involved in cell-to-cell contacts in gram-negative bacteria (Collinson *et al*., 1991). These fibers share biochemical and structural characteristics with amyloid fibers, formed of fibrous protein aggregates, causing Alzheimer's and Parkinson's diseases (Chapman *et al*., 2002). Unlike the misfolded protein amyloids, curli result from dedicated biosynthetic pathways by the T8SS, and are considered as important amyloids for bacterial survival (Desvaux *et al*., 2009).

1.3.8 Type IV Secretion System (T4SS)

The type IV secretion system (T4SS) is found in both, gram negative and gram positive bacteria to deliver DNA and protein substrates to both prokaryotic and eukaryotic cells. This system is the most complex and versatile secretion system with a significant impact on human health; it is the major driving force for infection and the spread of antibiotic resistance (Trokter *et al.,* 2014). According to the various functions these systems carry T4SSs are divided into three groups: A) conjugation systems that translocate single-stranded DNA to recipient cells in a contact dependent manner, such as the F plasmid of *E. coli* (Lawley *et al*., 2003; Hazes & Frost, 2008); B) DNA uptake and release into the extracellular environment independently of contact with another cell, as

seen in *C. jejuni*, *H. pylori* and *N. gonorhoeae* (Alvarez-Martinez & Christie, 2009; Christie *et al*., 2005); and C) effector molecule translocation that deliver protein substrates directly into eukaryotic cells, such as tumor inducing Ti plasmid of *Agrobacterium tumefaciens* (Christie *et al*., 2005), the pertussis toxin *Bordetella pertussis* (Shrivastava & Miller, 2009) or the CagA oncoprotein of *Helicobacter pylori* (Cascals & Christie, 2003) (Figure 1.11). The ability of T4SSs to deliver proteins and nucleic acids is what makes them unique.

The most studied T4SS are the *A. tumefaciens*, a tumor inducing (Ti) system, consisting of 12 core proteins named VirB1-VirB11 and VirD4, and the *E. coli* systems encoded by the conjugative plasmids F, R1, and pKM101 with up to 20 conserved proteins (Christie, 2001; Lawley *et al*., 2003; Alvarez-Martinez & Christie, 2009). Although they deliver different factors, many of the T4SS are closely related to one-another forming a similar translocation apparatus, varying in subunit number and

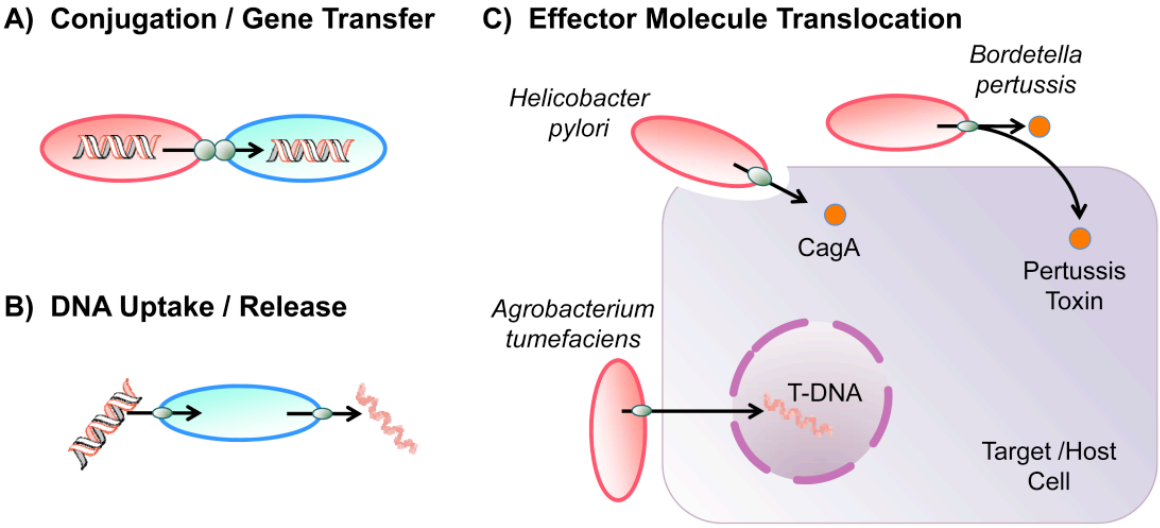


Figure 1.11. The versatile T4SS. Gram negative bacteria utilize T4SSs for a range of functions including (a) conjugation (i.e. the *E. coli* F plasmid), (b) DNA uptake and release (i.e. in *C. jejuni*, *H. pylori* and *N. gonorhoeae*) and (c) effector translocation (i.e. in *A. tumefaciens*, *B. pertussis* or *L. pneumophila*). Reproduced from Shala *et al*., 2015.

composition. Table 1.2 summarizes the conserved proteins and their homologous counterparts encoded by the three different T4SS plasmids, Ti, Cag and F.

While a few puzzle pieces of the T4SS assembly and its complex structure have been put together, there are still many unknown properties of these systems. Proteins VirB3, VirB6-B10 span both membranes forming the supporting structure for the pilus assembly and translocation (Trokter *et al*., 2014). Pili are the extracellular appendages that are thought to promote contact between donor and recipient cells, and in addition serve as a channel for ssDNA delivery, but not required for substrate secretion (Lawley *et al.*, 2003) The pilus itself is made of multiple subunits of VirB2, and the VirB5 pilus-tip adhesin, not found in all T4SS types. VirB1 as a periplasmic protein that functions by degrading the peptidoglycan layer, but is also required for pilus synthesis (Trokter *et al*., 2014). ATPases VirB4, VirB11, VirD4 located in the cytoplasm at the base of the structure power substrate delivery and pilus formation. Figure 1.12a shows a schematic view of the T4SS assembly, whereas Figure 1.12b shows the available structures and the proposed arrangement of numerous T4SS components determined by X-ray crystallography or Cryo-EM. While structural information for numerous T4SS proteins are available, there are many other proteins encoded by the F plasmid with no homologues found in other T4SSs, such as TrbC and TraW from the F plasmid of *E. coli*.

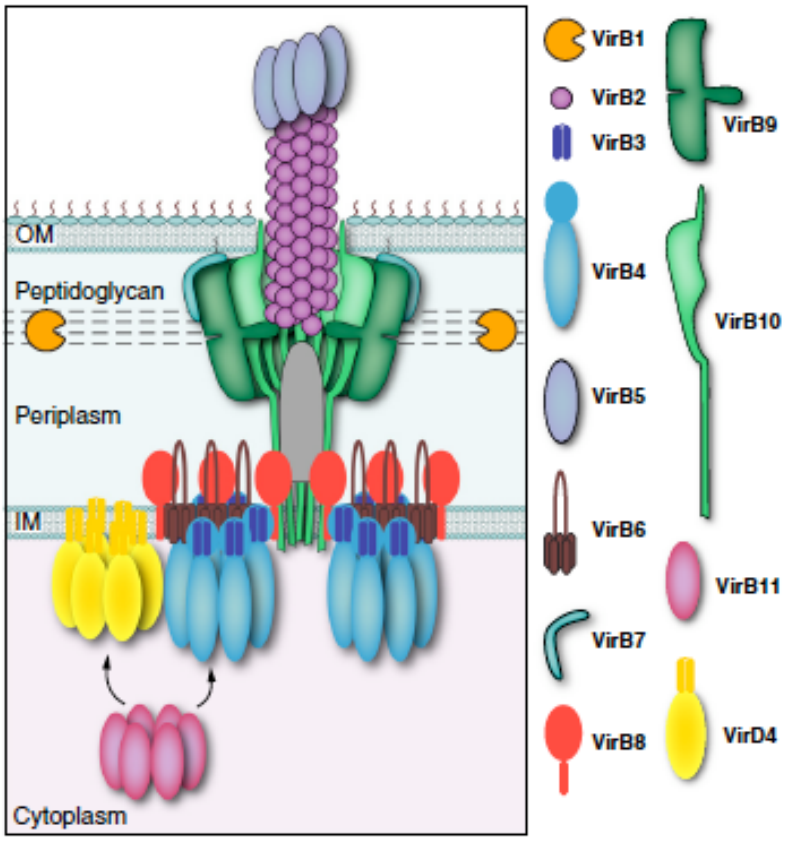
Among other DNA trafficking means, conjugation is known to have contributed to the widespread of the resistance genes, such as β-lactamase expressing genes in the presence of high antibiotic levels, within the non-resistant microbes in nosocomial and

Table 1.2. Proposed function and homologies between proteins encoded by three different T4SS plasmids from different bacteria

Function / Homology	T4SS System		
	Ti	**Cag**	**F**
Pilin	VirB2	CagC	TraA
Cyclase			
Acetylase			TraX
Lysozyme	VirB1	Cagγ	Orf169
Pore	VirB3	CagE	TraL
Secretion/ATPase	VirB4	CagE	TraC
Pore	VirB5	CagL	TraE
Pore	VirB6	CagW	TraG
Lipoprotein	VirB7	CagT	TraV
Pore	VirB8	CagV	
Secretion/pore	VirB9	CagX	TraK
Pore	VirB10	CagY	TraB
Secretion/ATPase	VirB11	Cagα	
Secretion/ATPase	VirD4		
Pore			TraF
Pore			TraH
Pore			TraW
Pore			TrbC
Pore			TraU
Adhesin			TraN
Relaxase			TraI
Transport			TraD
Relaxase			TraM
Relaxase			TraY
Accessory to relaxase			TraJ

Copied from: Lawley *et al*., 2003; Renshaw *et al.,* 2005; Fischer 2011; Trokter *et al*., 2014).

A.


OM
Peptidoglycan
Periplasm
IM
Cytoplasm
VirB1
VirB2
VirB3
VirB4
VirB5
VirB6
VirB7
VirB8
VirB9
VirB10
VirB11
VirD4

B.

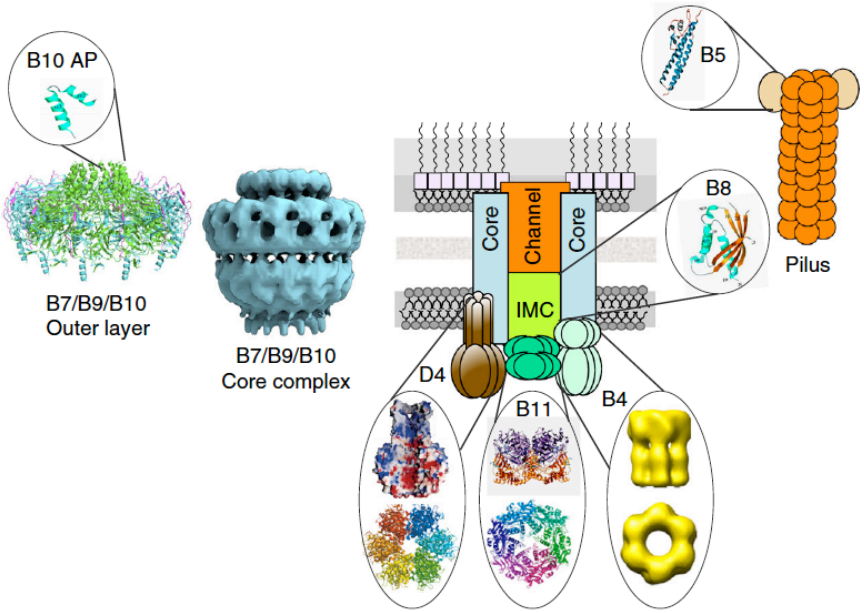

B10 AP
B7/B9/B10
Outer layer
B7/B9/B10
Core complex
B5
Pilus
B8
Core
Channel
Core
IMC
D4
B11
B4

Figure 1.12. **(A)** A schematic diagram of the T4SS system, using *A. tumefaciens* VirB/VirD4 as an example (Copied from Trokter *et al*., 2014). VirB7, VirB9, VirB10 form the core complex, and the pilus formation and extension assembles within the centre of the core complex. The complex spans through both inner and outer membranes. Other components are shown as localized in/around the complex. **(B)** The proposed combination of T4SS assembly subunits as determined by X-ray or CryoEM (Copied from Christie *et al*., 2014). The ATPases D4, B11,and B4 each forming hexameric structures (brown, green, cyan) are shown at the base of the complex; IMC-inner membrane core (lime) and the structure of its subunit B8; Channel and the pilus extension (orange) made of B2 subunits, and the structure of B5, pilus tip protein is shown. In addition the core B7/B9/B10, forming a tetradecamer, spans through both inner and outer membrane forming a cap at the outer membrane. The structure of the pore-forming B10, antennae projection (AP) protein is magnified.

other community settings (Yagi *et al*., 1997). Conjugation, as the most important process for the horizontal gene transfer between bacterial cells, requires independently replicating conjugative plasmids that encode and facilitate their own transfer as well as the transfer of other cellular DNA (Frost *et al.*, 1994). The F plasmid of *E. coli* was identified by Lederberg & Tatum almost 70 years ago (Lederberg & Tatum, 1946), and it is still the paradigm for studying bacterial conjugation. All genes necessary for F-mediated conjugation are located in the 33.3 kb transfer (*tra*) region of the ~100 kb F plasmid (Frost *et al*., 1994) The system assembly in F plasmids is composed of a DNA-processing complex, the relaxosome, that assembles on the plasmid's origin of transfer (*oriT*), and a T4SS, the transferosome, a secretion channel through which the DNA is transferred (Lawley *et al*., 2003); in addition a coupling protein exists as a link between the relaxosome and the transferosome (de la Cruz *et al.,* 2010).

The mating pair is stabilized through the interactions of $TraN_F$ and $TraG_F$, and F plasmid DNA is transferred from donor to recipient cell through a $TraB_F$-$TraK_F$-$TraV_F$ core complex (Harris *et al*., 2001) by the ATP-dependent coupling protein $TraD_F$. DNA transfer is initiated through the interaction of the relaxase TraI with the F plasmid at the origin of transfer (*oriT*), which is recruited by TraJ through a binding cleft bounded by a 5-stranded β-sheet and flexible α-helical domain (Alvarez-Martinez & Christie, 2009). Following interaction with TraI and the nicking of plasmid DNA, TraM, which binds near the *oriT*, is recruited. TraM contains a core eight helical bundle which interacts with the C-terminal domain of TraD (Lu *et al*., 2008) for ATP-dependent transfer of the ssDNA through the core T4SS complex for transfer. F plasmid transfer is regulated by a

few known environmental stressors, such as the Cpx system, which regulated response to extracytoplasmic stress, the cyclic AMP receptor protein Crp, and leucine-responsive regulatory protein Lrp that regulate gene expression upon nutritional availability (Frost & Koraimann, 2010). Additional modes of F plasmid regulation is done by the FinO-FinP fertility inhibition system, which regulates TraJ expression. FinP depends upon action of the plasmid encoded FinO, in F-like systems *finO* gene is interrupted, therefore transfer to F containing plasmid cells is interrupted (Gubbins *et al*, 2003). While structures of the main scaffolding proteins of the F conjugative plasmid are known, there are still many unclear events, many other unknown proteins necessary for the assembly of the F encoded plasmid into the ssDNA injection syringe.

1.4 Thesis Overview

Both TCS and T4SS are important means acquired by bacteria to ensure their survival and adaptation in stressful or new environments. While both of these systems have evolved independently, there is a link between the two, where one is required for the activation of the other. For example in *Bartonella henselae*, pH restricted expression of VirB/D4 T4SS is regulated by the TCS BatR/BatS, which regulate genes necessary for the T4SS expression (Québatte *et al*., 2013). In another example the TCSs EnvZ/OmpR and GrrS/GrrA from *Erwinia amylovora*, besides regulating polysaccharide production and motility, these systems are found to down-regulate flagellar gene expression as encoded by the T3SSs in *E. coli, Salmonella typhimurium* and other bacteria (Li *et al*.,

2013). In addition the phenotypes of the phototaxis mutants suggest that Che-like polypeptides control both cell orientation with respect to light and pilus biosynthesis (Bhaya *et al.,* 2001). These systems are connected by numerous examples, although the process of either is not well established learning more about one could guide us to understanding the other.

In Chapter 2 of this dissertation, the X-ray structural determination of the TCS LytR receiver domain from *Staphylococcus aureus*, in both native and Beryllium fluoride complex forms is detailed. In Chapter 3, interaction studies between TrbC and TraW, two important components for the F T4SS pilus assembly, will be discussed. In Chapter 4, the expression, purification and crystallization trials of TrbC and TraW from the F plasmid T4SS of *Escherichia coli* will be examined. Concluding remarks and a discussion of future directions is in Chapter 5, and two appendices on X-ray crystallographic methods (Appendix A), and on relevant protein sequence details (Appendix B) are presented.

Chapter 2

Structural Analysis of the Apo and Beryllium Fluoride bound N-terminal Receiver Domain of LytR from *Staphylococcus aureus*

2.1 Introduction

Two-component systems (TCSs), that function as phosphorylation pathways, have enabled pathogenic bacteria to respond and adapt to changing environmental conditions. These signal transduction systems in addition to regulating gene expressions they can also regulate protein-protein interactions (Galperin, 2010). The simplest type of TCS contains a membrane bound sensory histidine kinase (HK), and a two-domain response regulator (RR) while more complicated systems can include multiple phosphotransfer reactions (Toro-Roman *et al*., 2005). The diversity of these systems is reflective of their diverse RRs. While most RRs share a similar phosphoacceptor, receiver domain (N-terminal domain), they differ and are classified into subfamilies according to their DNA-binding effector domain (C-terminal domain) (Galperin, 2010). The most abundant transcriptional regulators belong to the OmpR/PhoB family, characterized with a winged helix-turn-helix (*w*HTH) DNA binding domain (Buckler *et al*., 2002; Galperin, 2010). The second most abundant family is NarL/FixJ, characterized by a typical HTH domain (Noriega *et al*., 2010; Galperin, 2010). Both of these types of transcriptional regulators are further divided into subfamilies based on the similarities of their output domains covering 95.3% of the characterized RRs. A much different family of transcription

factors is the AlgR/AgrA/LytR family, characterized by a non-HTH DNA binding domain consisting of mainly β-strands, named the LytTR domain, covering the remaining 4.7% of the DNA-binding response regulator proteins (Nikolskaya & Galperin, 2002; Sidote *et al.,* 2008) involved in regulation of important virulence factors and toxin production in pathogenic bacteria (Rood, 1998; Lizewski *et al*., 2002; Novick, 2003; Dawid *et al*., 2007; Galperin, 2008).

The variability of TCS in bacteria, and their response to the stress induced by the overuse and often misuse of antibiotics may be responsible for the emergence of the methicillin-resistant strains of *Staphylococcus aureus* (MRSA) (Ippolito *et al.*, 2010). The two component system LytSR, consisting of LytS and LytR, has been recently proposed to function as a sense-response system for detecting subtle changes in the cell's transmembrane potential (Peschel *et al.,* 2001; Patton *et al.,* 2006; Li *et al.,* 2007; Sharma-Kuinkel *et al.,* 2009). Changes in the transmembrane potential may be a result of the release of numerous cationic antimicrobial peptides (CAMPs) by the human host cells defence mechanisms (Peschel, 2002; Voyich *et al.*, 2005). In a typical TCS, the sensor HK, upon detection of transmembrane potential decrease (Sharma-Kuinkel *et al.,* 2009), transduces the signal intracellularly by transferring a phosphate to the conserved Asp residue of its cognate receiver domain of the RR (Gao *et al*, 2007; Gao & Stock, 2010). The phosphorylation initiates a series of conformational changes in the RR that often leads to the dimerization of the RR, which may result in a tighter binding of the effector domain to the target promoter, that ultimately modulate the output response of the proteins (Gao & Stock, 2010). LytR is a transcription factor that directly regulates the

lrgAB operons, thus affecting *cidABC* operon (Brunskill & Bayles, 1996), both of which are involved in the control of programmed cell death and cell lysis (Groicher *et al.,* 2000; Rice *et al.,* 2003). The gene products of the *cid* operon enhance murein hydrolysis activity (Rice *et al.,* 2003) and antibiotic tolerance, while the *lrg* genes inhibit these processes (Groicher *et al.,* 2000). Sequence alignment of the receiver domains from different families of RRs using Clustal Omega (Sievers *et al*., 2011) shows the most conserved key residues involved in the activation of the protein such as: Asp7, Asp8, Asp53, Asn55, Met57, Thr81, His83, Tyr98, Lys101, Pro102 highlighted in black, using numbering for LytR (Figure 2.1). However, variability in the residues involved in the dimerization interface, involving α4-loopα4β5-α5, such as Ala87, Val88, Phe91, Glu92, Asn94, Arg107, Gln110, and Lys114, shown by the residues highlighted in grey for LytR (Figure 2.1), indicates that despite the conserved activation of the proteins, they employ different strategies to regulate their target genes. Moreover, considering the diversity of RRs there is a significant variability in the primary structure of receiver domains, thus limiting the cross-talk among the receiver and effector domains of the two-component systems (Barbieri *et al.,* 2010).

Currently, structures for three members of AlgR/AgrA/LytR subfamily have been reported: LytTR DNA-binding domain of AgrA from *Staphylococcus aureus* bound or unbound to DNA (PDB ID: 3BS1 and 4G4K) (Sidote *et al*., 2008, Leonard *et al*., 2012), the DNA-binding domain of a putative methyl-accepting/DNA RR from *Bacillus cereus* (PDB ID: 3D6W), and the full length $ComE^{D58A}$, in addition to the receiver domains

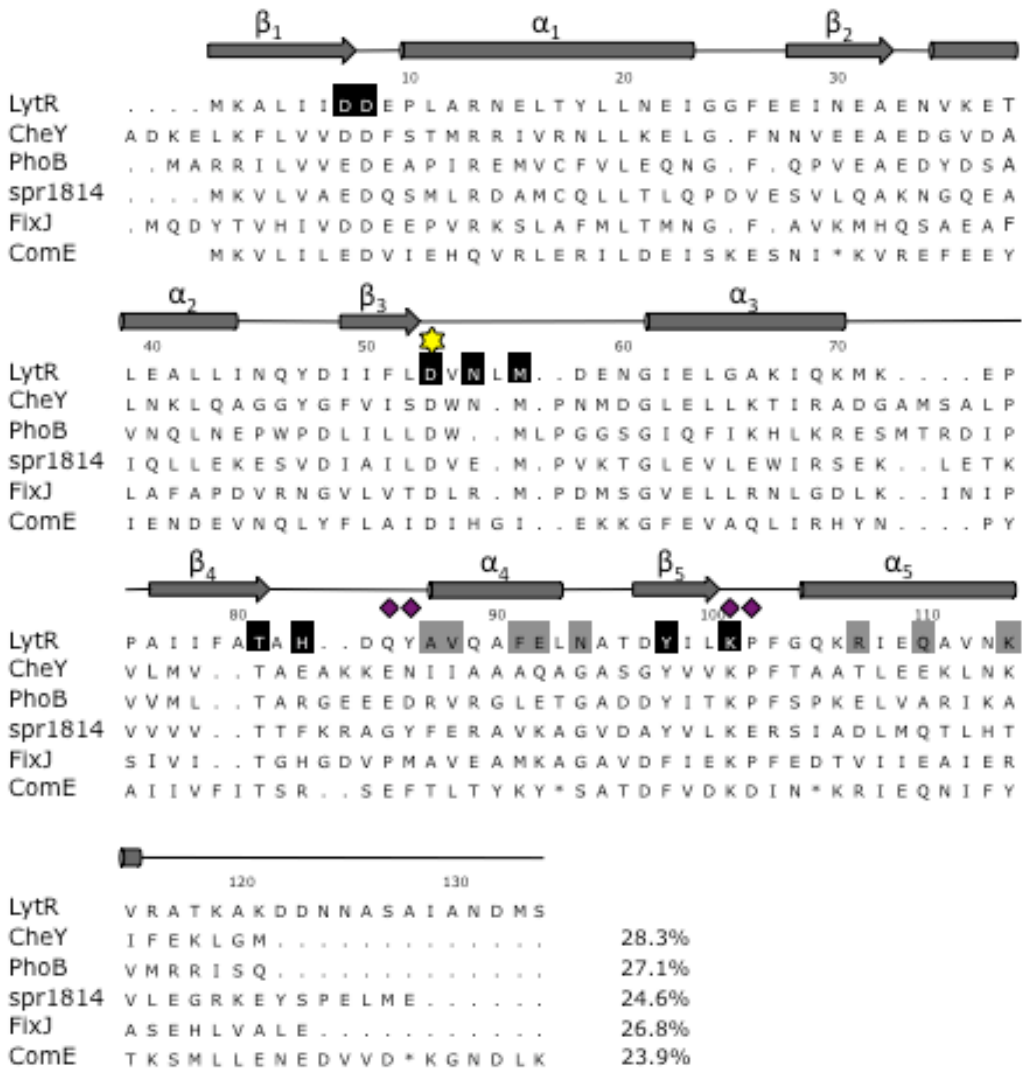


Figure 2.1. Sequence alignment of $LytR^{N}$ with the following receiver domains of other known response regulators: LytR (this work), CheY (PDB ID: 3CHY; Volz & Matsumura, 1991), PhoB (2IYN; Sola *et al*., 2006), spr1814 (4E7O; Park *et al*., 2012), FixJ (1DBW; Gouet *et al*, 1999), ComE (4MLD; Boudes *et al*., 2014). β-strand arrows and α-helix sheets correspond to LytR secondary structure. The star (yellow) indicates the conserved Asp, Asp53 in LytR, which gets phosphorylated by LytS. Highlighted in reverse video are residues involved in the activation and stabilization of the active site. Highlighted in grey are the residues that are involved in the dimerization interface. Diamonds (purple) indicate residues that are affected by the activation/dimerization of the protein; star (*) not aligned residues. Sequences were aligned using Clustal Omega (Sievers *et al.,* 2011), and the percentage sequence similarity relative to LytR is shown at the end.

REC^{D58A}, REC^{D58E} of ComE from *Streptococcus pneumoniae* (PDB ID: 4CBV, 4ML3, and 4MLD) (Boudes *et al.,* 2014). While the receiver domain of different RRs has been found to exist in equilibrium between the active and inactive conformations, commonly the phosphorylation of these domains on the conserved Asp residue by the HK, favors the active state of the protein (Gardino & Kern, 2007).

Here we report the structural analysis of the LytR receiver domain, both in its apo form (apo-$LytR^N$) and in complex with beryllium fluoride (BeF-$LytR^N$). Beryllium fluoride (BeF_3^-), a noncovalent complex that resembles the phosphoryl group, when bound to the oxygen atom of an Aspartate residue (Chabre 1990; Yan *et al*., 1999; Cho *et al*., 2001), is routinely used to secure RRs in their active conformations for structural characterization. Previous studies in solution and utilizing native polyacrylamide gel electrophoresis indicate that while apo-$LytR^N$ remains monomeric, the acetyl phosphate activated protein forms dimers (Patel, 2014), suggesting that BeF-$LytR^N$ should be dimeric thus facilitating an investigation of the mechanism of dimerization by $LytR^N$. The structure of $LytR^N$ was determined to consist of the general fold of a α/β doubly wound pattern (Richardson & Richardson, 1990; Solà et al., 1999), consisting of a central five-stranded parallel β-sheet surrounded by five α-helices; helices 2, 3, 4 on one side and 1 and 5 on the other. $LytR^N$, in complex with BeF_3^-, was observed to dimerize utilizing similar dimerization interface as the one observed in ComE, from the same LytTR family (Boudes et al., 2014).

2.2 Materials and methods

The methods herein describing the purification of $LytR^N$ and the preparation of the BeF-$LytR^N$ complex were primarily done at 4 °C in the laboratory of Prof. Dasantila Golemi-Kotra. All protein crystallization experiments, X-ray diffraction, structure solution, refinement and analysis was done in the laboratory of Prof. Gerald F. Audette.

2.2.1 Expression and Purification of $LytR^N$

The expression and purification protocols for $LytR^N$ were initially developed by Kevin H. Patel in the lab of Dr. Golemi-Kotra (Patel, 2014). To express $LytR^N$, the pET26b::$LytR^N$ vector was previously introduced into *E. coli* BL21 (DE3) cells (Patel, 2014). One mL of a 5 mL overnight culture containing a streak of these cells was used to inoculate 1 L of Luria-Bertani (LB) broth supplemented with 50 μg/mL Kananmycin (Kan). Cells were cultured at 37 °C and 200 rpm to mid-log phase growth (OD_{600}=0.5-0.8). Prior to induction with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 25 °C for 12 hrs, the culture was cooled to 4 °C. The cells were harvested by centrifugation at 3,300 x g for 20 min at 4 °C.

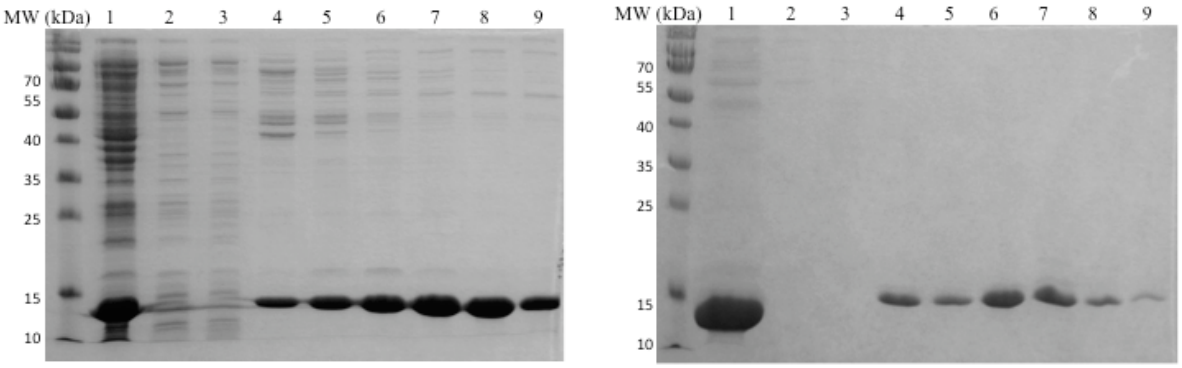
To purify $LytR^N$, the cell pellet was resuspended in 1:10 (w/v) lysis/loading buffer [20 mM Tris pH 7.5, 5mM $MgCl_2$]. The resuspended cells were lysed by sonication (Sonic Dismembrator 500, Fisher Scientific) while cooling on ice for 5 min (10s on/15s off). The insoluble cell lysate was removed by centrifugation at 34,000 x g for 30 min at 4 °C. The supernatant was loaded onto a pre-equilibrated 40 mL DEAE column (GE Healthcare, in Dr Golemi-Kotra's lab), and the column was subsequently washed with at

least five column volumes of the loading buffer to remove any contaminants. The bound protein was eluted over ten column volumes in a linear gradient to a final concentration of 100 % elution buffer [500 mM Tris pH 7.5, 5mM $MgCl_2$] at a flow rate of 3 mL/min, where the protein elutes with 50-70% of elution buffer (250-350 mM Tris) as indicated by the protein band corresponding to $LytR^N$ (molecular weight (MW) of 15,028 Da) (Figure 2.2a). $LytR^N$ elution peaks were pooled and concentrated by centrifugation (1,240 x g at 4°C) using an Amicon Ultra-3 concentrator (Millipore) to a final volume of 5 mL. $LytR^N$ was then further purified using ÄKTA Purifier at 4 °C (GE Healthcare, in Dr. Golemi-Kotra's lab) and buffer exchanged into crystallization buffer [100mM Tris pH 7.5, 5 mM $MgCl_2$] by size exclusion chromatography (SEC) on a HiPrep 26/60 Sephacryl S-200 HR gel-filtration column (GE Healthcare, in Dr. Golemi-Kotra's lab) (Figure 2.2b). Fractions corresponding to $LytR^N$ were collected and concentrated to 13 mg/mL prior to crystallization. The homogeneity of the protein was determined by Coomassie blue staining of an 15% SDS polyacrylamide gel (Figure 2.2).

2.2.2 Preparation of $LytR^N$ complex with BeF_3^-

Previous studies have shown that phosphorylation induces $LytR^N$ dimerization (Patel, 2014), and BeF_3^- a noncovalent complex that resembles the phosphoryl group when bound to the side chain oxygen atom of an Aspartate residue is routinely used as a phosphorylation analog to secure RRs in their active conformations for structural characterization (Chabre, 1990; Yan *et al.*, 1999; Cho *et*

al., 2001). The generation of BeF-LytR^N was achieved by adding 5.3 mM $BeCl_2$, 33 mM NaF and 1 mM $MgCl_2$ to the purified LytR^N (0.5 mg/mL initial concentration) in an ultracentrifugal filtration device (5 kDa MWCO) (Millipore). This solution was concentrated (3,500 x g / 4°C) to a final protein concentration of 7 mg/mL, which was then loaded into a 15% native–polyacrylamide gel electrophoresis (PAGE) to investigate protein oligomerization (Figure 2.3), and used for crystallization (Section 2.2.4). Native-PAGE analysis of the BeF-LytR^N complex was performed as follows: purified protein (0.47 mM) was mixed with 2X loading dye (40 mM Tris, pH 6.8, 80 mM DTT, 0.08% Bromophenol blue, 10% glycerol) at 2:1, 1.5:1, and 1:1 dilution. The native gel prepared contains a 15% Acrylamide-Bis acrylamide (19:1) resolving gel, and a 4% stacking gel. After loading the protein samples, they were electrophoresed, at 100 V for 4 hrs at 4°C, using a native running buffer [85 mM Tris, pH 8.3, and 33 mM glycine], and stained with Coomassie blue stain for visualization of the protein bands (Patel, 2014). When the protein was diluted 2:1 (10 µL protein + 5 µL dye = 0.31 mM protein) the smearing was visible, contrarily to 1.5:1 (0.28 mM) or 1:1 (0.24 mM) dilution of the protein with the gel loading dye. Since a specific amount of NaF and $BeCl_2$ is required for the formation of BeF_3^-, the dilution of these components in the solution does not allow for the complex formation. Alternatively, these results could be due to the dilution of the protein concentration. Considering that activation/dimerization of RRs is a revesible process and the proteins exist in equilibrium between the two states, it is most likely that the LytR^N in this case oligomerizes in the presence of BeF_3^- rather than just due to the higher protein concentrations.



A **B**

Figure 2.2. Purification of $LytR^N$ as analyzed by SDS-PAGE using a 15% gel. (**A**) Samples analyzed after purification with DEAE column: Lane 1, cytosolic fraction; lane 2, flough through; lane 3, wash of unbound; lanes 4-9 fractions during 50-70% of elution buffer (Patel, 2014). (**B**) Samples analyzed after purification with SEC: Lane 1, concentrated fractions from DEAE column (lanes 6-8, in figure a); lanes 2-3, elution with 75-125 mL of 100 mM Tris volume; lanes 4-9, elution of $LytR^N$130-175 mL volume of buffer.

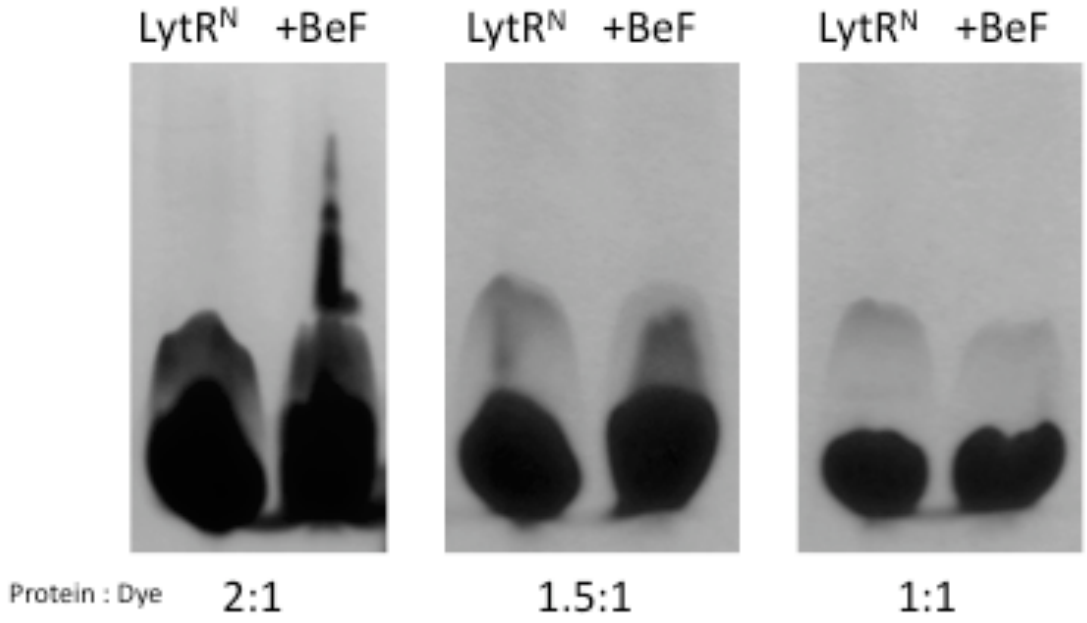


Figure 2.3. Native-PAGE analysis of the apo-LytR^N and BeF-LytR^N, both at the same concentration of 0.47 mM. At 2:1 (protein:dye; 0.31 mM) dilution the BeF protein smears off the gel indicating oligomerization of the protein; when both samples were diluted to 1.5:1 (0.28 mM) this smearing of the +BeF sample is less obvious, but still there; with the 1:1 (0.24 mM) dilution of both samples there is no difference between the two, indicating that the oligomerization of the protein +BeF is abolished, when the sample is further diluted.

2.2.3 Crystallization of apo-LytR^N

The crystallization and initial diffraction studies of apo-LytR^N have been reported previously (Shala *et al*., 2013). Briefly, initial crystallization experiments were performed by screening conditions in commercially available kits Crystal Screens 1 and 2 from Hampton Research (Jancarik & Kim, 1991), and JCSG II from Qiagen (Lesley & Wilson, 2005). The trials were set in 96-well sitting drop plates (Axygen) by hand with 1 μL protein solution concentrated to 6-13 mg/mL, and mother liquor in a 1:1 ratio over a reservoir containing 70 μL mother liquor. The crystal trays were stored at 22 °C. These trials yielded a number of hits producing numerous thin needle-like crystals (Figure 2.4a) from various conditions, most of which contained ammonium or lithium sulfate as the precipitant, within 1 day. The most promising initial crystals were obtained from 0.1 M MES monohydrate pH 6.5, 1.6 M Magnesium sulfate heptahydrate. Optimization of the crystallization conditions was performed by varying the precipitant concentration and the buffer pH using hanging drop EasyXtal 15 well plates with drop guard crystallization supports (Qiagen) and incubating the plates at a lower temperature of 4 °C. Better diffracting crystals grew using 7 mg/mL protein mixed with an equal amount of reservoir solution [0.1M MES monohydrate pH 6.5, 1.6 M magnesium sulfate heptahydrate] equilibrated against 0.5 mL of the reservoir solution at 4 °C over 6 days. The single crystals grew into rods with maximum dimensions of 410 x 166 x 10 μm^3 (Figure 2.4b) diffracting up to 2.3Å (Figure 2.4c) (Shala *et al*., 2013).

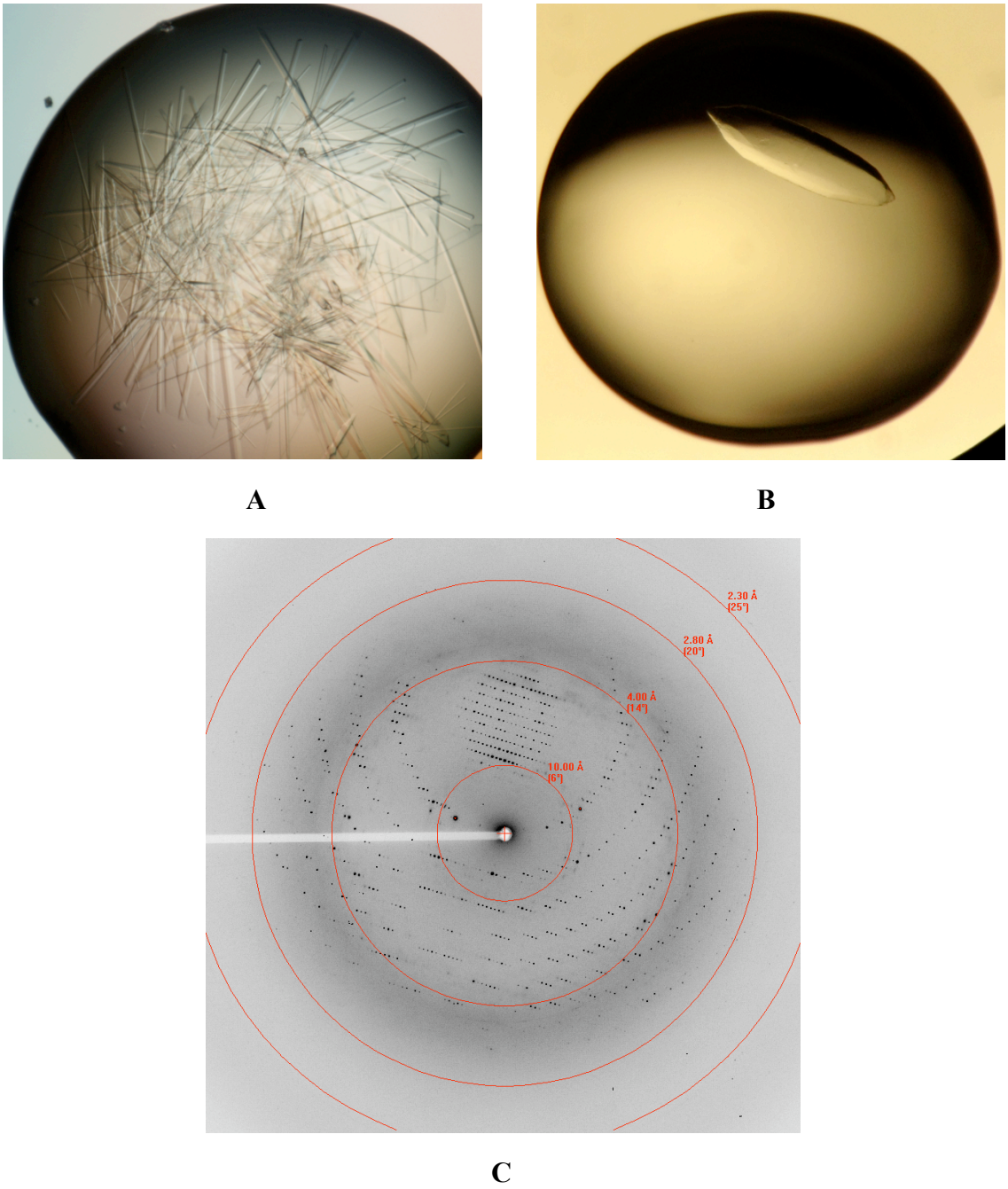


Figure 2.4. Crystals of $LytR^{N}$ (**A**) Initial crystals formed at room temperature looking as thin-like needles, that diffract to ~12 Å. (**B**) Crystals formed after refining against concentration of the precipitant and the pH with the dimensions of 410 x 166 x 10 μm^3, at 4°C. (**C**) Diffraction image of $LytR^{N}$. Resolution rings showing up to 2.30Å.

2.2.4 Crystallization of $LytR^N$ complexed with BeF_3^-

The BeF-$LytR^N$ complex was formed by adding 5.3 mM $BeCl_2$, 33 mM NaF and 1mM $MgCl_2$ to 7 mg/mL apo-$LytR^N$ (previously purified and SEC buffer exchanged into 100 mM Tris, pH 7.5) in an ultracentrifugal filtration device with a molecular mass cutoff of 5 kDa (Millipore) using a protocol initially developed by Prof. Dasantila Golemi-Kotra. Initial crystallization trials for BeF-$LytR^N$ were performed using the commercially available Crystal Screen and Crystal Screen 2 (Hampton Research) and the JCSG Core Suite II (Qiagen) kits. Similar conditions as for the native $LytR^N$ crystallization resulted in BeF_3^--$LytR^N$ complex crystals. The protein complex formed unstable crystals in numerous conditions with ammonium, lithium or magnesium sulfate after 1 day of incubation and diffracting only to a resolution of 8 Å. However, after optimization of the protein concentration, the pH and precipitant concentration, better crystals more sutable for X-ray diffraction formed in 2 to 3 days, with the dimensions 300 x 130 x 5 μm^3 diffracting to 3.5 Å. Crystallization of BeF-$LytR^N$ in this form was achieved by mixing equal volumes of protein (7 mg/mL $LytR^N$ in 100mM Tris-HCl, pH 7.5, 33mM NaF, 5.3 mM $BeCl_2$, and 1 mM $MgCl_2$) and reservoir solution (1.3 M $MgSO_4 \cdot 7H_2O$, 0.1 M MES, pH 5.55) and incubating the plates at 4 °C (Figure 2.5).

To prepare the crystals for data collection, crystals were cryoprotected by soaking them in a new drop of reservoir solution supplemented with 20% (v/v) glycerol for 45-60 s prior to vitrification in liquid nitrogen, in the cold room (4 °C). Crystals were initially screened at home, after which they were shipped to the Canadian Macromolecular

Crystallography Facility (CMCF) at the Canadian Light Source (CLS), where they diffracted to a resolution of 3.5 Å.

2.2.5 Data Collection for apo-LytR^N and BeF-LytR^N

X-ray diffraction data for both LytR^N and BeF-LytR^N complex were collected at the CMCF at the CLS. Diffraction data for the inactive and BeF_3^- protein were collected on CMCF Beamline 08ID1-1 at 100 K and a wavelength of 0.9795 Å. The data were processed using MOSFLM (Leslie & Powell, 2007) and integrated and scaled using AIMLESS and CTRUNCATE (Evans, 2011; Padilla & Yeates, 2003) from the CCP4 program suite (Winn *et al*., 2011; SERC, 1994). The native crystals belong to the space group $P6_122$ with unit cell dimensions as follows: a = b = 84.82, c = 157.3 Å, α = β = 90°, γ = 120° with a Matthew's coefficient (V_M)= 2.77, corresponding to one molecule per asymmetric unit. Crystals of BeF-LytR^N belong to $C222_1$ with unit cell dimensions: a = 86.76, b = 231.91, and c = 158.35 Å, α = β = γ = 90° and V_M = 1.70, corresponding to four LytR^N molecules per asymmetric unit. Diffraction data collection statistics are listed in Table 2.1.

Figure 2.5. Crystal of BeF-LytR^{N} complex protein, formed in 3 days at 4 °C with the dimensions of 300 x 130 x 5 μm, and a different morphology than the native LytR^{N}.

Table 2.1. Summary of Data Collection Statistics

	Apo-LytR^N	**BeF-LytR^N**
Space group	$P6_122$	$C222_1$
Unit-cell parameters (Å, °)	a = b = 84.82, c = 157.3 α = β = 90, γ = 120	a = 86.76, b = 231.91, c = 158.35 α = β = γ =90
Resolution range (Å)	66.56–2.34 (2.44–2.34)	65.39-3.5 (3.83-3.50)
Total No. of observations	310181 (32139)	99722 (23649)
No. of unique reflections	14816 (1619)	20588 (4839)
Completeness (%)	100 (100)	99.9 (100)
Average $I/\sigma(I)$	14.3 (1.5)	7.2 (1.7)
R_{meas}[a]	0.114 (2.573)	0.187 (1.712)
$R_{p.i.m.}$[b]	0.025 (0.571)	0.079 (0.711)
$CC_{1/2}$	0.997 (0.937)	0.997 (0.607)
B Wilson (Å^2)	72.2	283.4
Multiplicity	20.9 (19.9)	4.8 (4.9)
Mosaicity	0.37	0.75

[a]$Rmeas = \sum_{hkl}\{N_{hkl}/[N_{hkl}- 1]\}^{1/2} \sum_i|I_{hkl,i} - \langle I_{hkl,i}\rangle| / \sum_{hkl}\sum_i I_{hkl,i}$.
[b]$Rp.i.m. = \sum_{hkl}\{1/[N_{hkl} - 1]\}^{1/2} \sum_i|I_{hkl,i}-\langle I_{hkl}\rangle| / \sum_{hkl}\sum_i I_{hkl,i}$, where $I_{hkl,i}$ and $\langle I_{hkl}\rangle$ represent the diffraction-intensity values of the individual measurements and the corresponding mean values, respectively.

2.2.6 Structure Determination of apo-LytR^N

The structure for the apo-LytR^N was determined using data from 42.7-2.34 Å by molecular replacement with Phaser (Storoni *et al.,* 2004) integrated in the CCP4i interface, using the structure of MrkE from *Klebsiella pneumoniae* (PDB ID: 2QV0) as the search model. The SigmaA-weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps revealed the position of a sulfate ion coordinated by the active site residues. After several cycles of rigid body and restrained refinements using REFMAC (Murshudov *et al*., 2011), TLS refinement and COOT (Emsley *et al*., 2004) for visual inspection and model building, the final model was refined to 2.34 Å with R/R_{free} values of 0.209/0.25, respectively. The final model has 960 non-hydrogen atoms with 13 water molecules and 1 sulfate ion, and refinement statistics are summarized in Table 2.2. The average B-values for the main chain and side chain are 89.7 and 98.3 (Å^2), respectively, and for the water and the sulfate ions are 94.3 and 132.7, respectively. While these values seem high, there are other reported structures in the PDB with similar B factor values and resolution, such as the full length ComE from the LytTR family of RRs (PDB ID: 4CBV; Boudes *et al*., 2014). In addition, the large observed B-Wilson value from the data reduction (Table 2.1) suggests relative disorder in the crystal, resulting in higher overall B-values in the final structural model. The residues and ordered waters were built using the automated model-building program ARP/wARP (Cohen *et al*., 2008) followed by manual inspection in COOT. Residues 119-134 at the C-terminus of the protein were not modeled due to the missing electron density for those residues. The data quality and the stereochemistry of the final model were validated using SFCHECK (Vaguine *et al*.,

1999), PROCHECK (Laskowski *et al.,* 1993), and EBI-PISA (Krissinel & Henrick, 2007). PROCHECK analysis indicates that 97.4% of the residues appear in the most favorable regions of stereochemical space, with the remaining 2.6% being in allowed regions (Figure 2.6).

2.2.7 Structure Determination of BeF-LytR^N

The structure for the BeF_3^- complex with LytR^N was determined using data from 14.0-3.50 Å by molecular replacement using the structure of the apo-LytR^N as the search model. With four molecules per asymmetric unit, and due to the low resolution, automatic non-crystallographic symmetry (NCS) restraints were used during refinement. The model was built by iterative cycles of NCS-restraints automatically determined by REFMAC and manual rebuilding using COOT using SigmaA-weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps after every cycle. The final model was refined to 3.50 Å and had R/R_{free} values of 0.178/0.20 (refinement statistics summarized in Table 2.2). The final model of four monomers, chains A, B, C and D of BeF-LytR^N, in the asymmetric unit contains 3718 atoms, with 4 BeF_3^- ions, and 3 Mg^{2+} ions. The average B-values (main chain / side chain) for chains A, B, C, and D are 130.4/148.1, 129.2/147.8, 134.3/152.3, and 180.3/194.3 (Å^2), respectively, and for the BeF_3^- and Mg^{2+} ions are 143.4 and 114.9 (Å^2), respectively. Due to the overall lower density observed in chain D, the Mg^{2+} ion for that chain could not be modeled. In addition residues E9, R14, L56, M57, D58, and E59 are missing in chain D, with residues, Y18, K36, K72 and E73

missing side chains. The C-terminus of $LytR^N$, which is part of the linker region between the $LytR^N$ and $LytR^C$ (residues 119 – 134) was not modeled in any chain due to the poor electron density. The data quality and the structure stereochemistry were analyzed using SFCHECK (Vaguine *et al*., 1999), PROCHECK (Laskowski *et al.,* 1993), and EBI-PISA (Krissinel & Henrick, 2007). From the PROCHECK analysis, 97.1% of the residues appear in the most favorable regions and 2.9% occur in the allowed regions (Figure 2.7).

2.2.8 Accession Numbers

The coordinates and structure factors for the apo- and BEF-$LytR^N$ structures have been deposited in the Protein Data Bank with accession codes 4O40 and 4PQM, respectively.

2.3 Results and Discussion

2.3.1 Overall Structures

The structures of the N-terminal receiver domain of LytR, without and with the phosphoryl group analog BeF_3^- (apo-$LytR^N$ & BeF-$LytR^N$) were determined at 2.34 Å and 3.5 Å resolution, respectively. The apo- receiver domain of LytR, (apo-$LytR^N$ residues 1-134), crystallized in the hexagonal space group, $P6_122$ with one molecule in the asymmetric unit (Figure 2.8a). The structure of the protein was determined by molecular (MR) replacement using the regulatory domain of MrkE from *Klebsiella*

pneumoniae (PDB ID: 2QV0) as a search model, and refined at 2.34 Å (refinement statistics summarized in Table 2.2). The stereochemical quality of the final model was analyzed using PROCHECK (Laskowski *et al*., 1993); 97.4% of the residues appear in the most favorable regions of stereochemical space, with the remaining 2.6% being in allowed regions (Figure 2.6). The BeF_3^- complex receiver domain of LytR, (Bef-LytR^N residues 1-134), crystallized in the orthorhombic space group, $C222_1$ with four molecules in the asymmetric unit, chains A, B, C, and D (Figure 2.8b). The structure of the protein was determined using molecular (MR) replacement with apo-LytR^N as a search model, and refined to a resolution of 3.5 Å (refinement statistics summarized in Table 2.2). Stereochemical quality analyzed using PROCHECK indicated that 97.1% of the residues appear in the most favorable regions and 2.9% occur in the allowed regions of stereochemical space (Figure 2.7).

The LytR receiver domains (Figure 2.8) displays the typical α/β fold and a $(\beta\alpha)_5$ arrangement as observed in majority of other RRs (Park *et al*., 2012; Milani *et al*., 2005), which is built by five parallel β strand, assembled in a central β sheet and surrounded by five α helices: α2, α3, and α4 on one side, with α1 and α5 on the other. While the apo-form of LytR^N forms crystallographic dimer with another molecule from a different unit cell, the BeF-LytR^N with four molecules in the asymmetric unit forms two non-crystallographic dimers, all with the same dimerization interface using α4-loopα4β5-α5 region.

Table 2.2. Summary of Refinement Statistics

	Apo-LytR^N	**BeF-LytR^N**
Resolution range (Å)	42.7-2.34	14.0-3.50
Reflections used	13967	18683
R_{work}[a]	0.209	0.178
R_{free}[b]	0.25	0.20
Average B-factors (Å^2)	89.6	99
No. atoms	960	3718
Water	13	0
Sulfate	1	0
BeF_3^-	0	4
Magnesium	0	3

[a]$R_{Work} = \sum_{hkl}||F^{obs}_{hkl}| - |F^{calc}_{hkl}|| / \sum_{hkl}|F^{obs}_{hkl}|$

[b]R_{Free} is the cross validation R factor using 5% of reflections

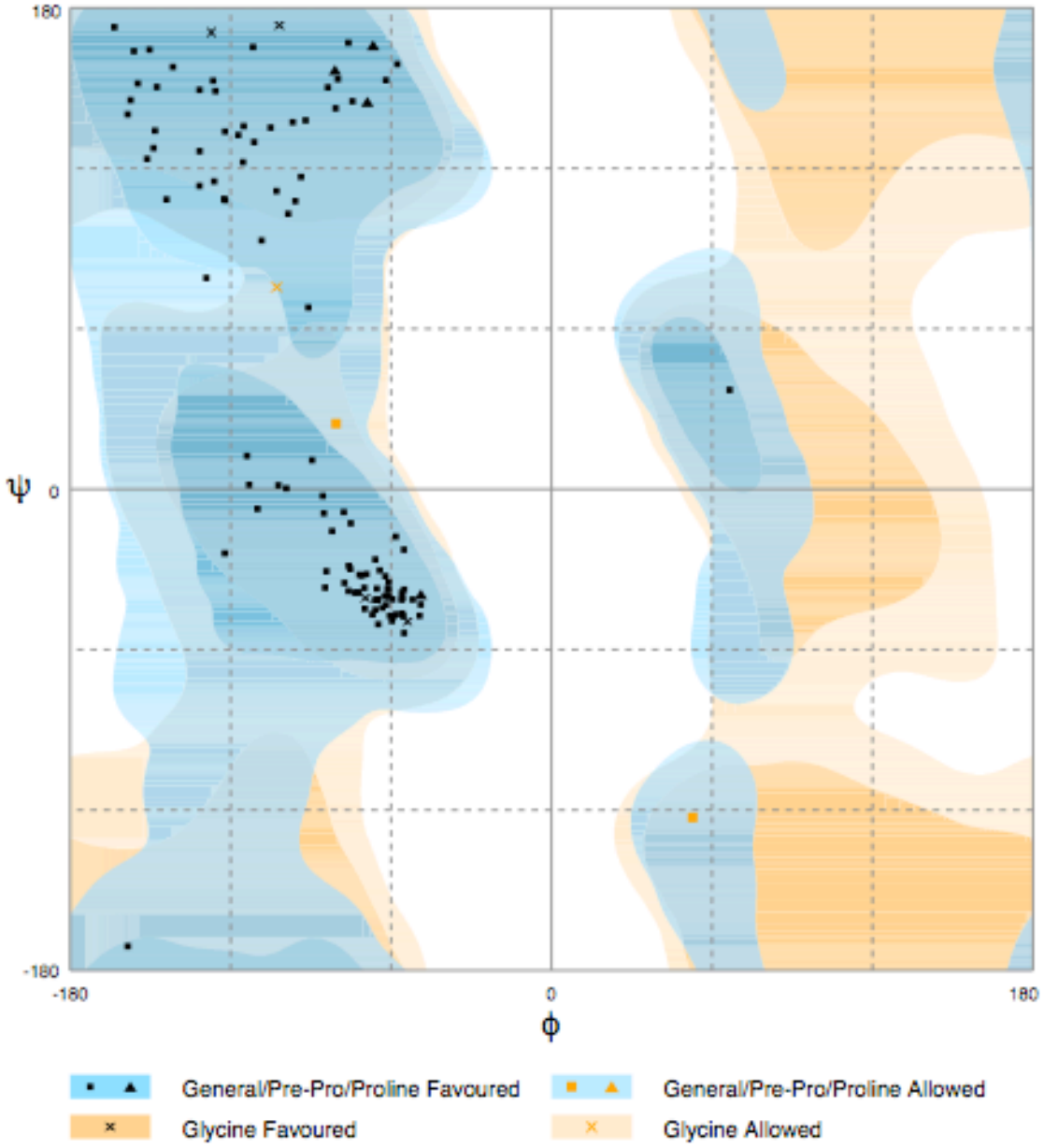


Figure 2.6. Ramachandran plot (Ramakrishnan & Ramachandran, 1965) of the apo-LytR^{N} structure. One hundred and thirteen (97.4%) of residues appear in favoured region; three (2.6%) in allowed region; 0 residues in outlier region.

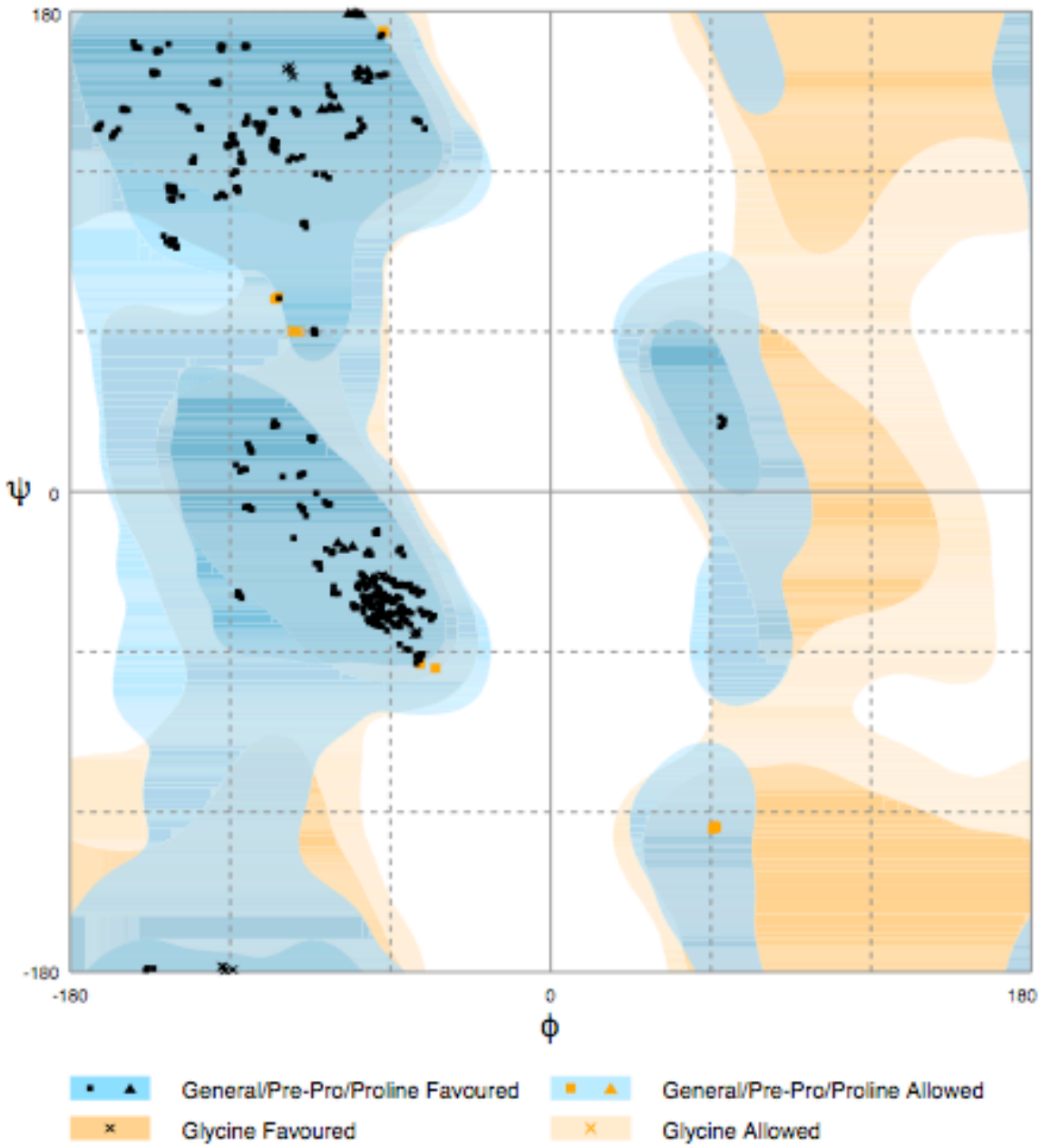
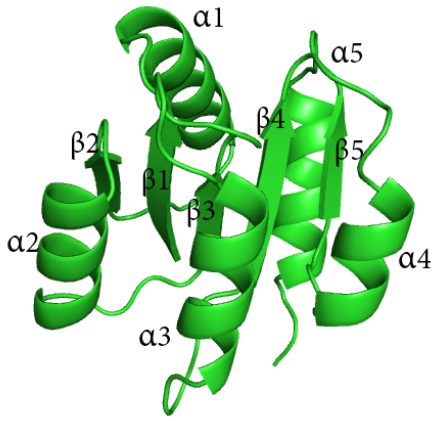


Figure 2.7. Ramachandran plot (Ramakrishnan & Ramachandran, 1965) of the BeF_3^- complexed $LytR^N$ structure. Four hundred and thirty nine (97.1%) of residues appear in favoured regions; thirteen (2.9%) in allowed regions; zero residues in outlier regions.

A



B

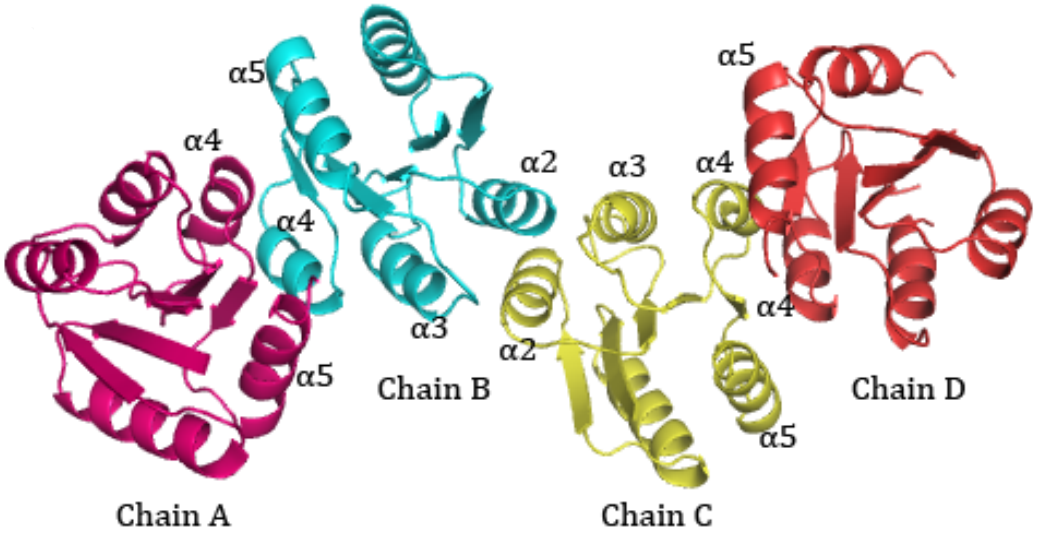


Figure 2.8. The two crystal structures of LytR^{N}. (**A**) Single molecule of apo-LytR^{N} in the asymmetric unit. (**B**) The four monomers observed in a single asymmetric unit of the BeF-LytR^{N} complex protein crystals. Chains A (pink), B (cyan), C (yellow), D (red), shown as cartoon; the helices involved in the interface interaction between chains A and B, C and D are α4 and α5; helices involved in the interface interaction between chains B and C are α2, and α3.

All $LytR^N$ structures are similar with Root-Mean-Square Deviation (RMSD) values between the monomer apo-$LytR^N$ and BeF-$LytR^N$ 0.407, 0.410. 0.410, and 0.388 Å, between chains A, B, C, and D, respectively. While chains A, B, C, and D from BeF-$LytR^N$ are almost all identical with each other, there is a difference between these structures as well as when compared to the apo-$LytR^N$ form.

2.3.2 Apo-$LytR^N$ Crystallizes in an "Active-like" Form in the Presence of Sulfate

The final native protein structure of $LytR^N$ contains residues 1-118, 13 water molecules, and one SO_4^{2-} (Table 2.1). Residues 119-134 were not modeled due to the absence of electron density suggesting general disorder at the C-terminal end of the protein. The solvent density site over the residue Asp8, was initially mapped as Mg^{2+} ion. However, examination of the B-values of active site side chains during initial stages of refinement, namely Nε2 from His83 (107.63 $Å^2$), and Nδ2 of Asn55 (123.00 $Å^2$) relative to the SO_4^{2-} oxygens O4 (126.67 $Å^2$) and O2 (133.94 $Å^2$) respectively (Table 2.3), as well as the large nonspherical electron density within the active site, and ionic environment provided by His83 and Asn55, that a sulfate ion occupies the space. Indeed, the B-value for the refined sulfate ion within the active site (312.7 $Å^2$) is similar to that of the above mentioned side chains, with appropriate interatomic distances to those side chains, supporting the placement of the sulfate ion within the active site.

The well-ordered sulfate ion is centrally located in the most accessible region of the active site. The sulfate ion is coordinated by the Nε amino nitrogen of His83, Nδ of the Asn55 amide side chain, and the carboxyl oxygen atoms from the Asp8, and Asp53 and at least 2 water molecules (Figure 2.8). The interaction of His83 (located in the β4-α4 loop) with the sulfate ion, leads to a conformational adjustment of the loop by positioning it closer to the active site similar to what is observed in other known phosphorylated proteins such as spr1814 and PhoB; this conformational change also induces structural alterations on α4 allowing for the stabilization of the α4-β5-α5 interface (Park *et al*., 2012; Bachhawat *et al*., 2005). In addition, a water molecule (W7) bridges between the hydroxyl of the Thr81 side chain, and Nζ of Lys101. The conserved residue Thr81 (Figure 2.9) adopts an active like conformation, moving closer to the phosphorylated Asp53 residue, similar to other RRs where it is found to be required for signal transduction upon phosphorylation (such as in FixJ; Brick *et al*., 1999). In the apo-form of $LytR^N$, both Thr81 and His83 have similar conformations as seen in other BeF-bound RRs (Brick *et al*., 1999). It appears that since the sulfate ion is chemically and physically similar to PO_4^{3-}, its presence at the active site core in combination with the residues surrounding the active site area, trigger a conformational change that is seen in other phosphorylated RRs. In addition, due to the close contacts between the symmetry related molecule, it was determined that the apo-$LytR^N$ forms a crystallographic dimer with another $LytR^N$ molecule using the same dimerization interface as the BeF-$LytR^N$ dimers.

Numerous structural findings of the receiver domains of RR suggest that RRs (full length or the N-terminal domain only) exist in equilibrium between active and inactive states, and phosphorylation allosterically shifts the equilibrium toward the active state (Stock and Da Re, 2000; Stock *et al.,* 2000; West and Stock, 2001; Volkman *et al*., 2001; Milani *et al*., 2005). In the apo-$LytR^N$ case the sulfate ion is able to have the same allosteric effect as phosphate, similar to what is seen by other ions such as Hg^{2+} in the case of unphosphorylated (or without BeF) StyR (PDB ID: 1YIO; Milani *et al*., 2005) and Ca^{2+} on $KdpE_N$ (PDB ID: 1ZH2; Toro-Roman *et al*., 2005) from FixJ and OmpR families, respectively. Similarly to the structures of PhoB (Bachhawat & Stock, 2007) and ComE (Boudes *et al*., 2014) crystallized in the absence of BeF_3^-, where the proteins were found to form crystallographic dimers, the apo-$LytR^N$ in the presence of the sulfate ion it dimerizes with a crystallographically related molecule.

2.3.3 Dimerization of $LytR^N$ when Crystallized in Complex with BeF_3^-

It has been shown the $LytR^N$ undergoes dimerization-induced activation in the presence of acetyl phosphate by using gel electrophoresis and gel filtration (Patel, 2014). In this case incubation of the protein in the presence of beryllium fluoride, which is a routinely used phosphoryl group analog (Chabre, 1990; Yan *et al*., 1999; Cho *et al*., 2001), resulted in a dimerization of the protein as well (Figure 2.3); which is most often associated with

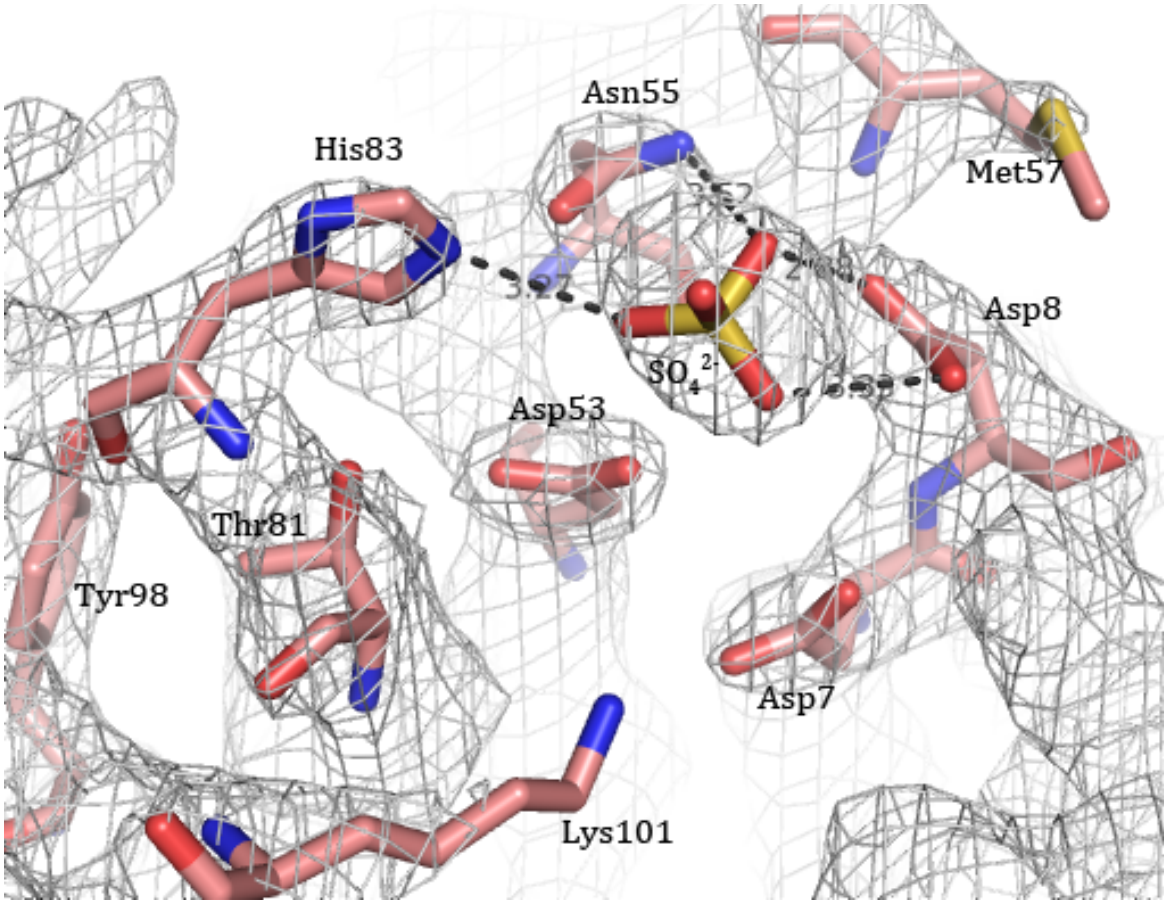


Figure 2.9. The active site of the apo-LytR^N structure. The conserved active site residues are shown as sticks, and the electron density map as a mesh is contoured at 1.5 σ. The sulfate ion coordination is shown by dotted lines, interacting with the active site residues, His83, Asn55 and Asp8, closely resembling active-like conformation of other RRs.

activation of RRs. In the presence of BeF_3^- and under very similar crystallization conditions, $LytR^N$ crystals formed with different morphology and space group than the apo-$LytR^N$ crystals. While the apo-$LytR^N$ crystals have one molecule, the BeF-$LytR^N$ complex crystals hold four molecules per asymmetric unit (Figure 2.10**)**.

Two sets of molecules (chains A and B, C and D) form identical two-fold symmetric dimers using the α4, loop-α4β5, and α5 interface (Figure 2.10a). In addition a two-fold non-crystallographic symmetry axis relates chains B and C through the α2α3 interface (Figure 2.10b); calculated bond lengths of the B-C interaction residues indicate that the interaction interface is too low for significance in functional dimerization.

The dimer interface as a result of BeF_3^- complex formation buries a 609 $Å^2$ of surface area from each monomer, and it involves the α-helix 4, the loop region between α4 and β5, as well as α5. The dimerization interface involves both hydrophobic and polar interactions stabilizing the dimerization of the protein. This interface, while resembling that observed in PhoB (PDB ID: 1ZES) and FixJ (PDB ID: 1D5W), is the closest to that observed in ComE belonging to the same subfamily of LytR proteins (PDB ID: 4MLD; Boudes, *et al*., 2014); that include the loop region between α4-β5 (Figure 2.11) unlike other family proteins where sheet β5 can be included as well (Park *et al*., 2012); or a completely different interface is observed involving only helices α1 and α5, in VraR and NarL (Leonard *et al*, 2003). Residues involved in hydrophobic interactions are Ala87, Val88, and Phe91, from both molecules A and B; all of them belonging to helix α4. In addition, Nδ2 of Asn94, which is part of the loop region between α4 and β5, forms a

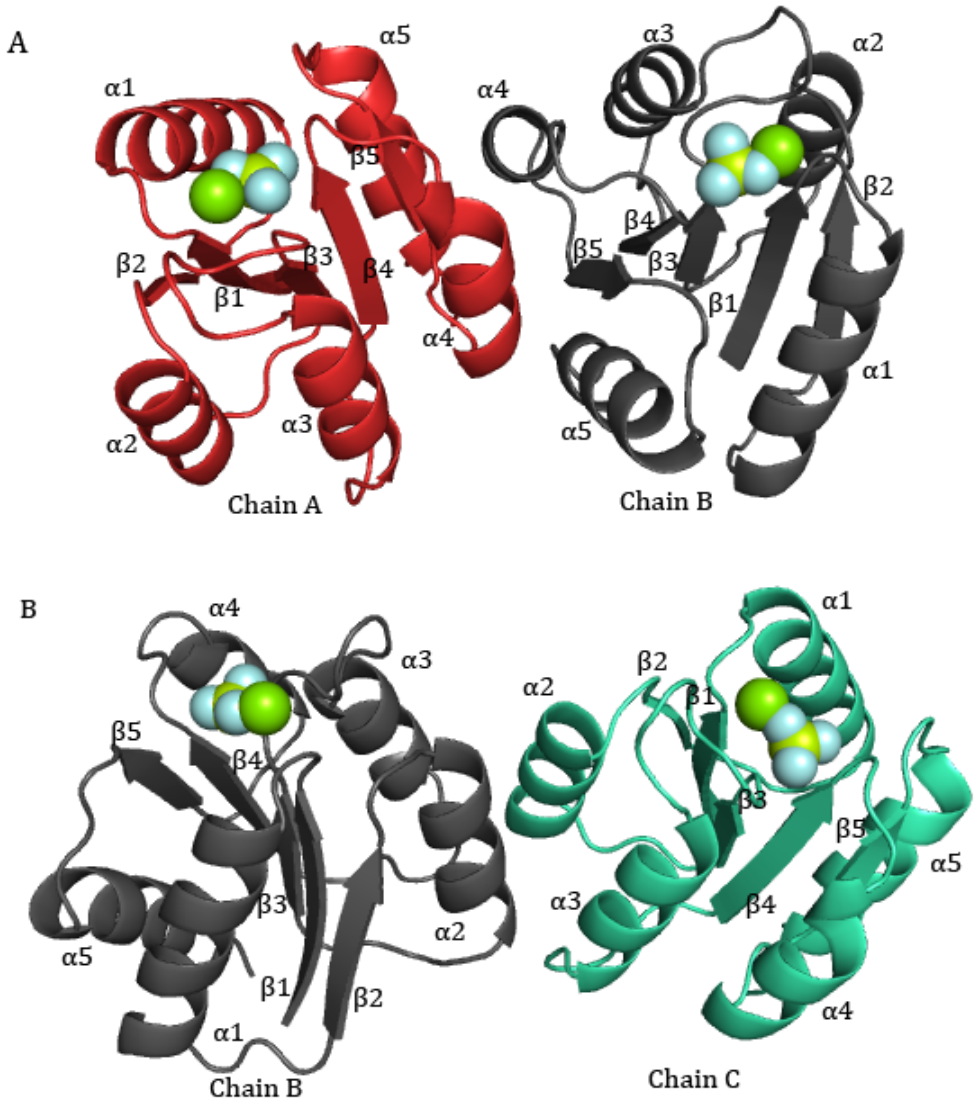


Figure 2.10. The dimerization interface of the BeF_3^- complexed $LytR^N$ structure shown in ribbon, relative to the ions at the active site as spheres (Mg, green; Be, yellow; F, cyan). (**A**) The dimerization interface between α4-loopα4β5-α5 as result of activation is observed between monomers A (red) and B (black), the same as observed between monomers C and D (not shown). (**B**) The non-crystallographic dimerization interface through α2-α3 between monomers B (black) and C (green) relates the two active $LytR^N$ dimers (AB and CD) in the asymmetric unit.

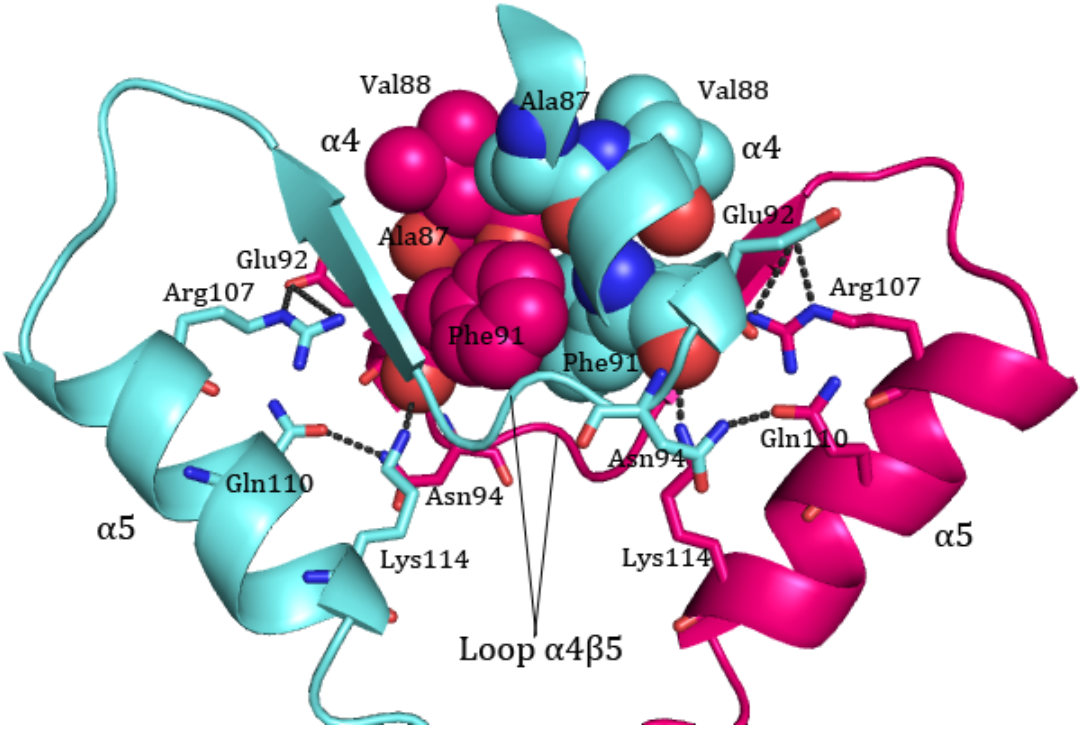


Figure 2.11. The $LytR^N$ dimer interface (AB dimer) highlighting the ionic and hydrophobic interactions in the BeF3- activated $LytR^N$. The α4-loopα4β5-α5 region of both monomers is shown as ribbons, with conserved residues that are involved in intermolecular salt bridges (dashed lines) shown in stick model. Residues involved in hydrophobic contacts are shown as spheres (carbons, cyan and magenta; nitrogen, blue; oxygen, red).

hydrogen bond (2.98 Å) with Oε1 from Gln110 in helix α5. Nζ of Lys114 (in helix α5) forms a hydrogen bond (3.40 Å) with the backbone Oxygen of Phe91. Oε1 of Glu92 forms salt bridge interactions with Arg107 Nε, and NH2 (3.77 and 3.11Å, respectively) connecting α4 and α5 (Figure 2.11). Similar interactions are observed between molecules C and D, which were also found in the apo-$LytR^N$ structure with a crystallographicaly-related molecule.

Dimers AB and CD form a tetramer generated by the BC interface (covering an area of 584 $Å^2$) related by non-crystallographic two-fold symmetry via the α2, and α3 face of monomers B and C. Chains A, B, and C are well defined with continuous density, compared to chain D which has an overall lower density, suggesting instability of the chain. Residues 119-134 were not mapped on any chain due to the absent density. Four BeF_3^- ions, and 3 Mg^{2+} ions were modeled, however one magnesium ion was not modelled due to the low density in the active site of the D monomer. It is worth noting that signs of conformational disorder by the segments of broken density in the loop regions β1-α1 and β3-α3 of the D monomer, in addition to the absent magnesium ion affect the packing of the molecules in the crystal lattice, contributing to the overall low resolution structure of the protein in the active form. In the active site of $LytR^N$ the presence of BeF_3^- and Mg^{2+} complex noncovalently bound to Asp53 was clearly visible in the initial electron density map. The Mg^{2+} ion is coordinated with the fluoride atom F2 from the BeF_3^-, the side chain carbonyl oxygens (Oδ) of Asp8 and Asp53, as well as the main chain carbonyl oxygen of Asn55, which are also found conserved in other RRs. Water molecules normally satisfy the octahedral coordination of the ion in other higher

resolution structures of RR, however due to the low resolution of BeF- $LytR^N$ waters were not modeled into the structure. Lys101 forms salt bridges (involving both polar and hydrogen interactions) with Asp7 and F1 of BeF_3^-, in addition F1 of BeF_3^- forms a hydrogen bond with the main chain carbonyl oxygen of the conserved Thr81. The main chain amide of Asn55 forms hydrogen bonds with F3 of BeF_3^- and the carbonyl oxygen of Asn55, which then interacts with the Nε, amino nitrogen of His83 (Figure 2.12). As a result, the orientation of His83 allows for the main chain carbonyl O of His83 residue to hydrogen bond with the –OH of Tyr98 (also a conserved residue, see Figure 2.1), stabilizing the extended conformation of the β4-α4 loop. The –OH group of Tyr98 points toward the N-terminus of helix α4, stabilizing helix α4, thus resulting in the stabilization of the dimerization interface in a similar manner to that observed in other NarL/FixJ or OmpR/PhoB family members (Park *et al*., 2012; Toro-Roman *et al*., 2005). Contrarily, Phe107 from ComE (PDB ID 4MLD; Boudes *et al*., 2014) is located in the same area as Tyr98 from LytR, interacting with helix α4 through hydrophobic contacts with Ala94, located on helix α4, instead of ionic interactions of Tyr98 as seen in $LytR^N$-BEF structure.

2.3.4 Insights into Phosphorylation-Induced Activation Mechanism of $LytR^N$

Two residues, Thr81 and Tyr98 ($LytR^N$ numbering), have been described as acting like switch residues during activation of various RRs through alteration in side-chain orientation upon phosphate binding (Bachhawat & Stock, 2007). In the apo-$LytR^N$

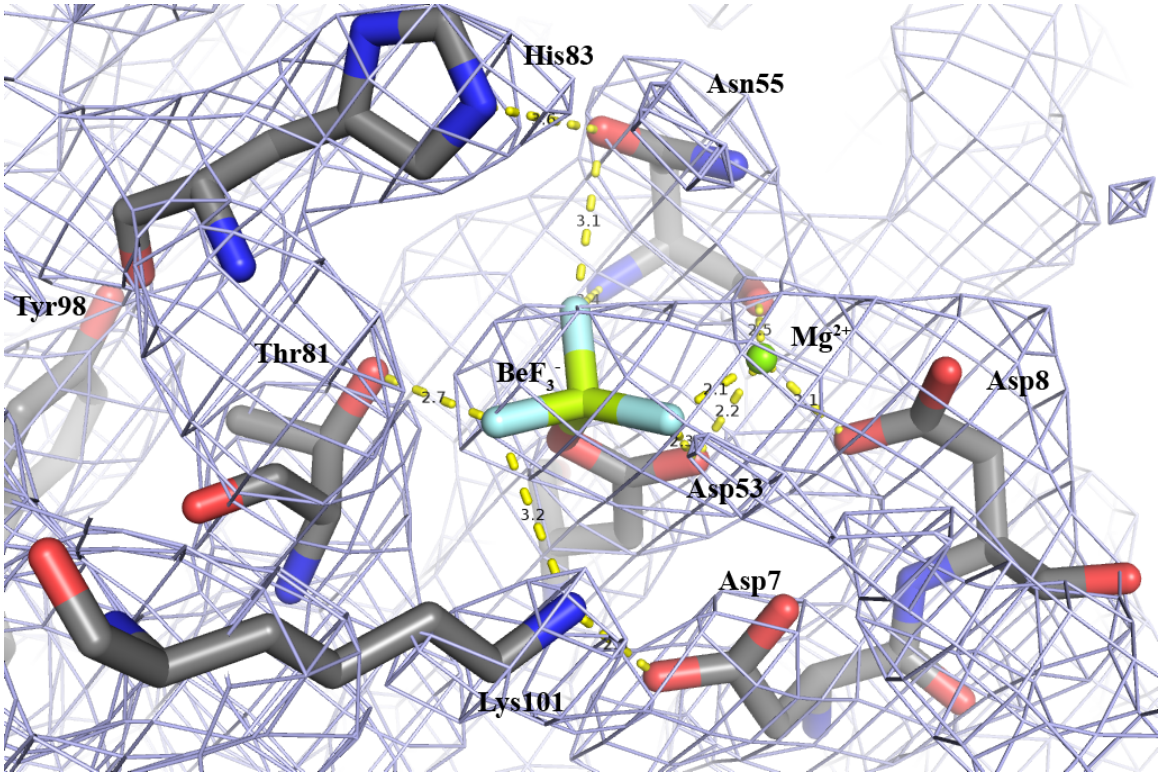
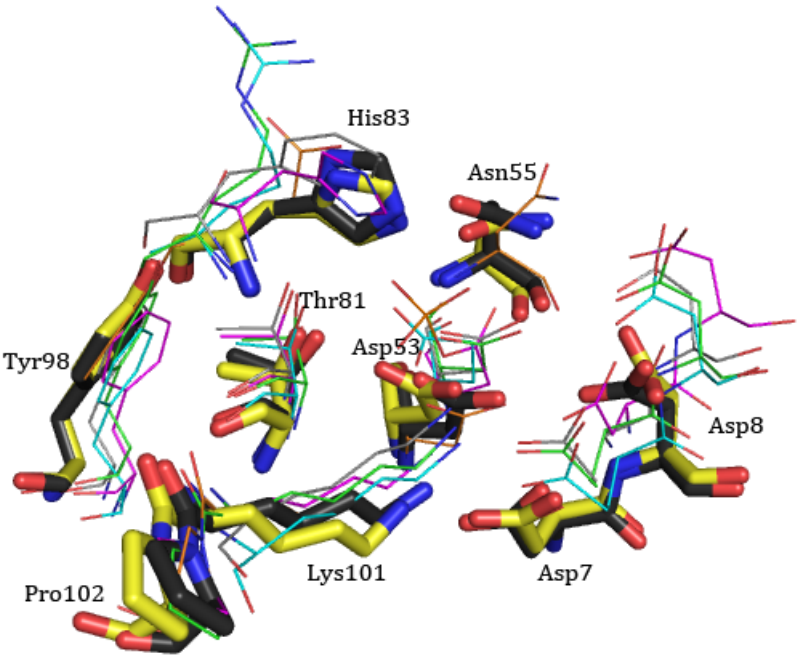


Figure 2.12. The active site of the BeF_3^- complexed $LytR^N$. The conserved monomer A active site residues are shown as sticks with oxygen atoms in red, and nitrogen atoms in blue. The electron density map is contoured at 1.0 σ. Coordination of the BeF_3^- and Mg^{2+} ions is shown by dotted lines.

and BeF-LytR^N structures both Thr81 and Tyr98 exhibit an active-like conformation, due to their direct or indirect interaction with the sulfate or beryllium fluoride at the active site, respectively (Figures 2.9 & 2.12). The key residue affecting the position of the two residues appears to be His83, which hydrogen bonds (2.62 Å) with the hydroxyl group of Tyr98 via the main chain carbonyl oxygen of His83. While this residue is not strictly conserved among RR receiver domains, for instance this residue is Arg85 in PhoB and Arg90 in ComE, Glu89 in CheY or Phe83 in spr1814, and His84 in FixJ (Figure 2.13), the respective side chains adopt similar conformations as His83 of LytR^N (Figure 2.13). These residues appear to have an important role in ion coordination at the active site upon phosphorylation (Bachhawat *et al*., 2005; Solà *et al*., 1999; Park *et al.,* 2012; Birck *et al*., 1999). Figure 2.13a shows an overlay of the side chains of the several receiver domains of RRs from different families, such as FixJ, PhoB, CheY, spr1814, and ComE in their active forms obtained in complex with BeF_3^-, or by D58E substitution in ComE, compared with the BeF-LytR^N (black) and apo-LytR^N (yellow), highlighting the differing positions that His83 (or equivalent) side chain take upon phosphorylation and their respective root-mean-square deviations (RMSD) (Maiorov & Crippen, 1994). There are very few differences between the active site residues of apo-LytR^N and BeF-LytR^N, which show an RMSD of 0.611 Å. Superposition of other BeF-activated RR receiver domains onto LytR^N show slightly greater differences with RMSD values of 2.283 Å, 1.954 Å, 0.790 Å, 2.266 Å, and 3.352 Å for BeF-FixJ, BeF-PhoB, BeF-CheY, BeF-spr1814 and ComE receiver domain with D58E, respectively (Table 2.3). The differences in side chain conformations between other RRs and BeF-LytR^N are similar to

A.


His83
Asn55
Thr81
Tyr98
Asp53
Asp8
Lys101
Asp7
Pro102

B.

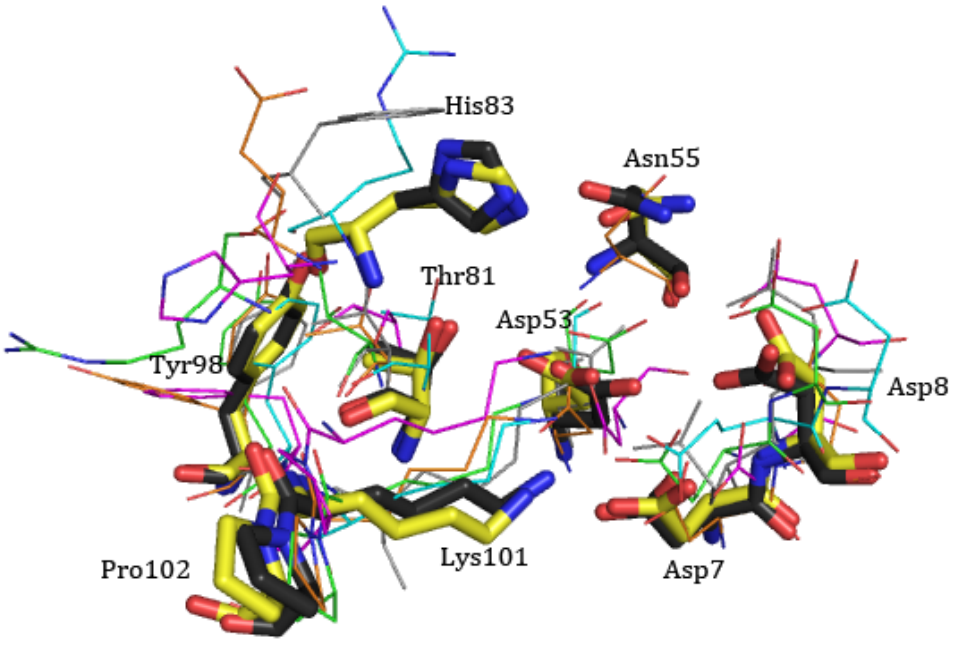

His83
Asn55
Thr81
Asp53
Tyr98
Asp8
Pro102
Lys101
Asp7

Figure 2.13. Comparison of the active site for both BeF-LytR^N (black) and apo-LytR^N (yellow) LytR^N labeled residues shown in sticks with other conserved RR residues shown as lines. (**A**) Comparing LytR^N (Asp7, Asp8, Asp53, Thr81, His83, Tyr98, Lys101, Pro102) with other active protein structures: FixJ (Asp10, Asp11, Asp54, Thr82, His84, Phe101, Lys104, Pro105; PDB ID: 1D5W, magenta), PhoB (Glu9, Asp10, Asp53, Thr83, Arg85, Tyr102, Lys105, Pro106; PDB ID: 1ZES, green), CheY (Asp12, Asp13, Asp57, Thr87, Glu89, Tyr106, Lys109, Pro110; PDB ID: 1FQW, orange), spr1814 (Glu7, Asp8, Asp53, Thr81, Phe83, Tyr100, Lys103; PDB ID: 4E7P, grey) and ComE (Glu7, Asp8, Glu58, Thr88, Arg90, Phe107, Lys110; PDB ID: 4MLD, cyan). Both apo-LytR^N and BeF-LytR^N display similar conformations of the key residues in other active RRs (RMSD values for the overlay of these residues are in Table 2.3), stressing that the apo-LytR^N displays active-like conformation. (**B**) Comparing the same LytR^N residues with the key residues of other apo-inactive RRs receiver domain structures: FixJ (PDB entry 1DBW, magenta), PhoB (2IYN, green), CheY (3CHY, orange), spr1814 (4E7O, grey), and ComE (4ML3, cyan). Again, both apo-LytR^N and BeF-LytR^N show similar conformations, larger differences are observed with other RRs in residues equivalent to Thr81, His83 and Tyr98 (RMSD values in Table 2.3).

Table 2.3. Comparison of Receiver Domain Active Site Side Chain Positions

	RMSD (Å)*	
LytR^N vs. Inactive RRs	**Apo-LytR^N**	**BeF-LytR^N**
Apo-LytR^N	0	0.611
FixJ	4.330	4.177
PhoB	3.672	3.551
CheY	1.319	1.060
spr1814	3.539	3.261
ComE	2.099	2.019
LytR^N vs. Active RRs		
BeF-LytR^N	0.611	0
FixJ	2.283	2.158
PhoB	1.954	2.027
CheY	0.790	0.467
spr1814	2.266	2.349
ComE	3.352	3.315

***RMSD** = $[\{SUM(d_{ii})^2\}/N]^{1/2}$, where d_{ii} is the distance between the i^{th} atom of structure 1 and the i^{th} atom of structure 2, N is the number of atoms matched from each structure (Maiorov & Crippen, 1994).

those observed for apo-LytR^N, mainly due to the similarity of the active sites between the two forms of LytR^N. On the other hand, the RMSD values for the superposition of LytR^N with the native forms of other RRs (Figure 2.13b) shows greater differences, with RMSD values of 4.330 Å, 3.672 Å, 1.319 Å, 3.539 Å, and 2.099 for FixJ, PhoB, CheY, spr1814, and ComE, respectively (Table 2.3). The difference in RMSD values for BeF-LytR^N and the same RRs is similar to those values with apo-LytR$^{N.}$ These observations indicate that both apo- and BeF-bound LytR^N active sites more closely resemble other active-like (phosphorylated) RRs than apo/native (unphosphorylated) forms of those receiver domains. With the exception of ComE active and inactive structures, which show lower RMSD values of 2.099 Å for inactive, and higher value of 3.352 Å for the active form of the protein when compared to LytR^N (Table 2.3). These values result mainly due to the H-bond interaction of Arg90 (from the loop region β4-α4) with Glu65 (loop region β3-α3), in the absence of the conserved Aspartate at the active site, and Histidine from the loop region β4-α4 interacting with BeF_3^- at the active site.

Figure 2.13 highlights that in both apo-LytR^N and BeF-LytR^N residues His83, Thr81, and Tyr98 adapt conformations similar to that of structures from other phosphorylated (active) RR receiver domains. Conversely, comparison of the LytR^N structures with other apo/un-phosphorylated RR receiver domains (Figure 2.13b) indicates that while most active site residues align well, His83, Thr85, and Tyr98 of LytR^N, differ significantly. In addition, the full structure alignment of LytR^N-BEF with ComE D58E (active form) from the same LytTR family results in an RMSD value of 3.416 Å. Crystallization of the native and active forms of the protein was done by

phospho mimetic mutation of D58E for active and D58A for inactive (Boudes *et al*., 2014), where both structures, similarily to apo-LytR^N and BeF-LytR^N, resemble each other with a low RMSD value of 0.610 Å.

In apo-LytR^N His83 is part of the β4-α4 loop region. Although this loop region is not involved in the dimerization interface in LytR^N, the main difference between the BeF-LytR^N and apo-LytR^N structures appears to be in the length of this loop region affecting the length of helix-α4 (when the loop is shorter, the helix is longer). In the native LytR^N residues Gln85 and Tyr86 are part of the β4-α4 loop, whereas in the active form the two residues belong to helix-α4, making the helix longer, while shortening the loop in the BeF-LytR^N structure (Figure 2.14). The extension of helix α4 is also observed in the difference of other active and inactive forms of RRs such as StyR from the FixJ family, where the inactive form has a shorter helix α4 (Milani *et al*., 2005; Birck *et al*., 1999). Similarly the active CheY structure was found to have an extended helix α4 from 6 to 9 residues (Lee *et al*., 2001). This extension of the α4 may allow for the stabilization of the dimerization interface through this helix, and it also found to be long in both active and inactive forms of ComE structures (Boudes *et al*., 2014).

In BeF_3^- activated structures, the Thr81 (LytR^N numbering) fluoride-hydrogen bond (Figure 2.12) operates as a switch (Birck *et al*., 1999), stabilizing the β4-α4 loop, and orienting helix α4 to stabilize the homodimer (Robinson *et al*., 2000). While the so-called switch interaction is not observed in apo-LytR^N in the absence of BeF_3^- ion, this

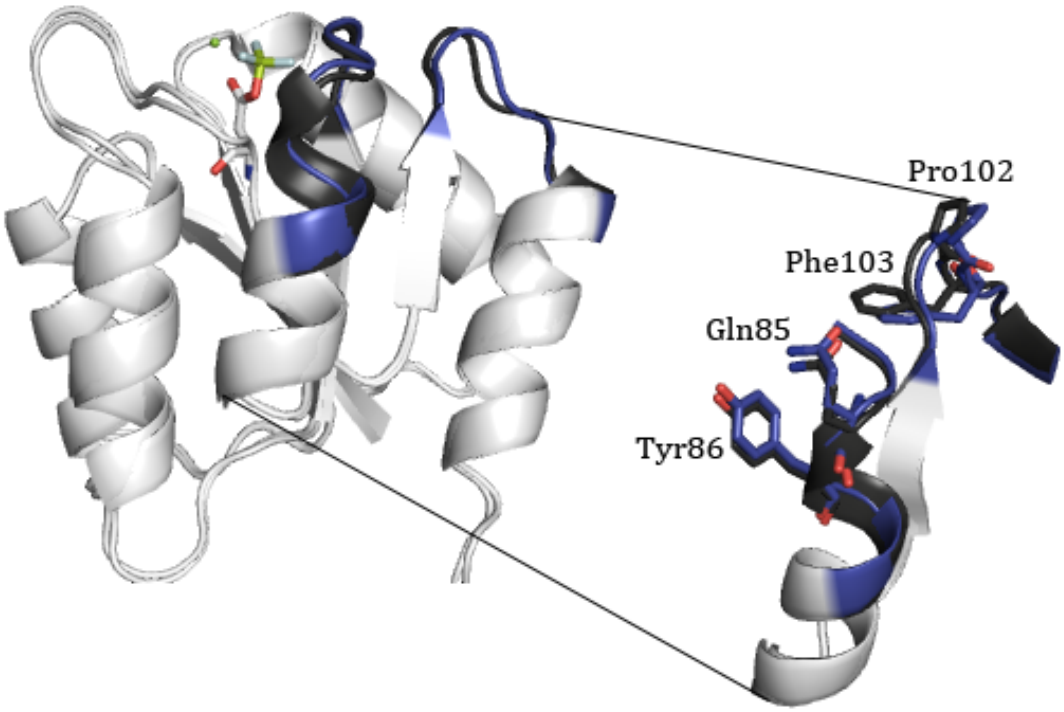


Figure 2.14. Comparing the conformation changes between the apo-LytR^N (blue) and BeF-LytR^N (black). The main differences occur in helix-α4, where residues Gln85, and Tyr86 become part of a helix upon activation. In addition loop region connecting β5 and α5 is affected by residues Pro102 and Phe103, where the Phe103 fills in the cavity between the loop and helix-α1, stabilizing the structure.

suggests different activation modes and dimerization between the two forms of $LytR^N$ proteins. In many RRs, such as FixJ and PhoB, complex formation with $BeF_3^=$ mediates dimerization of the receiver domains to promote DNA binding by the effector domains. Considering that these receiver domains of RRs exist in equilibrium between active and inactive forms stabilization of the protein with BeF_3^- and Mg^{2+} (Solà *et al.,* 1999, Makino *et al.,* 1989), which form shorter bonds than SO_4^{2-}, is more effective than with SO_4^{2-}. The sulfate, in the case of apo-$LytR^N$, has a similar effect on the active site of the protein as BeF_3^- and Mg^{2+} ions on the BeF-$LytR^N$; inducing crystallographic dimerization of the protein, which is associated with phosphorylation-induced activation, or with high protein concentration. Although many inactive structures of RRs have been determined in dimeric form, they are more likely to occur as a result of high protein concentrations (i.e. NtrC, StyR, PhoB) and usually have different dimerization interface than the active structures (Park *et al*., 2012,). The largest difference between apo-$LytR^N$ and BeF-$LytR^N$ is observed in the length of helix α4, and the loop region connecting β5 and α5. Although the loop between β5 and α5 is not directly part of the interface, it affects helix α5, which is involved in interface interactions and dimer stabilization. In BeF-$LytR^N$ the conformation of Phe103, which is located between α1 and loop β5-α5, brings the two regions of the protein closer, closing the cavity between α1 and loop β5-α5 (Figure 2.14). These data suggest that there can be more than one activation mode of RRs depending on the residues at the active site. In addition, although the receiver domains from different RR subfamilies are similar structurally, they employ different dimerization interfaces; the $LytR^N$ dimer interface more closely resembles ComE from the same family of RRs

members (Boudes *et al.,* 2014). The phosphorylated form of ComE, as described by Boudes *et al*. (2014), from the same family as LytR, was determined to have little effect on DNA-bnding affinity therefore activation of the protein, instead it induces oligomerization of the protein. On the other hand, they propose that the dimer interaction between two receiver domains of ComE may trigger activation of the LytTR RR, and the oligomerization of the inactive form occurs only due to the elevated protein concentrations. While the crystallization results of LytR are similar to ComE, where both apo- and BeF- complex structures of LytR resulted in active-like dimers using the similar dimerization interface as ComE, it is suggested that both LytTR family proteins ComE and LytR may share functional homologies. Identification of the dimerization interface residues that were determined in this work may lead towards new drug developments, for blocking LytR protein-protein interactions, in the fight against *S. aureus* infections.

Chapter 3

Probing the interaction of TraW and TrbC from the *Escherichia coli* F plasmid Type IV Secretion System

3.1 Introduction

The rapid establishment of antibiotic resistance within a bacterial population is largely due to the ability of prokaryotes to share their DNA, not only within the same species, but also between unrelated species. Conjugation, as the most important process for horizontal gene transfer, requires independently replicating conjugative plasmids that encode and facilitate their own transfer as well as the transfer of other cellular DNA between two cells (Lanka & Wilkins, 1995; Frost *et al*., 2005). Genes essential for conjugative single-stranded DNA transfer of the *Escherichia coli* K-12 fertility factor, F, are encoded within the 33.3kb transfer (tra) region of this 99.159 kb plasmid (Genbank accession number: NC002483) (Ippen-Ihler & Minkley, 1986; Moore *et al*., 1990; Frost *et al*., 1994) (Figure 3.1**)**. While many of the proteins encoded by this plasmid are homologous to other systems such as the tumor inducing (Ti) plasmid of *Agrobacterium tumefaciens* (Christie & Cascales, 2005), the pertussis toxin *Bordetella pertussis* (Shrivastava & Miller, 2009) or the CagA system of *Helicobacter pylori* (Cascales & Christie, 2003), there are twelve proteins that are unique to the F plasmid (Table 1.2).

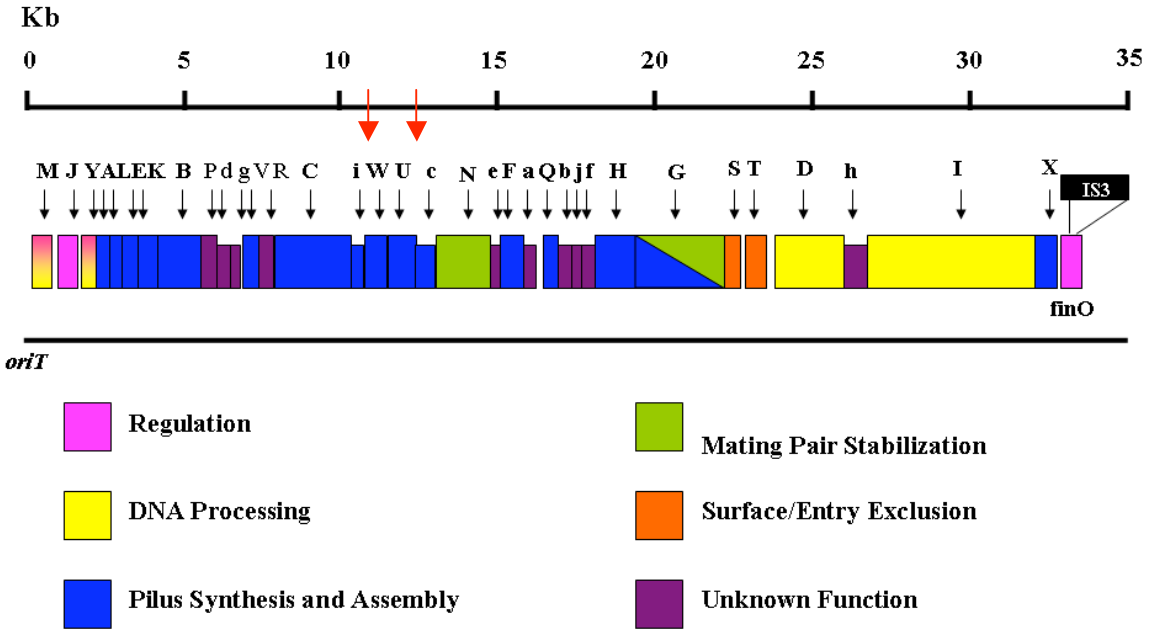


Figure 3.1. The 33.3 Kb transfer region of the F plasmid. The genes of the transfer region are noted, with capital letters denoting "tra" (i.e. *traW*) and lower case denoting "trb" (i.e. *trbc*), where the order of the genes is mapped. Protein function is denoted by colour, shown below the map. The position of TraW (W), and TrbC (c) are indicated by the red arrows; both proteins involved in the pilus synthesis assembly. Copied from Frost *et al.,* 1994.

A key component initiating the F-mediated conjugation is the F-pilus (made of multiple TraA subunits), which is assembled through a T4SS, that mediates contact between a F+ donor cell and a F-recipient. Subsequent to this contact, a mating pair bridge is formed that mediates DNA transfer through the pilus that also serves as a channel for ssDNA transfer during conjugation (Lawley *et al*., 2003; Schröder & Lanka, 2005; Babić *et al*., 2008). Although energetically expensive, conjugative DNA transfer is highly regulated responsive to physiological and environmental stimuli, controlled by the interaction and communication of numerous F plasmid proteins (Frost & Manchak, 1998). While DNA processing and plasmid transfer from a donor to recipient cell have been well studied, the details regarding the interactions occurring within the proteins of the T4SS leading to Mpf and plasmid transfer remain relatively unclear (Arutyunov & Frost, 2013). For many of the transfer proteins of the F T4SS, their overall role in F-mediated conjugation is known and can be related to other homologues of similar systems, however detailed information into their functions and particular interactions in the complex formation and orchestration of the pilus assembly, pore formation, and Mpf are less well characterized. The general proposed organization of the proteins forming the T4SS machinery in shown in Figure 3.2.

The periplasmic proteins TraW and TrbC (Figure 3.2, shown in cyan, W and c, respectively) are unique to F-like systems with no homologues in other similar systems and are required for F-pilus assembly and DNA transfer proficiency. They play an important role in Mpf during conjugation (Maneewannakul *et al*., 1987; Maneewannakul *et al*., 1991; Lawley *et al*., 2003). In several systems, including the R27 (*Salmonella*

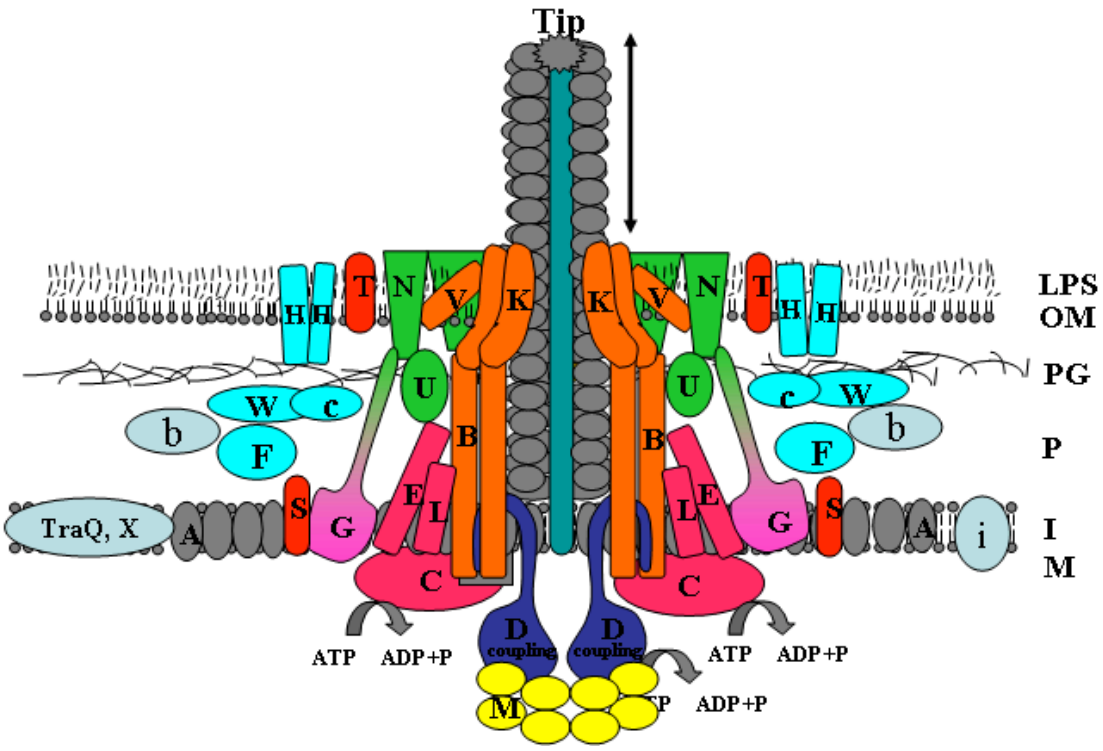


Figure 3.2. Model of T4SS system assembly as encoded by the F plasmid of *E. coli*. These proteins are involved in the retraction, of the F pilus (grey), stabilization of the mating pair, as well as conjugative DNA transfer. Copied from Lawley *et al*, 2003.

typhimurium (Sherburne *et al*., 2000), Rts1 (*Proteus vulgaris* (Murata *et al*., 2002)), R391 (*Providencia rettgeri* (Böltner *et al*., 2002)) and SXT (*Vibrio cholerae* (Beaber *et al.,* 2002)) plasmids, TrbC is fused into the N-terminal end of TraW. Conversely, in F (Frost *et al*., 1994), pED208 (*Salmonella typhimurium;* Lu *et al*., 2002), UWE25 (*Protochlamydia amoebophila*; Greub *et al*., 2004) and pNL1 (*Novosphingobium aromaticivorans*; Romine *et al*., 1999) plasmids, TraW and TrbC are expressed as separate proteins, where TraW is encoded upstream to TrbC (Figure 3.3). This suggests that these proteins may have a common function and potentially interact even when expressed separately (Anthony *et al*., 1999; Lawley *et al*., 2003; Greub *et al*., 2004; Arutyunov *et al*., 2010; Ramsey *et al*., 2012). Sequence alignment of TraW/TrbC proteins, when expressed as a single or two separate polypeptides, (Figure 3.4) suggests that (a) the proteins interact for proper T4SS function when expressed separately, and (b) that this interaction is likely through the C-terminal region of TrbC to TraW. In R27 and SXT systems, the gene sequence is continuous (Figure 3.4, the two top sequences), whereas TrbC and TraW, highlighted in grey and black, respectively, in pED208 and F are expressed separately as encoded by two different genes.

Yeast two-hybrid analysis has confirmed the interaction between some of the proteins in the periplasm, including TraW-TrbB and TraW-TraU interactions (Figure 3.5), but no association between TrbC and other proteins could be observed (Harris & Silverman, 2004). To date, there are no published results to indicate interaction between the two proteins, but in the current study we confirm the interaction between TrbC and TraW, and demonstrate the preferred interaction of TraW with the N-terminus of TrbC

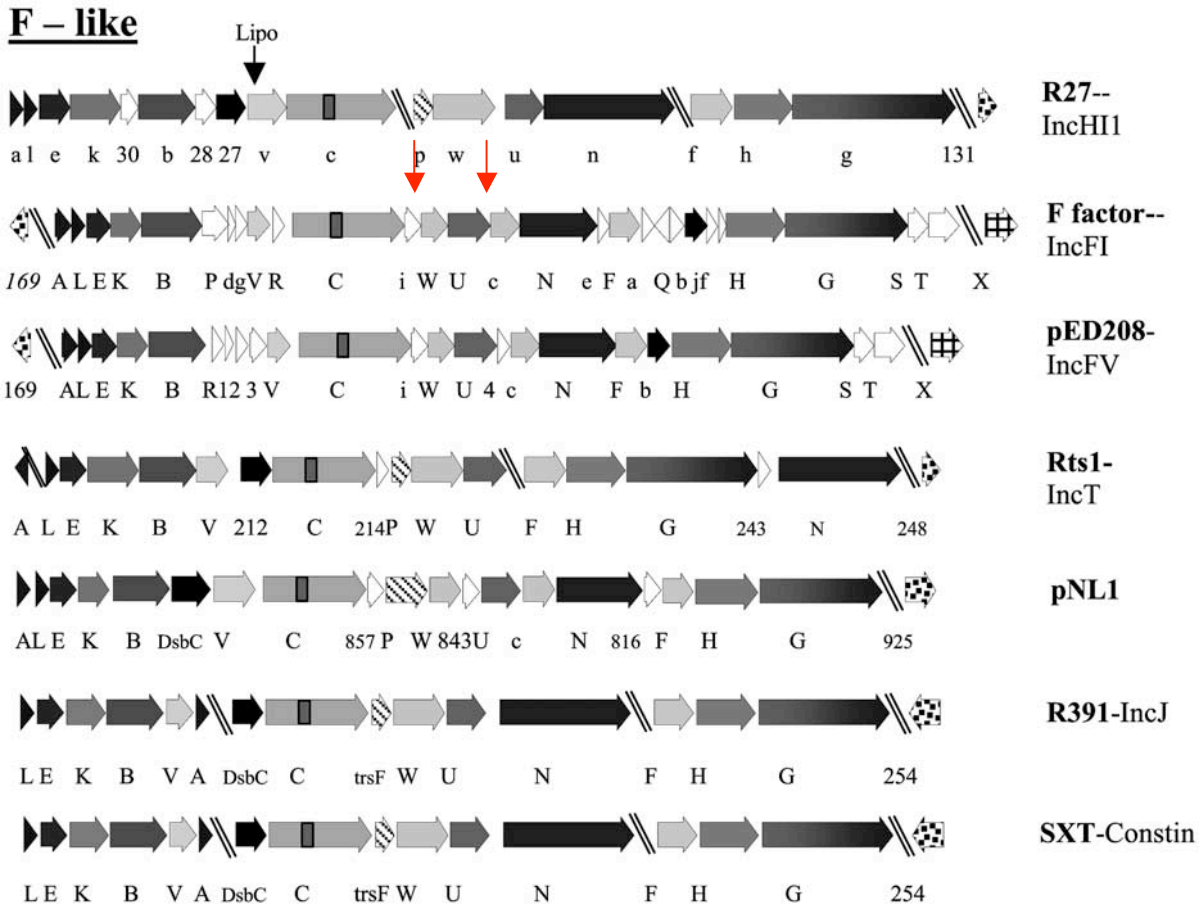


Figure 3.3. Comparing the genes of the F-like systems. TraW (W) and TrbC (c) (red arrows) are fused in R27, Rts1, R391, and SXT, whereas in F factor, pED208, and pNL1 they are separate polypeptides. Copied from Lawley *et al.,* 2003.

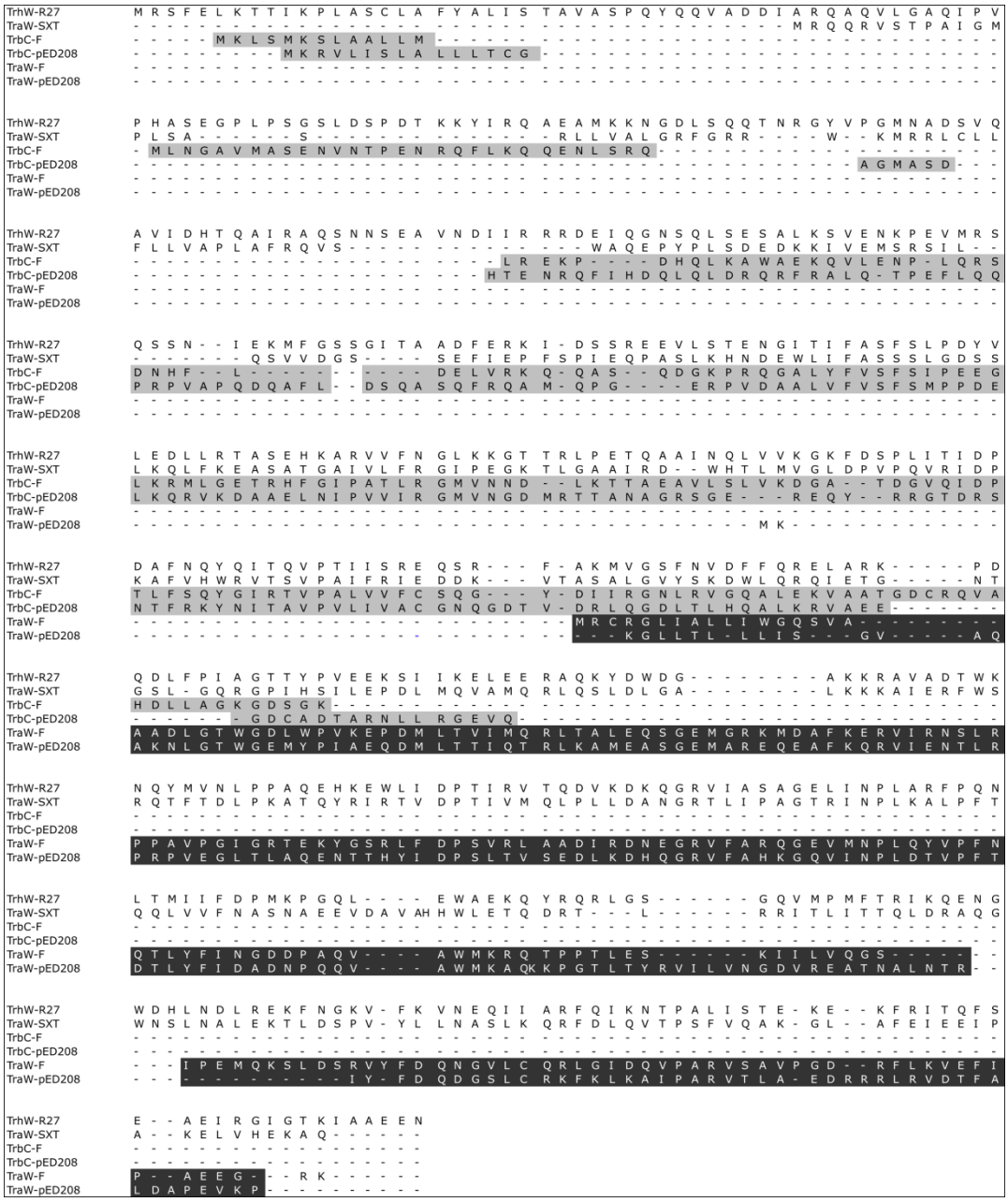


Figure 3.4. Sequence alignment of TraW and TrbC from several F-like T4SSs. TraW/TrbC is expressed as a single polypeptide in the R27 and SXT systems, however in F and pED208, TraW (highlighted in black) and TrbC (highlighted in grey) are expressed separately. The alignments was produced using Clustal Omega (Sievers *et al*., 2011) using the following Uniprot accession numbers: TrhW-R27 (NP_058221.1); TraW-SXT (AAL59671.1); TrbC-F (NP_061467.1); TrbC-pED208 (AAM90718.1); TraW-F (NP_061465.1); TraW-pED208 (AAM90715.1).

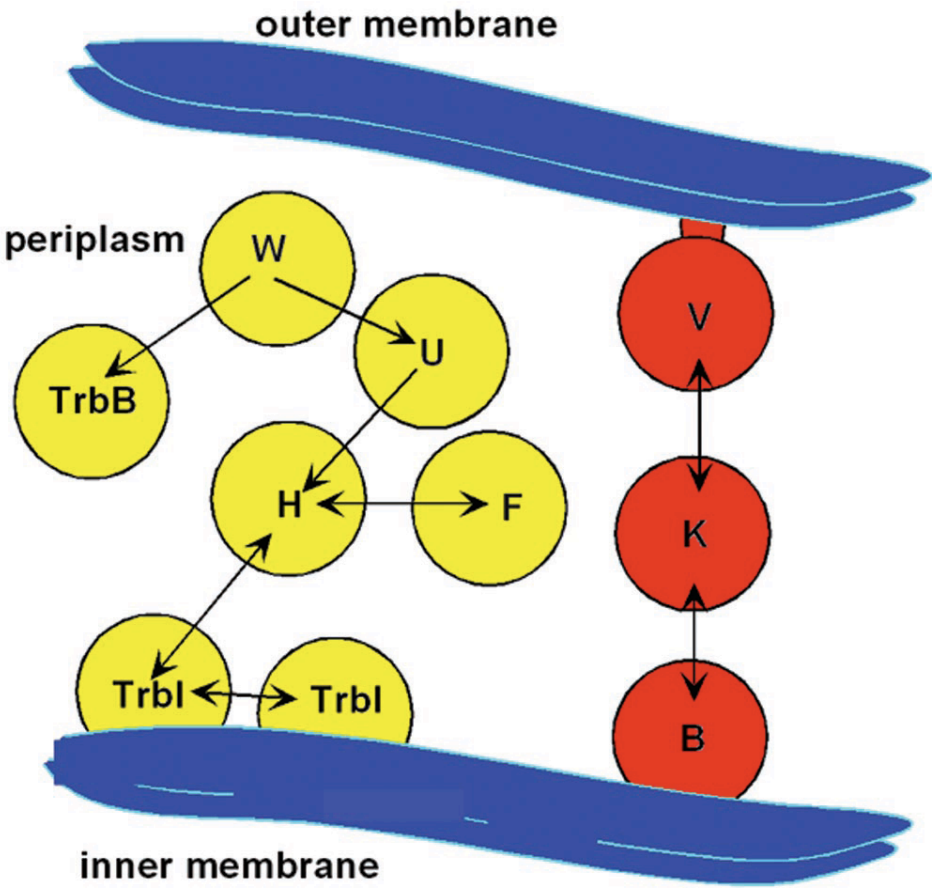


Figure 3.5. Periplasmic protein interactions as determined by yeast two-hybrid analysis. *tra* proteins are denoted in capital letters, such as W for TraW, which is shown to interact with TrbB and TraU; as indicated by arrows. Copied from Harris & Silverman, 2004.

versus the C-terminus. It was also determined that this interaction is independent of the presence or absence of N-terminal purification tags.

3.2 Materials and Methods

3.2.1 Construction of N-terminally 6-His tagged TraW and TrbC

Several initial constructs, including the pMALp2 vector harbouring N-terminally MBP-tagged TrbC (MBP-TrbC) and pT7-7 vectors harbouring C-terminally 7-His tagged TraW and TrbC (TraW-His_7 and TrbC-His_7), were kindly provided by Dr. Laura Frost (University of Alberta). The *traW* gene encoding the mature protein, from protein ID: NP_061465.1 (residues 19-210) was PCR amplified by Vent polymerase (New England Biolabs; NEB) using forward primer 5'-GCC*CATATG*GCCGATCTTGGTACCTGG-3' containing an NdeI site (italics) and reverse primer 5'-GGCC*AAGCTT*GTTATTTTCTGCCCTCC-3' containing an HindIII site (italics). Similarly, the *trbC* gene encoding the mature protein, protein ID: NP_061467.1 (residues 22-212) was PCR amplified using forward primer 5'-GCC*CATATG*TCAGAAAACGTGAACAC-3' containing an NdeI site and reverse primer 5'-GC*AAGCTT*TTATTTCCCGGAATCTCCT-3' containing an HindIII site. Following PCR amplification, the PCR products were isolated from an agarose gel using the QIAquick Gel Extraction Kit (Qiagen), digested with NdeI and HindIII and cloned into pET28a expression vectors (Novagen). The resulting expression plasmids

(pET28a::TraW and pET28a::TrbC) encode for the N-terminally six-histidine tagged His_6-TrbC and His_6-TraW.

3.2.2 TrbC In-Gel Trypsin Digestion for Mass Spectrometry Analysis

Identification of TrbC dimerization was performed using an in-gel tryptic digestion kit (Thermo Scientific-Pierce) as per the manufacturers protocol and mass spectrometric analysis. A 12% SDS gel was prepared and analyzed for the TrbC fractions. The protein band of interest with a size of ~45 kDa was then excized from the gel. The gel slices were destained to remove the Coomassie dye using a methanol/acetic acid solution, then incubated in 30 µL reduction buffer to reduce any bonds, incubated in 30 µL of alkyl buffer and washed twice with 200 µL of the destain solution at 37 °C. Incubation in 50 µL acetonitrile shrunk the gel slices, which were then dried in a speed vac prior to the addition of 35 µL of digestion buffer, containing trypsin for over 16 hrs at 30 °C. This extract solution containing the peptide fractions was then sent for analysis using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) analysis at the Center for Research in Mass-Spectrometry at York University, and peptide fragments were analyzed using ProteinProspector (Chalkley *et al*., 2005).

3.2.3 Expression of TraW and TrbC

Protein expression of pET vector-based constructs (pET28a::TraW (His_6-TraW), pET28a::TrbC (His_6-TrbC), pT7-7::TraW (TraW-His_7), pT7-7::TrbC (TrbC-His_7)), was

achieved using *Escherichia coli* BL21 (DE3) cells (Stratagene). Cells were grown with shaking in 1 L Luria-Bertani (LB) broth supplemented with 50 μg mL^{-1} Kanamycin (Kan) for pET28a expression vectors) or Ampicillin (Amp) for pT7-7 expression vectors at 37°C / 200 rpm until mid-log phase (OD_{600} = 0.7-0.9), at which point protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1.0 mM. Protein expression was allowed to continue post-induction for additional 4 hours at 37°C, at which time cells were harvested by centrifugation at 6,000 x g for 20 min at 4°C. The cells were kept frozen at -20°C until further use.

3.2.4 Expression of MBP-TrbC

The pMALp2::TrbC plasmid was originally provided by Dr. Laura Frost (University of Alberta). This plasmid, harboring MBP-TrbC was transformed into DH5α cells using the heat shock transformation protocol (Invitrogen). The DNA was then isolated and purified from the cells using QIAprep Spin Miniprep Kit (Qiagen). This DNA was then transformed into competent ER2508 (New England Biolabs) for the periplasmic expression of MBP-TrbC. Cells were grown in 1 L LB broth supplemented with 50 μg mL^{-1} Amp at 37°C / 200 rpm to mid-log phase (OD_{600} = 0.7-0.9). Protein expression was induced with the addition of 1 mM IPTG and was allowed to proceed for 3 hr at 37°C, and cells were harvested by centrifugation at 6,000 x g for 20 min / 4°C.

3.2.5 Purification of MBP-TrbC

The MBP-TrbC protein was isolated from *E. coli cells* using an osmotic shock protocol. The cells are washed with a hypotonic solution, [30 mM Tris, 20% sucrose, 1 mM EDTA, pH 8.0] at a ratio of 1:10 wash solution to expression culture (v/v) while incubated for 10 minutes at room temperature. The cells were centrifuged (8000 x g, 4°C, 20 min), then the wash solution was discarded while the pelleted cells were resuspended in a hypertonic solution [5 mM $MgSO_4$] at 4°C for 30 minutes, causing the release of the periplasmic content with MBP-TrbC. After centrifugation (15,000 x g, 4°C, 20 min), the supernatant with the periplasmic contents was filtered using 0.45 μm syringe filter to remove any cellular debris prior to purification with amylose resin (NEB).

The protein is purified using a gravity flow amylose column initially equilibrated with the loading buffer [10 mM Tris, 50 mM NaCl, 1 mM EDTA, pH 7.4]. Since MBP has a higher affinity for maltose than amylose, upon the addition of an elution buffer with maltose the protein is displaced from the column. Prior to elution, the column was washed with five column volumes (CV) of loading buffer to remove unbound components, then the protein was eluted with a 100% step gradient of elution buffer [10 mM Tris, 50 mM NaCl, 1mM EDTA, 10 mM maltose, pH 7.4] over two CV. Fractions collected were analyzed by measuring the absorbance at 280 nm, and were further analyzed by SDS-PAGE. In cases where the removal of the MBP tag was required, this was acheived by incubating 25 μg Factor Xa (GE Healthcare) with ~20mg fusion protein

at 4°C for 48 hours, followed by the purification of the protein with the amylose resin again.

3.2.6 Interaction of His_6-TraW with His_6-TrbC and Untagged TrbC

Frozen cell pellet was resuspended on ice in lysis/loading buffer [10 mM imidazole, 30 mM Tris-HCl pH 8.2, 500 mM NaCl] and sonicated for 30 min of 15 s on cycle pulses at 30% amplitude with 30 s cooling time between each pulse. Soluble cellular lysates containing His_6-TraW (23.98 kDa) or His_6-TrbC (23.52 kDa) were collected by centrifugation at 45,000 x g for 45 min at 4 °C, after which they were mixed and purified on a 10 mL Ni^{2+} charged Ni-NTA resin (Thermo Scientific) column equilibrated with the loading buffer. Proteins were eluted using a step-gradient containing 60% elution buffer [30 mM Tris-HCl pH 8.2, 500 mM NaCl, 500 mM imidazole]. For the removal of the six-histidine tag (His_6) of TrbC (21.64 kDa) and TraW (22.17 Da), His_6-tagged proteins were incubated for 16 hrs at 21°C for TrbC, and/or 2 hrs at 37°C for TraW with 10 units/mg of thrombin (Sigma); to prevent TraW from degrading it was incubated at 37° for two hours. The protein sample was analyzed by size exclusion chromatography (SEC) on a HiPrep 16/60 Sephacryl S-100 HR column (GE Healthcare; separation range 1-100 kDa) at a flow rate of 0.5 mL/min to remove residual thrombin and the cleaved tag using the interaction buffer [20 mM HEPES pH 7.5, 150 mM NaCl]. The column was standardized with Blue Dextran 2000 (Vo = 36.2 mL), 2 mg/mL BSA (MW = 66.5 kDa; Ve = 44.7 mL) and 150 mM NaCl (Vt = 106.2

mL). To observe interactions between histidine-tagged and untagged TrbC with TraW, the three protein constructs were incubated for 2-16 hrs following which the protein mixtures were subjected to SEC with the interaction buffer at a flow rate of 0.5 mL/min.

3.2.7 Interaction of TraW-His_7 with TrbC-His_7

C-terminally tagged TraW (TraW-His_7, 22.93 kDa) and TrbC (TrbC-His_7, 22.53 kDa), were expressed and purified separately, and the soluble cell lysates were loaded onto a Ni-NTA column. Proteins were eluted from the column with 100% of elution buffer (the same as for the 6-His tagged proteins). Following separate purifications, the proteins were mixed and analyzed by SEC (as per section 3.2.6) using interaction buffer at a flow rate of 0.5 mL/min and SDS-PAGE.

3.2.8 Interaction of MBP-TrbC with TraW-His_7

Soluble periplasmic solution containing MBP-TrbC was combined with a cytosolic solution containing TraW-His_7 and loaded into Ni-NTA resin for co-purification using binding buffer [5% (v/v) glycerol, 10 mM imidazole, 20 mM $MgCl_2$, 20 mM HEPES pH 8.2, 300 mM NaCl]. Proteins were eluted using a step-gradient containing 25% elution buffer [5% (v/v) glycerol, 20 mM $MgCl_2$, 20 mM HEPES pH 8.2, 300 mM NaCl, 500 mM imidazole] and analyzed by SDS-PAGE.

3.2.9 Immunoblot Analysis

Proteins were separated on a 12.5% SDS-PAGE gel and transferred to nitrocellulose membrane (Bio-Rad) for 1 h at 120 V, following which the membranes then blocked with a 1% gelatin solution containing 0.05% Tween-20 overnight at 4°C. Histidine-tagged proteins were detected using standard procedures employing a primary mouse anti-His antibody (ZYMED Laboratories) at 1:10,000 dilution, and secondary goat anti-mouse alkaline phosphatase (AP)-conjugated secondary antibody (Sigma) at 1:30,000 dilution. To visualize the proteins of interest (via the AP-conjugated 2° antibody), the blots were incubated in a solution of 5-bromo-4-chloro-3-indolyl phosphate (BCIP) and nitroblue tetrazolium (NBT). AP dephosphorylates BCIP, yielding bromochloro indoxyl intermediate, which is then oxidized by NBT to produce an indigoid dye. In addition, NBT is also reduced by the indoxyl producing insoluble blue precipitate. The combination of both coloured precipitates make the proteins visible in the western blot.

3.3 Results and Discussion

Previous sequencing studies have shown *trbC* directly upstream of *traW* result in two different polypeptides in the F plasmid (Frost *et al*., 1994). These two genes, while expressed as two proteins in some systems, in others they are expressed as a single polypeptide with TrbC fused to the N-terminus of TraW (Figure 3.4) (Lawley *et al*., 2003). These data suggest that TrbC and TraW share a coupled function and interact

during T4SS assembly. However, to date no direct evidence has been reported to support the hypothesis that TrbC and TraW do indeed interact; we therefore endeavoured to characterize the interaction using size exclusion chromatography (SEC) and pull-down affinity interaction analysis.

3.3.1 SEC Elution profiles of TrbC

The TrbC protein (21.64 kDa) exists in a mixture of oligomeric states in solution, resulting in different elution volumes when analyzed by SEC (Figure 3.6). The larger oligomers elute from the column in a peak close to the void volume of the column at 40.4 mL (Peak 1, Figure 3.6a), whereas a TrbC dimer elutes after 48.9 mL (Peak 2) followed by minimal amounts of the monomer 68.3 mL (Peak 3). Similar results are observed for TrbC with or without histidine tags, indicating that the tag does not influence the tertiary structure of the protein (Figure 3.6c). Interestingly, the higher molecular weight oligomers are observed both in the SEC elution profile, but also when the peaks are analyzed by SDS-PAGE, indicating that stable TrbC dimers/oligomers form through reduction-stable covalent bonds (i.e. not disulfide linkages) (Figure 3.6b,d).

Similar to His_6-TrbC, the higher oligomeric forms of the protein following the removal of the His-tag elute in Peak 1 (lanes 3-5, Figure 3.6d), followed by a TrbC dimer in Peak 2 (lanes 6-9, Figure 3.6d). The SEC elution profile of TrbC due to the different oligomeric forms of the protein present in solution was observed to be the same for the protein with an N-terminal His-tag (His_6-TrbC; 23.52 kDa), C-terminal His-tag (TrbC-

His_7; 22.53 kDa), or with no tag (TrbC; 21.64 kDa) (Figure 3.6). These results, coupled with the other elution profiles of the protein, suggest that the presence of the higher molecular weight oligomers of TrbC are not as result of a tag-induced oligomerization of the protein. Table 3.1 summarizes the TraW and TrbC constructs used in this study, of which most constructs have similar molecular weights; the full sequences of the protein constructs can be found in Appendix B.

3.3.2 Identification of Dimerization of TrbC by Mass Spectrometry

To confirm the oligomerization of TrbC and identify the protein bands of the higher mass, always present in SDS gels with purified TrbC, in-gel trypsin digestion was performed. The protein bands from the gel were sliced and processed for analysis with MALDI-TOF MS method. The identity of the band at ~45 kDa was confirmed to result from the dimerization of TrbC (Figure 3.7).

3.3.3 Elution Profile of TraW from SEC

Under similar experimental conditions, His_6-TraW (23.99 kDa) elutes from the column in a single peak at an elution volume of 63.2 mL (Figure 3.8), which is indicative of a globular protein according to its molecular weight. Considering the similar size of His_6-TraW, (23.99 kDa), and that of His_6-TrbC (23.52 kDa), under normal and monomeric conditions, monomeric TrbC would be expected to elute from the column in Peak 3 (Figure 3.6a) under identical experimental conditions. Similar elution profiles

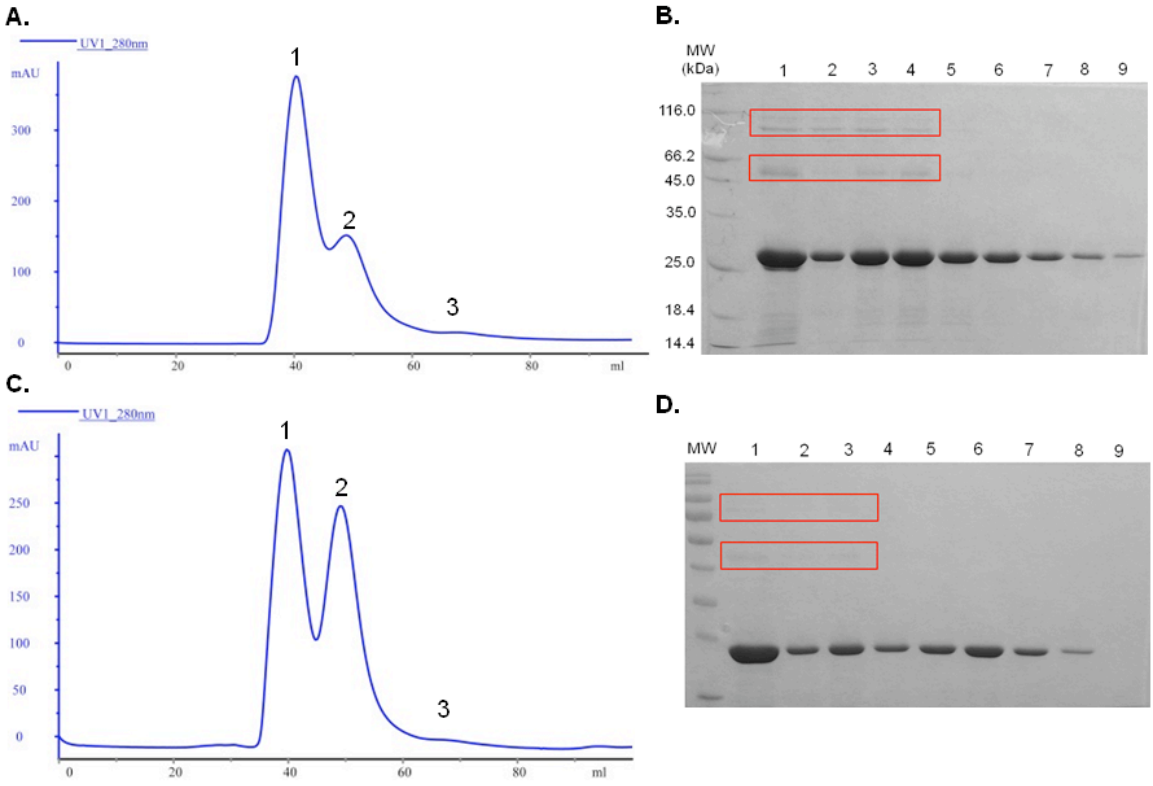


Figure 3.6. TrbC exists as multiple species in solution. **(A)** SEC analysis of purified His_6-TrbC at 4 mg/mL shows the presence of at least 3 oligomeric states of TrbC which elute in three peaks. **(B)** SDS-PAGE analysis of the SEC separated His_6-TrbC shows the presence of different oligomers of the protein in the corresponding elution fractions from the column. Lane 1, Ni-NTA isolated His6-TrbC; lane 2-4, Peak 1; lanes 5-7, Peak 2; lanes 8-9, Peak 3. In-gel trypsin digestion followed by mass spectrometric analysis of the contents of the higher molecular weight bands in lanes 1-4 (highlighted with red boxes) confirm that these species are indeed higher order oligomers of TrbC. **(C)** SEC analysis of TrbC following removal of the N-terminal His_6 tag results in a greater amount of TrbC oligomerization in solution, with a marked increase in the relative peak height of Peak 2 and decrease in Peak 3. **(D)** SDS-PAGE analysis of the SEC separated TrbC. Lane 1 Ni-NTA isolated TrbC (prior to removal of the His tag); lanes 2-4, Peak 1; lanes 5-8, Peak 2 and lane 9, Peak 3.

Table 3.1. The TrbC and TraW constructs used in this study*

Protein Construct	Molecular Weight (kDa)
TrbC	21.64
TrbC-His7	22.53
His6-TrbC	23.52
MBP-TrbC	66.93
TraW	22.17
TraW-His7	22.93
His6-TraW	23.99

*Purification tags are 6 or 7 histidine residues or Maltose binding protein (MBP). The tag on the right denotes C terminally tagged, the tag on the left denotes N-terminally tagged.

Match to: **gi|9507801** Score: **151**
conjugal transfer pilus assembly protein TrbC [Plasmid F]
Nominal mass (M_r): **23532**; Calculated pI value: **8.87**
Matched peptides shown in **Bold Red**

```
     1 MKLSMKSLAA LLMMLNGAVM ASENVNTPEN RQFLKQQENL SRQLREKPDH
    51 QLKAWAEKQV LENPLQRSDN HFLDELVRKQ QASQDGKPRQ GALYFVSFSI
   101 PEEGLKRMLG ETRHFGIPAT LRGMVNNDLK TTAEAVLSLV KDGATDGVQI
   151 DPTLFSQYGI RTVPALVVFC SQGYDIIRGN LRVGQALEKV AATGDCRQVA
   201 HDLLAGKGDS GK
```

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
68 - 78	1344.70	1343.69	1343.65	0.05	0	R.SDNHFLDELVR.K (Ions score 52)
90 - 107	2041.00	2039.99	2040.07	-0.08	1	R.QGALYFVSFSIPEEGLKR.M(Ions score 50)
114 - 122	1011.60	1010.59	1010.57	0.03	0	R.HFGIPATLR.G (Ions score 29)
142 - 161	2153.10	2152.09	2152.04	0.05	0	K.DGATDGVQIDPTLFSQYGIR.T(Ions score 20)

Figure 3.7. MALDI-TOF MS analysis for tryptic-digested SDS-PAGE band appearing at ~45 kDa which results due to the dimerization of the protein monomer TrbC ~23 kDa, even under the conditions of a denaturing SDS gel analysis.

were observed for TraW without or with both N-terminal or C-terminal His-tags, eluting in a single peak (Ve = 63 mL). It should be noted that TraW precipitates when stored at 4°C, yet degrades when stored at room temperature (21°C).

3.3.4 Interaction Studies of TraW-His_7 and TrbC-His_7 using SEC

C-terminally tagged TraW and TrbC (TraW-His_7 and TrbC-His_7) were purified separately prior to 3-4 hour incubation, concentration, followed by analysis using SEC. A shift in the elution profile of the proteins was observed (Figure 3.9a). Similar to the SEC elution of TrbC alone, three peaks were observed; TrbC is found primarily in Peaks 1 and 2 while TraW is observed to co-elute with TrbC in Peak 2 and elutes as a monomer protein in Peak 3 (Figure 3.9c). Considering the elution profile of TraW alone (Figure 3.8) where the protein elutes in a single peak at 63.2 mL, its elution in Peak 2 at 54.1 mL (Figure 3.9) can be associated with a larger molecular size as a result of its interaction with TrbC-His_7 . SDS-PAGE analysis of the contents of Peak 2 show that both TrbC and TraW are present **(**Figure 3.9b; lanes 5-7).

3.3.5 Pull-down Affinity Used for Probing Interaction of MBP-TrbC and TraW-His_7

TraW-TrbC interaction was confirmed with an affinity capture assay using TraW-His_7 as a bait. TraW-His_7 was isolated from cell lysate on a Ni-NTA column, following which a periplasmic cell lysate containing MBP-TrbC (66.93 kDa) was added.

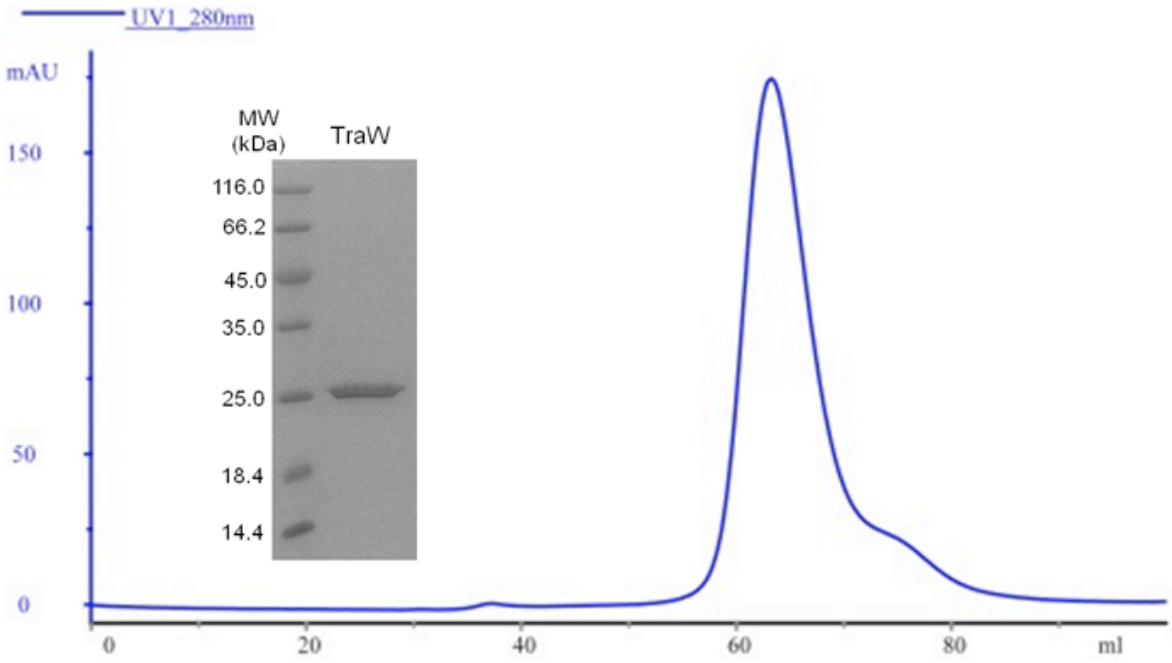


Figure 3.8. SEC analysis of His_6-TraW. Unlike TrbC, TraW (4 mg/mL) elutes in a single peak at an elution volume of 63.2 mL, corresponding to a monomeric species in solution. SDS-PAGE analysis of the contents of the single elution peak (inset) confirm the presence of a 23.99 kDa TraW monomer.

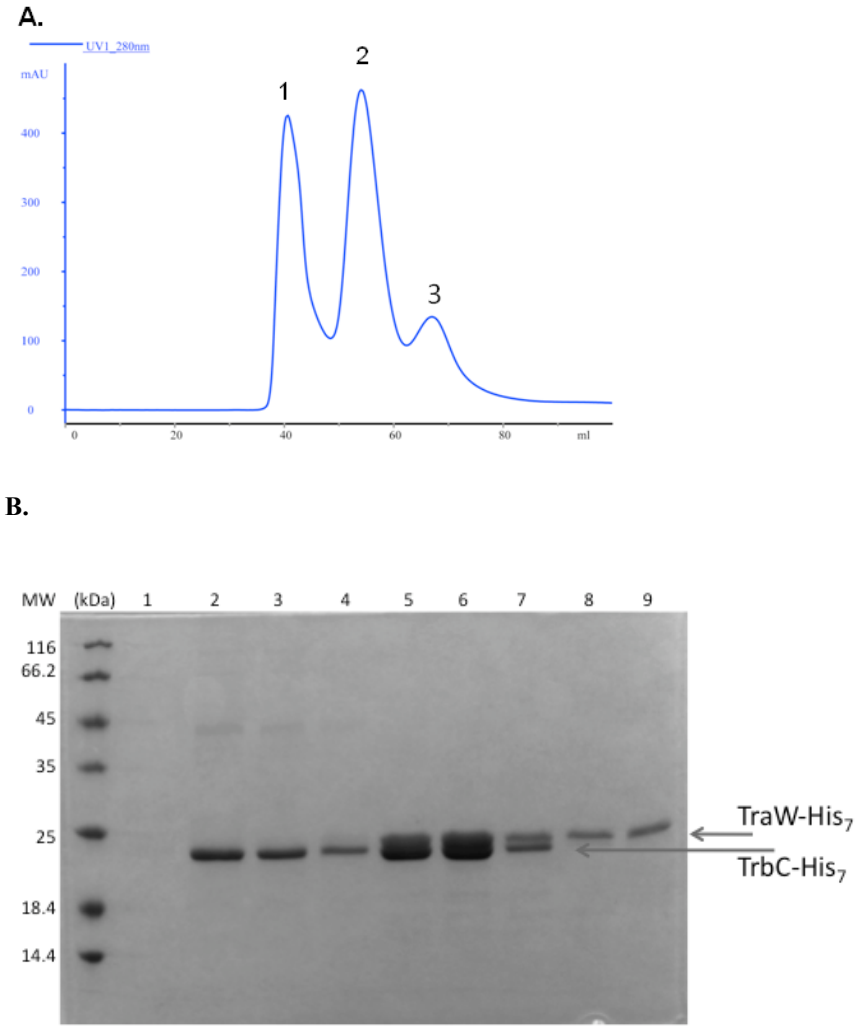


Figure 3.9. Interaction of C-terminally His-tagged TraW and TrbC. **(A)** SEC analysis of a TraW-His_7 (3 mg/mL) and TrbC-His_7 (3 mg/mL) mixture. The three observed peaks are indicative of multiple species being present in solution, and the contents of the peaks were confirmed by SDS-PAGE. **(B)** SDS-PAGE analysis of the SEC-separated TraW-TrbC peaks (graph A). Lanes 2-4 are from Peak 1, lanes 5-7 are from Peak 2 and lanes 8-9 are from Peak 3. The differences in molecular weight between TraW-His_7 and TrbC-His_7 allow for identification of the protein contents of each peak. Accordingly, peak 1 is composed primarily of TrbC oligomers (as determined by MS analysis), peak 2 is a TraW-TrbC dimer and peak 3 is primarily composed of TraW.

After a subsequent wash step for the removal of unbound proteins, TrbC-His_7 was eluted from the column with 125 mM imidazole, and the peak elution contents analyzed by SDS-PAGE. In aliquots corresponding to the imidazole eluted peak, a high molecular weight MBP-TrbC band is observed in addition to the expected TraW-His_7 band (Figure 3.10a). A negative control experiment was done to rule out the possibility of MBP-TrbC binding to the Ni^{2+} charged resin (Figure 3.10b).

It is interesting to note that this observation both confirms the SEC observations, and provides an insight into the orientation of the preferred TraW-TrbC interaction. Sequence alignment predictions (Figure 3.4) and gene expression (Frost *et al*., 1994) suggested that the C-trminal domain of TrbC interacts with the N-terminal of TraW. The presence of the large N-terminal MBP tag on TrbC and proximal location of the Ni-NTA resin to small C-terminal His_6 tag of TraW indicates that the TraW-TrbC interaction likely occurs between the C-terminal region of TrbC and N-terminal region of TraW. Steric bulk of the TrbC tag and Ni-NTA resin to which TraW is immobilized precludes any other orientation by which the two proteins could interact.

It is important to mention the unstable nature of TrbC, which has been found to degrade in a short period of time when purified. In addition, the low expression levels of the protein using a periplasmic system with the MBP tag imposed a challenge. However,

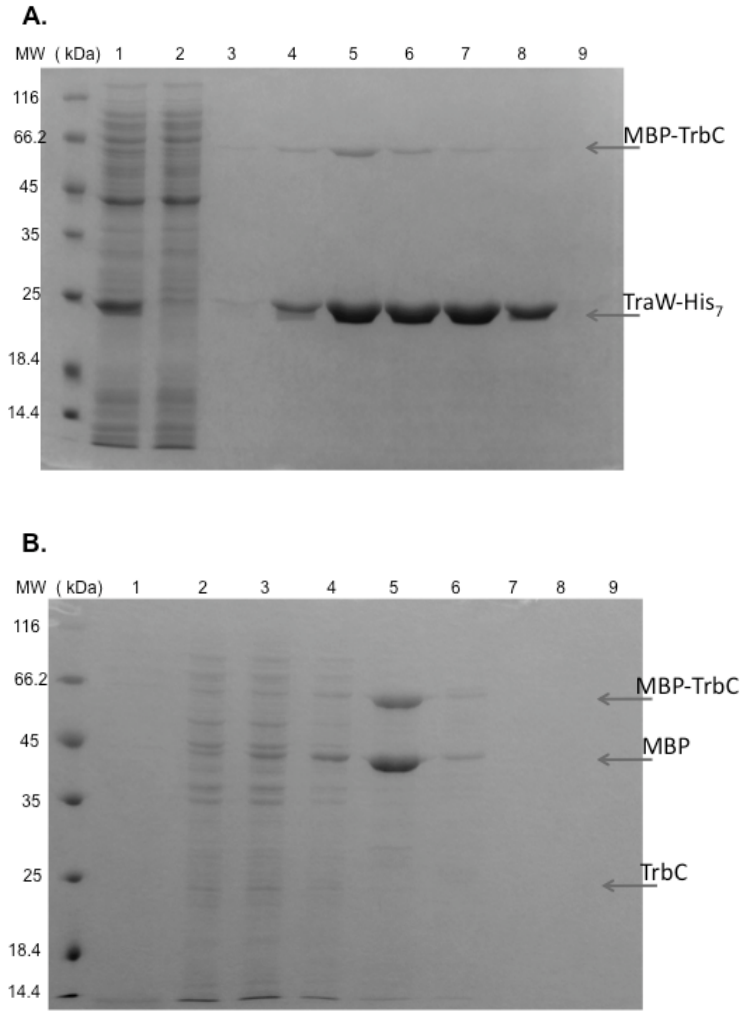


Figure 3.10. SDS-PAGE analysis of an on-column capture of MBP-TrbC by TraW-His_7. (**A)** TraW-His7 was bound to Ni-NTA resin, following which MBP-TraW containing cell lysate (Lane 1) was added to the column; Lane 2, flow through; lane 3, wash; Lanes 4-9, the TrbC-TraW complex eluted with 125 mM imidazole. (**B**) Negative control SDS-PAGE gel, Lanes 2-3, flow through; lanes 4-5, wash of unbound; lanes, 6-7, elution with 50 mM imidazole; lanes 8-9, elution with 125 mM imidazole. The presence of the N-terminal MBP tag on TrbC and proximal location of the Ni-NTA resin to the C-terminus of TraW may suggest that the TraW-TrbC interaction occurs between the C-terminal region of TrbC and N-terminal region of TraW.

despite these challenges, it can be seen that the interaction of TraW with TrbC also has a stabilizing effect on TrbC, as suggested by the amount of MBP tag cleaved in the negative control gel, without the presence of TraW (Figure 3.10).

3.3.6 Preferred Interaction of TraW with the Free N terminal end of TrbC

While our results show the proteins interacting in SEC, based on their elution profiles with the same tags, and also in Ni-NTA with different affinity tags, we aimed to confirm if there is a preference between the interaction of TraW with the C or the N terminal domain of TrbC. The F plasmid map reported by Frost *et al.* (1994), shows that the *traW* gene is encoded before *trbC*. On the other hand, when the two proteins are fused into one gene (Figure 3.4), the TrbC sequence is encoded before TraW (Lawley *et al*., 2003). This indicates a higher probability of TraW interacting with the C-terminal end of TrbC in cases when the proteins are expressed separately, as in the F plasmid. Additionally, in the UWE25 plasmid, (*Protochlamydia amoebophila)* where the proteins are encoded separately, *traW* is followed by *trbC*, without *traU* separating the two as seen in the F plasmid (Greub *et al*., 2004).

When the three different proteins His_6-TraW (23.99 kDa), His_6-TrbC (23.52 kDa), and TrbC (21.64 kDa) are co-eluted and separated by SEC, a preference for interaction between His_6-TraW with untagged TrbC is observed. (Figure 3.11). His_6-TrbC and untagged TrbC proteins behave similarily when analyzed individually by SEC in the presence of TraW. However, His_6-TraW co-elutes from the mixture with untagged TrbC

in Peak 2 at 50.0 mL elution volume (Figure 3.11a) and not His_6-TrbC (Figure 3.11b, lanes 5 & 6). This suggests that TrbC is preferentially interacting with TraW, and that N-terminally tagged TrbC has a lower potential to interact with TraW due to its histidine tag. This also could suggest that the interaction between TrbC and TraW is stronger than that between TrbC homodimer.

To confirm the identity of the tagged proteins a Western Blot was done in parallel with a 12.5% SDS-PAGE using anti-His primary antibody to detect His_6-TraW and His_6-TrbC (Figure 3.11b). Monomeric His_6-TrbC was observed in lanes 9 and 10, which were collected in Peak 3 (Ve = 63.7 mL). His_6-TraW (detected on the transfer blot) with TrbC (not detected on the blot, but present from the SDS-PAGE) is observed in lanes 5 and 6 (Figure 3.11b) due to the co-elution of the proteins from the SEC column (Figure 3.11a; Peak 2).

To put all this into perspective, Figure 3.12 shows the overlay (bottom panel) of the different elution profiles (top panels). Monomeric His_6-TraW (brown) elutes at Ve = 63 mL. Untagged TrbC (Blue) elutes at Ve = 39 mL as a tetramer (86.56 kDa) and at Ve = 50 mL as a dimer (43.28 kDa). When TraW is incubated with His_6-TrbC and untagged TrbC (red), they co-elute in Peak 2 at Ve = 50 mL, due to the interaction between TraW with untagged TrbC (total MW = 45.63 kDa).

3.3.7 Interactions of TrbC with Untagged TraW

Similar experiments using N-terminally tagged constructs of His_6-TrbC (23.52 kDa) and His_6-TraW (23.99 kDa) were analyzed with SEC to show interaction between TrbC and TraW. The removal of the His-tag from TraW (22.17 kDa) was never desirable, since the tag was never completely cleaved when incubated over 16-20 hrs at 21°C with thrombin, while longer incubation times would cause protein degradation. After purifying both proteins separately with Ni-NTA resin, cleaving His_6-tag from TraW, followed by the removal of thrombin with a SEC for TraW only, the cleaved TraW and His_6-TrbC were incubated prior to loading into SEC jointly. The results also confirm the interaction between TraW and His_6-TrbC (Figure 3.13). TraW is found in Peak 1 as well as Peak 2, but not in Peak 3 where the homogeneous monomeric TraW elutes. On the other hand, His_6-TrbC is found mainly in Peak 3, while the rest is in Peak 1 with TraW, as indicated by the gel lanes 2 and 3 (Figure 3.13a,b). It is not clear whether TrbC is found in Peak 2, since all what is visible from the gel is TraW and residual His_6-TraW. The presence of TraW in Peak 1 with His_6-TrbC confirms the protein-protein interaction; this analysis was conducted in triplicate, with similar results.

These observations presented here clearly indicate an interaction between TrbC and TraW of the F-plasmid T4SS. While there is an interaction between different His-tagged proteins on either terminal end, there is preferential interaction of TraW and the untagged N-terminal end of TrbC. Our findings suggest that during the F pilus assembly,

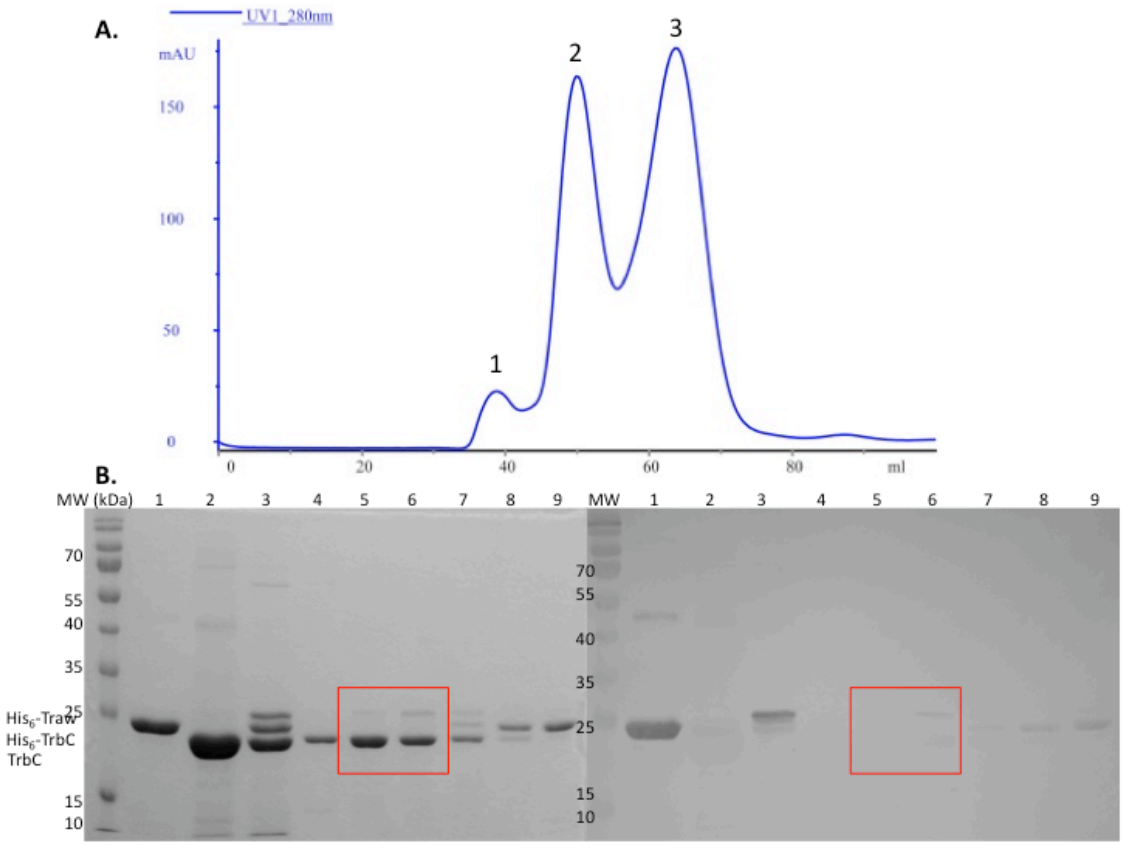


Figure 3.11. Interaction of N-terminally His-tagged TraW and TrbC in the presence of His-tagged TrbC. SEC elution profile of a His_6-TraW/His_6-TrbC/TrbC mixture lined up in decreasing size order. **(A)** SEC elution profile of the His_6-TraW/His_6-TrbC/TrbC mixture, again indicating multiple species in solution. **(B)** SDS-PAGE (left) and Western blot (right) analysis of SEC separated peaks. Western blot analysis was conducted using an anti-His_6 primary antibody. Lanes 1-3 are His_6-TrbC, TrbC and the His_6-TraW/His_6-TrbC/TrbC mixture, respectively; lane 4 is from Peak 1 corresponding to TrbC; lanes 5-6 from Peak 2 corresponding to His_6-TraW/TrbC; lanes 7-9 from Peak 3 corresponding to mainly His_6-TrbC. Comparison of the SDS-PAGE and Western blots shows the presence of distinct TraW bands in lanes 5-6 (highlighted red box), with corresponding untagged TrbC. This indicates that TraW preferentially interacts with untagged TrbC, with residual TraW then interacting with His6-TrbC (lane 7); the remaining unbound His6-TrbC elutes in Peak 3 (lanes 8-9) as confirmed by the anti-His antibody.

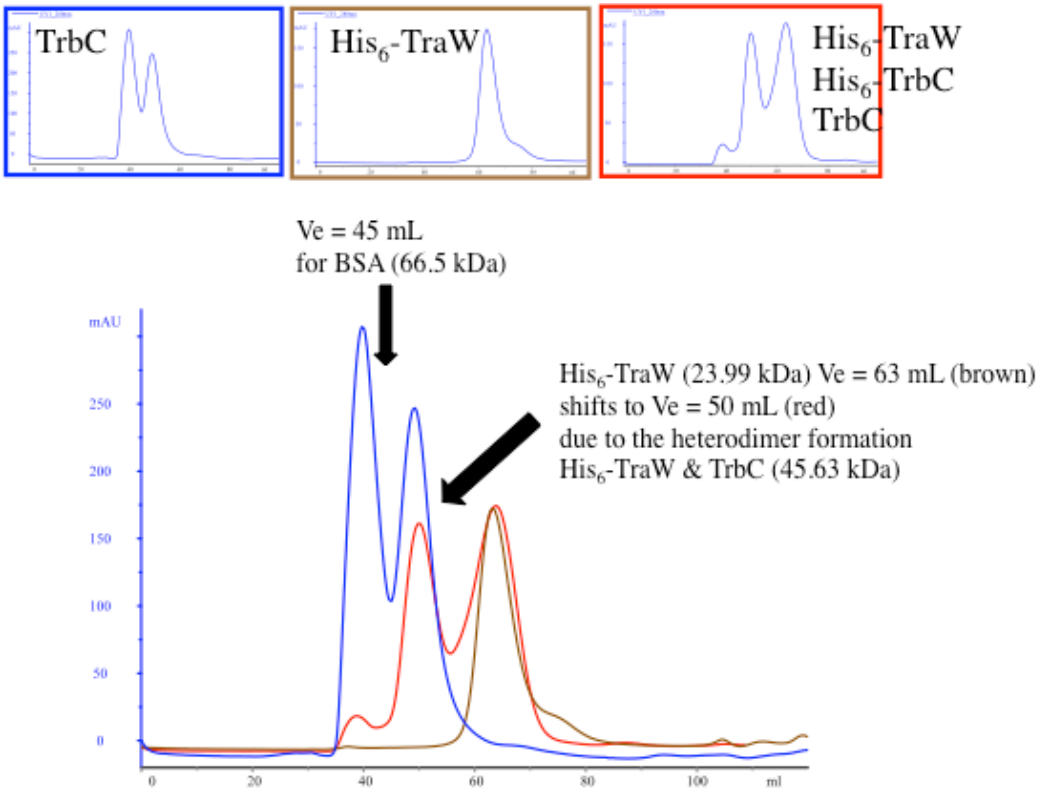


Figure 3.12. Overlaying of the individual elution profiles of TrbC (blue), His_6-TraW (brown), and all three His_6-TraW, His_6-TrbC and TrbC mixed in (red). His_6-TraW (brown) where the TraW elutes in a single peak 63 mL. Untagged TrbC (blue) has a similar profile to the tagged TrbC, the protein elutes in two peaks 39 mL (tetramer) and 50 mL (dimer). The co-elution profile all three proteins is shown by the graph in red, where the elution occurs in three different peaks, 39 mL, 50 mL, and 63 ml; in this case however TraW co-elutes with TrbC in the second peak at 50 mL (heterodimer) while His_6-TrbC elutes in the third peak at 63 mL (monomer). The black arrow at 45 mL indicates the elution of the BSA standard with the MW = 66.5 kDa.

mating pair formation and bacterial conjugation the two proteins interact and have a common role even when they're expressed separately. Further characterization of this interaction will lead to a better understanding of F-mediated bacterial conjugation, and potentially leading to more effective strategies targeting microbial plasmid delivery, which could offer benefits for public health in addition to more effective treatments of bacterial infections and reduction of the spread of multidrug resistance.

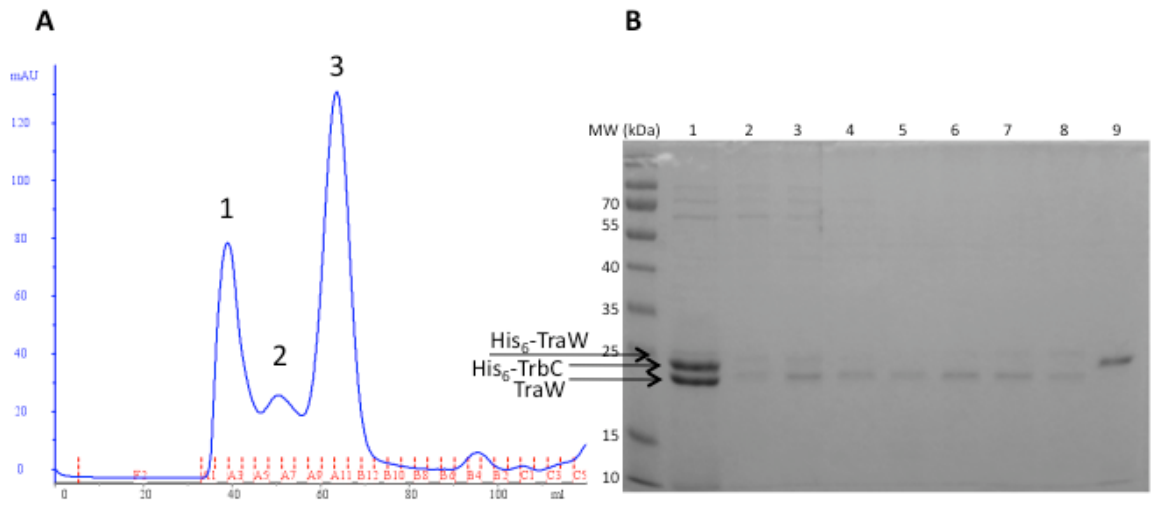


Figure 3.13. Interaction of N-terminally His-tagged TrbC and untagged TraW, in the presence of residual His_6-TraW. **(A)** SEC analysis of the protein mixture with mostly His_6-TrbC and TraW. The three observed peaks are indicative of multiple species being present in solution, and the contents of the peaks was confirmed by SDS-PAGE. **(B)** SDS-PAGE analysis of the SEC-separated TraW-TrbC peaks. Lane 1, mixed proteins before SEC; Lanes 2-4 are from Peak 1, lanes 5-8 are from Peak 2 and lane 9 is from the centre of Peak 3. Accordingly, peak 1 is composed primarily of TrbC oligomers including TraW and His_6-TrbC, peak 2 is a TraW and residual His_6-TraW and peak 3 is primarily composed of His_6-TrbC.

Chapter 4

Purification of TrbC and Crystallization of TraW from the *Escherichia coli* F Plasmid

4.1 Introduction

In gram-negative bacteria, construction of the secretion system pore requires localization of various structural and catalytically active proteins at the bacterial membranes and within the periplasm. To ensure proper folding and protein assembly in the periplasm, bacteria produce proteins that promote disulfide bond formation, generally between cysteine-containing proteins (Hemmis & Schildbach, 2013). The F plasmid of *E. coli* encodes for thirty-five different proteins, where ten of them have at least one cysteine residue, localized or partially associating with the periplasm (Frost *et al*., 1994). Recent data suggests that TrbC and TraW, as periplasmic proteins, both containing two and one cysteine residues, respectively, play an important role in the conjugative processes as mediators in folding and assembly of the pore complex proteins by inter- and intramolecular thiol redox reactions (Hemmis & Schildbach, 2013).

As two hallmark proteins from the *E. coli* F plasmid, TrbC and TraW are important for bacterial conjugation, due to their role in the F pilus assembly (Anthony *et al*., 1999; Arutyunov *et al*., 2010). Based on its amino acid sequence, TrbC is a 22.5 kDa hydrophilic protein with 32% hydrophobic content, which suggests peripheral membrane associations (Frost *et al.,* 1994) and exists as a dimer/oligomer (Chapter 3, in particular

Figure 3.7). In addition, TrbC is correctly processed in the presence of TraN, suggesting an interaction between these two proteins, which are encoded on adjacent genes of the F plasmid (Figure 3.3). On the other hand TraW is a 210 amino-acid, 23.6 kDa protein (Maneewannakul *et al*., 1992) found to interact with TrbB (Harris & Silverman, 2004), a disulfide bond isomerase (Hemmis & Schildbach, 2013). In addition, TrbC and TraW have thioredoxin-like signature folds, βαβαβ or ββα, as predicted by Phyre^2 (Kelley & Sternberg, 2009) (Figure 4.1) but they lack the signature sequence C-X-X-C motif that is specific to thioredoxins (Martin, 1995). Based on a 13% matched sequence identity (mainly attributed by the C-terminal domain of TrbC, where the two cysteine residues are found) with 90% confidence TrbC is predicted to be a thioredoxin-like protein. Similarly, TraW with 11% sequence identity and 45% confidence in predicted to be a thioredoxin-like protein.

Based on the predicted redox properties of the proteins due to the two cysteine residues, TrbC could potentially be similar to a substrate for thiol-oxidase DsbA (Figure 4.2a), whereas TraW (with one free cysteine) could be a substrate to a disulfide bond isomerase/sulfenate (–SOH) DsbG, which oxidizes free thiols (Figure 4.2b). However, without evidence of partricular protein function other than being required for F plius formation (Frost *et al*, 1994; Lawley *et al*, 2003), efforts to crystallize TrbC and TraW were undertaken to provide a structural analysis of the proteins to facilitate functional characterization. While challenging to obtain suitable quantities of solbule protein of significant puriity, initial crystals of TraW have been obtained from which optimization can be undertaken.

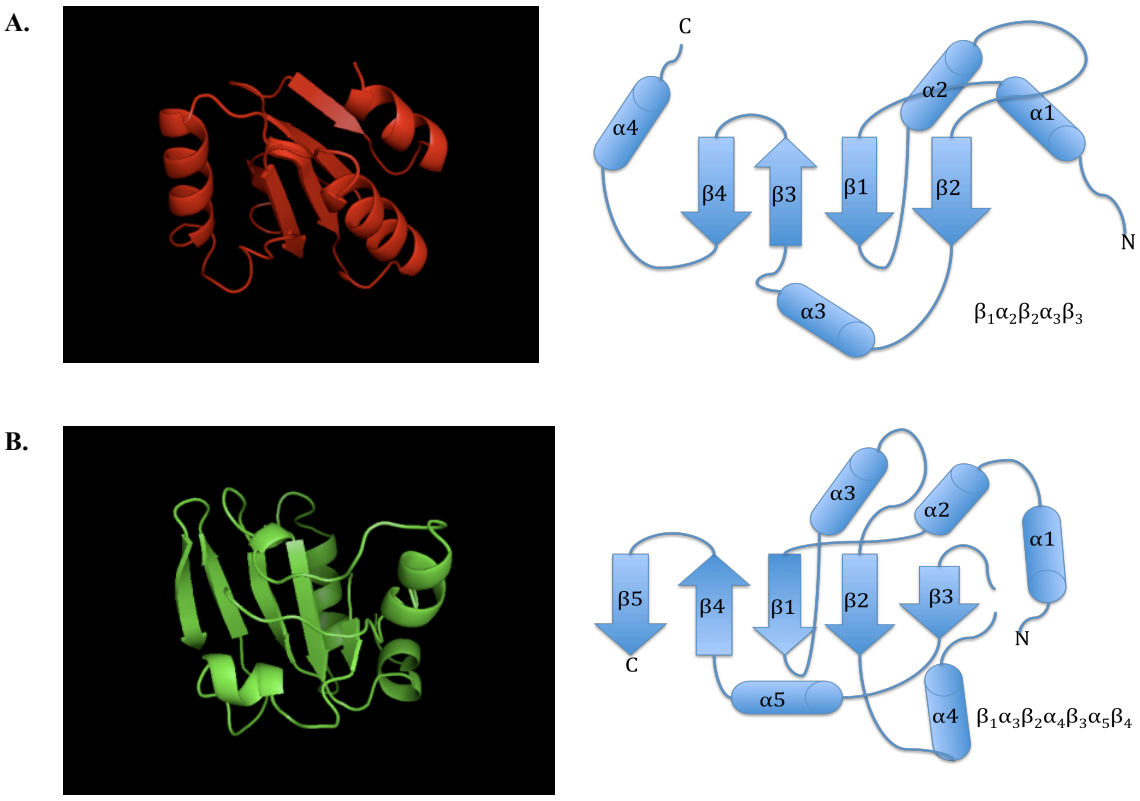


Figure 4.1. Structure predictions for TrbC and TraW. **(A)** Structure similarity of TrbC with c2iWtA protein, with an oxidoreductase fold, belonging to thioredoxin isoform superfamily and thioredoxin with a disulfide complex family. 3-D structure generated by PyMol, and 2-D topology diagram showing the signature βαβαβ thioredoxin-like fold. **(B)** Structure similarity of TraW with d1zzoa1 protein, with a thioredoxin fold, belonging to Thioredoxin-like superfamily, and Glutathione peroxidase-like family. 3-D structure generated by PyMol and 2-D topology showing the signature thioredoxin fold βαβαβαβ.

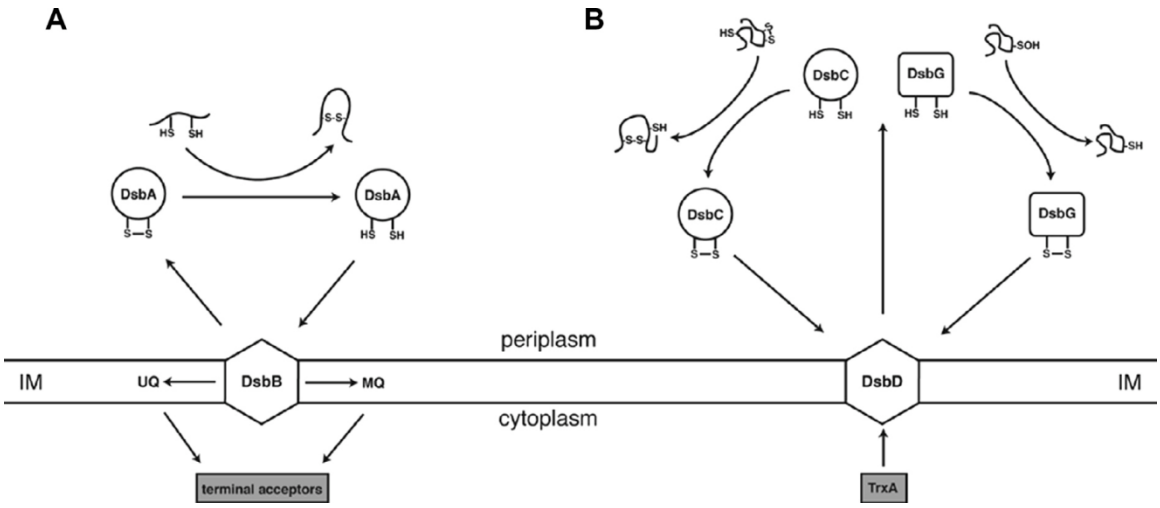


Figure 4.2. Periplasmic redox systems in *E. coli*. (**A**) Thiol-oxidase DsbA oxidizes a substrate in the periplasm with multiple cysteine residues that are not in close sequence proximity, such as TrbC. The reduced DsbA donate electrons to the membrane bound DsbB, to recycle the electrons, and resume active state. (**B**) Disulfide bond isomerase DsbC and disulfide bond sulfenate (-SOH) reductase DsbG become oxidized upon isomerization from DsbC, or sulfenate reduction of a substrate with a free thiol, such as in TraW, by DsbG. Both proteins accept electrons from the cytoplasm via TrxA into the periplasm through DsbD, to resume active states. Copied from Hemmis & Schildbach, 2013.

4.2 Materials and Methods

4.2.1 Cloning and purification of the C-terminal domain of TrbC ($TrbC^C$)

Phusion® High-Fidelity DNA Polymerase (NEB) was used to amplify the C-terminal sequence of TrbC (NP_061467.1; residues 102-212) (Frost *et al.*, 1994) using pT7-7::TrbC construct as a template, with PCR primers: forward, 5'-GCTT*CCATGG*GCGAAGAGGGGCTGAAACGAATG-3' and reverse, 5'-GCGTT*CTCGAG*TTTCCCGGAATCTCCTTTCCC-3', using restriction sites NcoI and XhoI respectively (shown in italics). The resulting PCR product was purified from a 1% agarose gel using Qiaquick Gel extraction kit (Qiagen), then double digested with NcoI-HF and XhoI restriction enzymes. Following a reaction cleanup with Gene Jet PCR kit (Fermentas), the digested PCR product (332 bp), was ligated at 16°C overnight into a pET28a similarily digested with NcoI and XhoI. The resulting construct was used to transform into DH5α cells for storage. DNA sequencing analysis and positive colony PCR reactions determined the correct cloning of $TrbC^C$. The final pET28a::$TrbC^C$ product was transformed into *E. coli* BL21 (DE3) cells for expression. The $TrbC^C$ construct contains a C-terminal 6-His tag; purification of $TrbC^C$-His_6 proceeded under identical conditions as for the full length TrbC or TraW (Chapter 3). SEC analysis (Figure 4.3a) indicates that $TrbC^C$ possesses the same properties as the full length protein, and exists as oligomers in solution, which was confirmed by SDS-PAGE (Figure 4.3b; Peak 1 = lanes 3-4, Peak 2 = lanes 7-9).

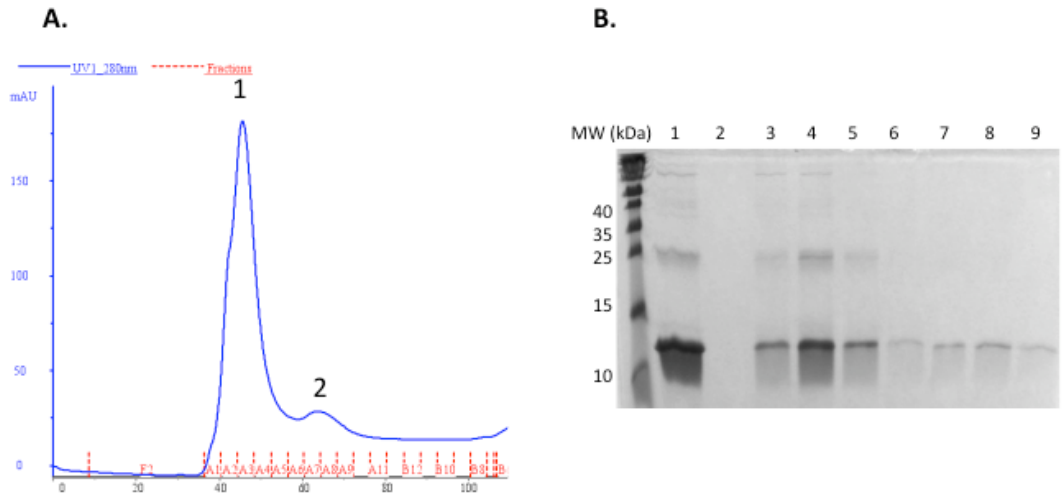


Figure 4.3. TrbC^{C} exists as multiple species in solution. **(A)** SEC analysis of purified TrbC^{C}-His$_6$ shows the presence of at least 3 oligomeric states of TrbC which elute in two peaks. **(B)** SDS-PAGE analysis of the SEC separated TrbC^{C}-His$_6$ shows the presence of different oligomers of the protein in the corresponding elution fractions from the column. Lane 1, Ni-NTA isolated TrbC^{C}-His$_6$; lane 3-5, Peak 1; lanes 6-9, Peak 2.

4.2.2 Expression and purification of TrbC

Full-length TrbC was expressed and purified following the same methods as discussed in Chapter 3. Gravity flow purification or purification using automated methods with ÄKTApurifier 10 (GE Healthcare) resulted in similar amounts of protein of similar purity. A typical purification and analysis of TrbC is shown in Figure 4.4.

Various additives were used during the purification of TrbC in an effort to increase the solubility and reduce aggregation of the protein. Additives included 0.05-0.2% NP-40, 0.05%-1% Triton X-100, 1-3 mM Dithiolthreitol (DTT), 1-5 mM Tris(2-Carboxyethyl)-phosphine Hydrochloride (TCEP HCl), 4 mM Glutathione (GSH), 5-15% glycerol as well as 1-2 % n-octyl-β-D-glucopyranoside (OG).

4.2.3 Expression and purification of TraW

Expression of full-length TraW was carried out as described in Chapter 3. To purify TraW for crystallization the frozen cell pellet was resuspended in 1:10 (w/v) lysis/loading buffer [500 mM NaCl, 30 mM Tris, 10 mM imidazole, pH 8.2]. The resuspended cells were lysed by sonication (Sonic Dismembrator 500, Fisher Scientific) while cooling on ice for 5 min (10s on/15s off). The insoluble cell lysate was removed by centrifugation at 34,000 x g for 30 min at 10 °C. The supernatant was loaded onto a pre-equilibrated 12 mL Ni^{2+} charged Ni-NTA resin (Thermo Scientific) and the column was subsequently washed with at least 5 column volumes (CV) of the loading buffer to

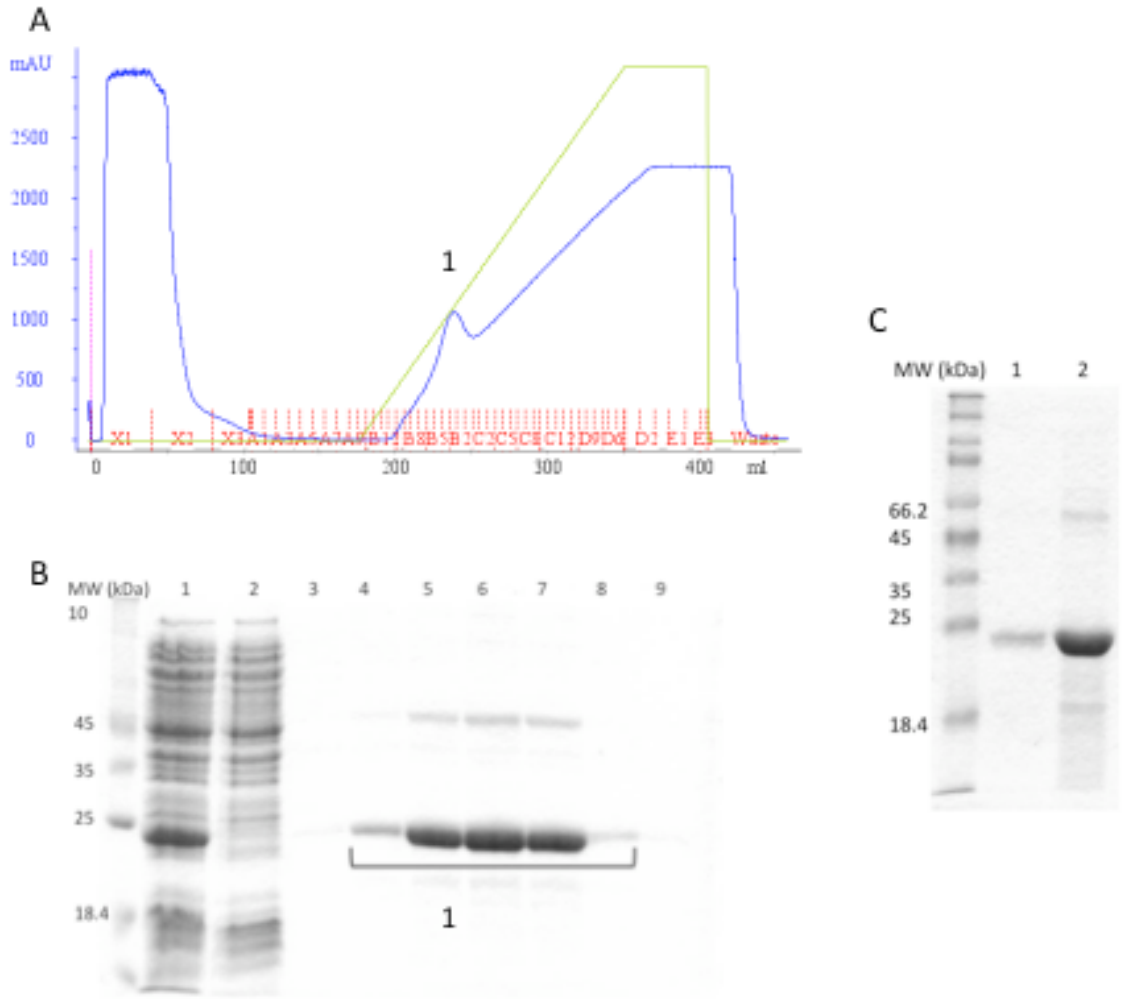


Figure 4.4. TrbC-His$_7$ purification using Ni-NTA affinity chromatography. (**A**) Chromatogram showing the elution of the protein during a linear gradient from 10-500mM imidazole, where the protein elutes during peak 1, with 125-150 mM imidazole. (**B**) The corresponding SDS-PAGE results. Lane 1, cell lysate; lane 2 flow through; lanes 4-8 protein elution from the column with the addition of 125-150 mM imidazole. (**C**) The final protein product after purification with SEC, lane 1; and concentrated to 4 mg/mL.

remove any contaminants. The protein was purified from the column using a step gradient with elution buffer [500 mM NaCl, 500 mM imidazole, 30 mM Tris, pH 8.2]. After washing the column with 5 CV of loading buffer, followed by 5 CV of 10 % elution buffer. The bound protein was eluted over 5 CV with 60 % elution buffer (300 mM imidazole) at a gravity flow rate of 0.5 mL/min. The fractions containing TraW were incubated with 10 U/mg thrombin overnight at 21°C for the cleavage of the His tag. The protein solution was concentrated using a 10 kDa MWCO centrifugal concentrator (Millipore) at 4,000 x g and 14°C to a final volume of 5 mL. TraW was further purified using SEC on a HiPrep 16/60 Sephacryl S-100 HR column (GE Healthcare) for buffer exchange [200 mM NaCl, 20 mM Tris, pH 7.0] and the removal of the His tag. The isolation of the protein in this manner yields TraW in over 90 % purity (Figure 4.5), as monitored by SDS-PAGE. Fractions corresponding to TraW without the His tag were collected and concentrated to 5-8 mg/mL prior to crystallization.

4.2.4 Crystallization of TraW

Initial crystallization experiments were performed by screening conditions in commercially available kits: Classics Suite, JCSG I, JCSG II, JCSG III, JCSG IV, and JCSG + from Qiagen (Lesley & Wilson, 2005). The trials were set in 96-well sitting drop plates (Axygen) by hand with 1 μL protein solution concentrated to ~8 mg/mL, and mother liquor in a 1:1 ratio over a reservoir containing 150 μL mother liquor. The crystal trays were stored at 21 °C and 4°C. Crystal hits formed were found only in one condition

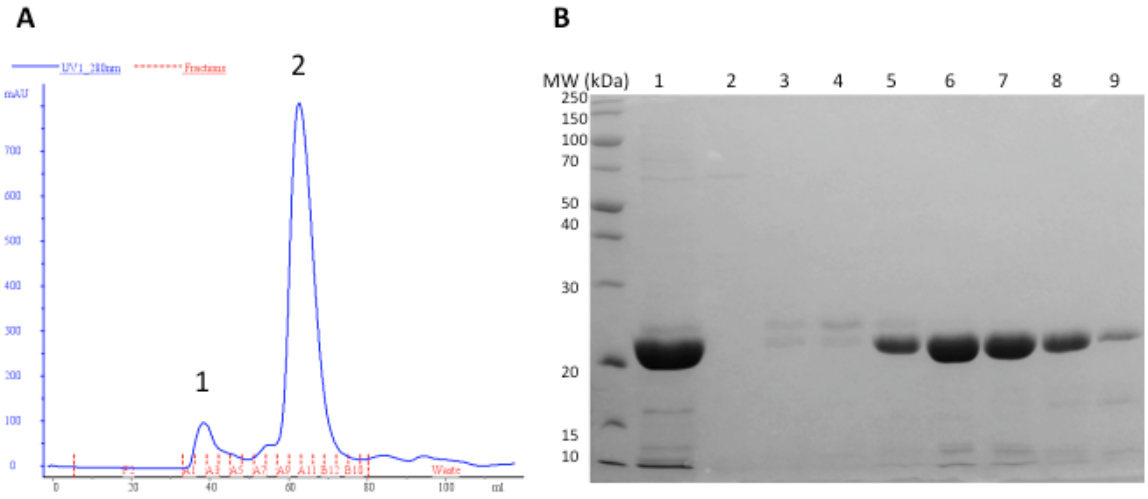


Figure 4.5. TraW purification using SEC gel filtration, after purification with Ni-NTA and thrombin cleavage of the His tag. (**A**) Chromatogram showing purification of the protein, peak 1, high molecular mass impurities, whereas the ~90% pure TraW elutes during peak 2. The shoulder on the left of peak 2 is a result of the mixed cleaved and uncleaved His tag (**B**) The corresponding 12% SDS-PAGE gel. Lane 1, concentrated TraW after isolation with Ni-NTA; Lane 2, contents of peak 1 from the chromatogram; lane 3-4, peaks corresponding to the shoulder on the left of peak 1; lanes 5-9, fractions corresponding to peak 1, where untagged TraW elutes from the column.

from JCSG II condition # 20 [0.2 M Magnesium chloride, 0.1 M imidazole pH 8.0, 40% (v/v) MPD] and only at 4°C. Optimization of the crystallization conditions was attempted by varying the precipitant concentration and the pH, while using hanging drop 24 well plates with cover slips (VWR), in addition to varying the ratio of protein to the mother liquor on the drop equilibrated against 0.5 mL of reservoir; the Additive Screen (Hampton Research) was also used for optimization; however, the attempts to get larger and more stable crystals were unsuccessful. Very thin and 15-50 μm long crystals appeared, which would dissolve in 2-3 days; while reproducible, the crystals did not improve upon optimization. The same crystals were obtained using TraW with the C-terminal 7-His tag (TraW-His_7).

4.2.5 Circular Dichroism Spectroscopy (CD) of TraW

To determine the effect of n-octyl-β-D-glucopyranoside (OG) on TraW, CD spectra (200-260 nm) were acquired using a Jasco J-810 CD spectrometer, with a 0.1 cm path length cuvette at 22°C. The protein was prepared using Ni-NTA, followed by SEC with [150 mM NaCl, 30 mM Tris, 10% (v/v) glycerol, pH 7.6] and concentrated to 5 mg/mL. Spectra were obtained of the protein alone, and in the presence of 1% (w/v) OG to compare the protein in the presence or absence of the detergent. Spectra were obtained in duplicate and the scans were corrected for the buffer contribution through the use of a buffer-only spectra.

4.3 Results and Discussion

4.3.1 TrbC Thioredoxin-Like Properties

With the high probability that TrbC is a thioredoxin-like protein, purification and crystallization of the protein under various reducing conditions was attempted in an effort to obtain soluble, homogeneous, and therefore crystallizable protein (McPherson, 1999). Under non-reducing conditions the protein exists as a mixture of monomer and oligomers (Chapter 3), making crystallization incredibly challenging. During purification with Ni-NTA and in the presence of high concentrations of imidazole, the protein was found to reduce the Ni^{2+} in the column, as indicated visually by the change in colour of the resin from green to yellow, suggesting redox properties of the protein. Different reducing agents were utilized when purifying the protein; while they helped with increasing the protein solubility, they were not effective in that they induced degradation of the protein (Figure 4.6a). Similar observations were noted with BME, DTT, as well as TCEP; which is more effective and more commonly used for crystallization because of its stability as a reducing agent (Han & Han, 1994). On the other hand purification under the same conditions with SEC and buffer [150 mM NaCl, 30 mM Tris-HCl, pH 8.0] and addition of 4 mM GSH proved to be more effective. While GSH does not increase the solubility of the protein it reduces its oligomerization, and increases the homogeneity of the sample (Figure 4.6b).

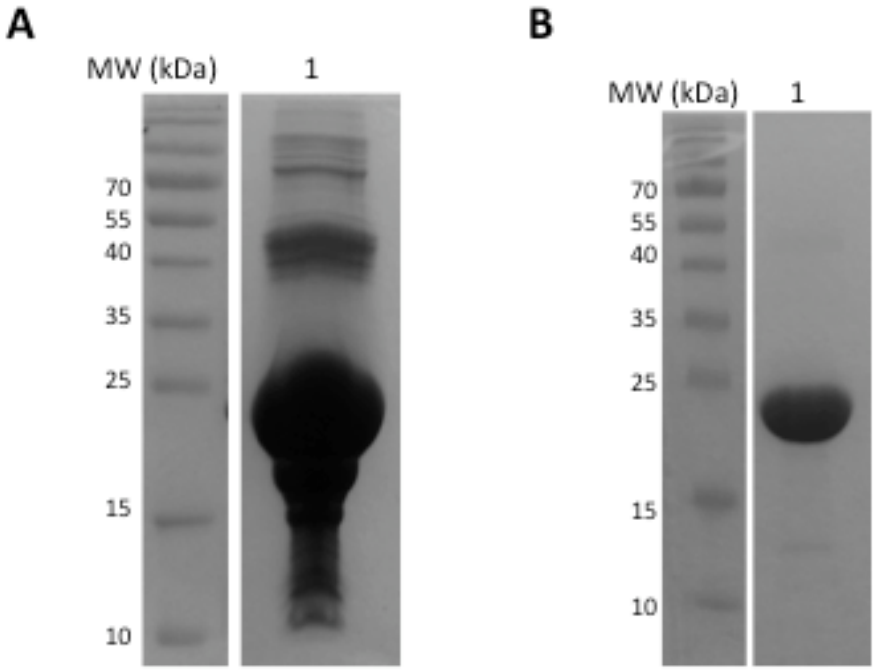


Figure 4.6. The affect of different reducing agents on TrbC. (**A**) TrbC after purification with SEC and [150 mM NaCl, 30 mM Tris-HCl, pH 8.0, + 5 mM BME]; the protein becomes more soluble but denaturation is increased. (**B**) TrbC after purification by SEC in the presence of 4 mM GSH. While the solubility of the protein is not affected, the stability of the portein is increased.

Although crystallization trials of TrbC in the presence of GSH as an effective reducing agent were undertaken, protein crystals were not observed. Given the properties of thioredoxin-like proteins, which can take months to crystallize while the protein slowly oxidizes (Weichsel *et al*., 1996), coupled with the unstable nature of TrbC, we were not able to obtain TrbC protein crystals suitable for X-ray diffraction analysis.

While the full-length TrbC was difficult to work with, constructs of the N-terminal domain and the C-terminal domain of the protein were cloned. The N-terminal domain construct was not a successful target as the protein would not express in *E. coli* (BL21) cells. On the other hand the C-terminal domain, $TrbC^{C}$, did not express well, however it has the same properties as the full-length protein (Figure 4.3) and it oligomerizes; crystallization of this construct was also attempted. Taking into account that both cysteine residues reside in the C-terminal domain of the protein (see Appendix B), it is possible that this region of the protein is responsible for the thioredoxin-like properties of TrbC.

4.3.2 Crystallization of TraW

With one free thiol, as a result of a single cysteine residue in its sequence, TraW was not as difficult to express and purify. Crystallization of the protein, on the other hand is a challenge. TraW is found to aggregate and precipitate at 4°C, whereas at room temperature at 21°C the protein degrades. This could due to the protein's thermal

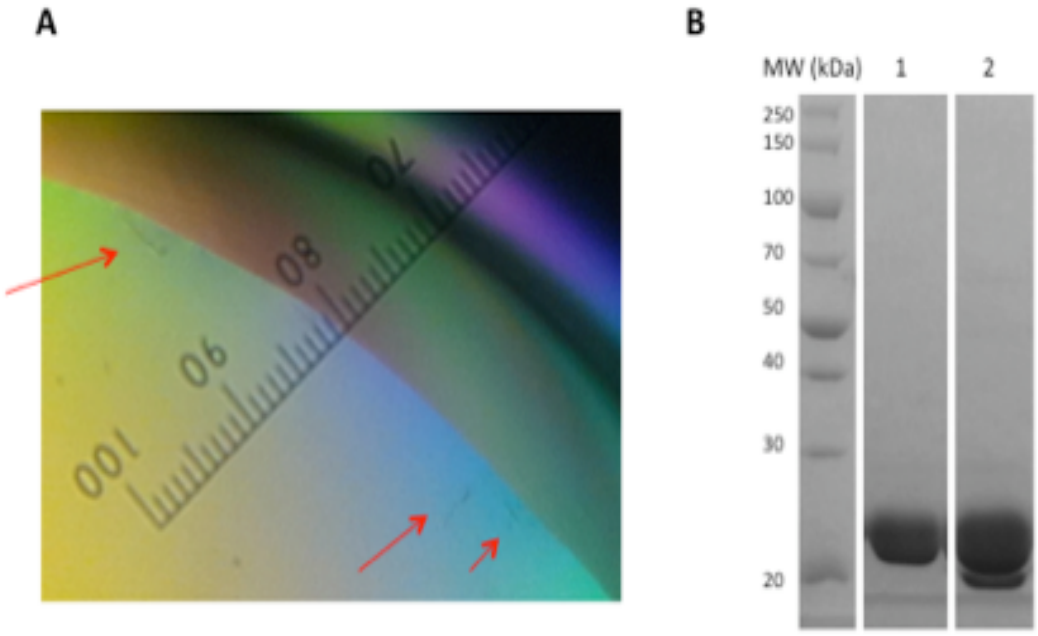


Figure 4.7. Crystallization of TraW at 4°C. (**A**) Small crystals with 15-50 μm length formed in 2-3 days. (**B**) SDS-PAGE gel of purified TraW (22.17 kDa). Lane 1, the protein in [200 mM NaCl, 20 mM Tris, pH 7.0], concentrated to 8 mg/mL for crystallization trials; lane 2, the same sample after 1 week incubation at 4°C.

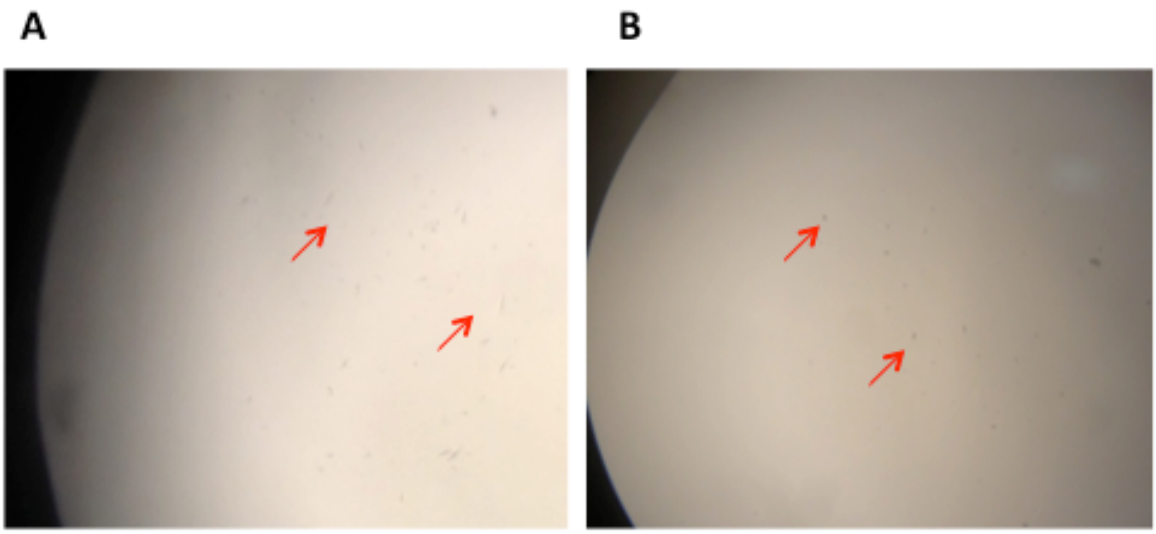


Figure 4.8. Evolving TraW crystals. (**A**) Microcrystals in areas indicated by the red arrows (**B**) Crystals of the same drop turned to precipitate over time as shown in areas indicated by the red arrows.

instability in the absence of intramolecular disulfide bonds (Wang *et al.*, 2004; Trivedi *et al*., 2009). The precipitation of the protein at 4°C was reversible, however not preferred since protein degradation still occurs. For this reason the protein was handled at 14 °C, a temperature that was determined to be less problematic at least over a short period of time. Tiny 15-50 μm TraW protein crystals formed in 2-3 days only at 4°C (Figure 4.7a). The protein sample used for crystallization was over 90% homogeneous, however when incubated for a 7 days at the same temperature as the crystallization plates, at 4°, the protein was found to degrade (Figure 4.7b). The same phenomenon was observed in the crystals as well (Figure 4.8). With continuous monitoring it was found that the protein crystals turned into an amorphous precipitate as shown in Figure 4.8b, indicating protein precipitation due to instability (McPherson, 1999).

While various additive agents were added in an attempt to increase the stability of the protein, OG detergent was found to have a positive effect. In the presence of 1% (w/v) detergent TraW was less autolytic. This effect of the detergent was also investigated using CD spectroscopy (Figure 4.9). The CD spectra do not show appreciable difference between the two protein solutions, however, there is a small reduction of the minima at 222 nm indicating a slight loss of α-helical structure of the protein (Wallace & Janes, 2009). Although the detergent appears to be an effective agent for stabilizing the protein, in its presence or without the same crystals formed and disappeared.

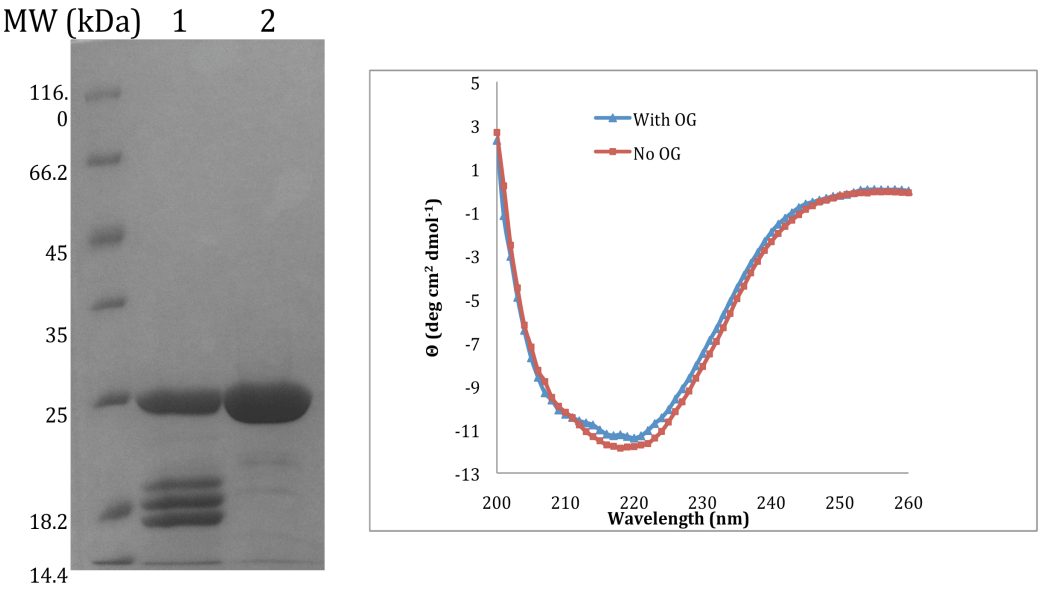


Figure 4.9. The effect of OG on TraW. **(A)** Protein purified in the absence (lane 1) and the presence of 1% (w/v) OG (lane 2). **(B)** CD spectroscopy scan showing overall similar protein folding, with the exception of the reduced minima at 222 nm.

The observations presented here indicate that while TraW is difficult to express and purify the protein is crystallizable. Further optimization of the purification and crystallization conditions could lead to TraW crystals suitable for structure and function determination using X-ray crystallography.

Chapter 5

Concluding Remarks and Future Work

5.1 LytR from TCSs

Two component systems play an important role in bacterial survival, adaptation, and virulence. These systems sense environmental changes via the HK and in response to a variety of stimuli regulate genes expression through RRs. *S. aureus* as an opportunistic pathogen, which is normally found in healthy human skin and mucous membranes, has become a threat primarily to the infected immunocompromised individuals (Zhu *et al*., 2010). This threat is magnified by the fact that the bacteria are becoming multi-drug resistant and as such they pose serious concerns in human health. The enhanced resistance to antibiotics and to the components of the innate host defence mechanisms is attributed by the ability of microbes to form thick multilayered biofilms (Donlan & Costerton, 2002).

S. aureus genome contains two genes *lytS* and *lytR* that have sequence similarities with sensor and response regulators, respectively, from bacteria TCSs (Brunskill & Bayles, 1996). The conserved structural arrangement and their ability to regulate *lrg*AB and *cidABC* genes that affect murein hydrolase activity, stationary-phase survival, antibiotic tolerance and biofilm formation (Sharma-Kuinkel *et al*., 2009), indicate that these genes have highly conserved functions and are represented by TCSs (Brunskill & Bayles, 1996). Two regulatory networks control expression of *cidABC* and *lrgAB*, one is

as a consequence of carbohydrate metabolism in response to acetic acid accumulation (Rice *et al*., 2005) and the other as a result of change in membrane potential (Patton *et al*., 2006). In addition TCS RRs have been found to activate upon phosphorylation by the HK, as well as phosphorylation by small-molecule phospho donors, such as acetyl phosphate (Lukat *et al*., 1992; McCleary, 1996). For these reasons it has been suggested that the LytSR may control *lrgAB* transcription by two mechanisms; one via sensing of the membrane potential change and the other by direct phosphorylation with acetyl phosphate in the presence of acetate accumulation (Sharma-Kuinkel, 2009).

In the absence of structural information of LytS and LytR from *S. aureus* TCS, we have determined the X-ray crystal structure of N-terminal domain of LytR. While this information does not provide us with sufficient clues as to how LytR regulates DNA binding and *lrgAB* and *cidABC* gene transcription, it confirms that the protein structure closely resembles other known RRs. In addition, the apo-$LytR^N$ form shows striking similarity to the BeF-$LytR^N$ complex (representing a phosphorylated form of the protein), with an overall RMSD value 0.616 Å. In the structure of $LytR^N$ the sulfate ion occupies the active site, where a phosphate would be located in the active form; the active site residues adopt conformations similar to other phosphorylated N-terminal domains of RRs. For this reason both BeF-$LytR^N$ and apo-$LytR^N$ have similar active-like structures as compared to other RRs. While the overall differences between the two crystal structures of $LytR^N$ are small the main difference between the two structures exists in the extension of helix α4 in the presence of BeF_3^- as induced by the interaction of Thr81 with a fluoride of the BeF_3^- complex. In addition the dimerization mode of the N-terminal

receiver domain resembles more closely to the same family of RR such as ComE from the same LytR subfamily (Boudes, *et al*., 2014), which involves the loop region between helix-α4 and sheet-β5, not seen in other family of RRs. Our findings suggest a conserved mode of dimerization amongst same family members.

Phosphorylation-induced dimerization is known to enhance DNA binding 500-fold (Koenig *et al*., 2004). A proposed model based on the structure of $AgrA^{C}$ (LytTR, C-terminal domain) bound to DNA (which consists of two direct repeats separated by 12 bp), suggests that only the dimeric N-terminal domain must mediate interaction of RR. While both monomers sit along the same face bound to the DNA the closest distance between the two protein monomers was found to be 10.1 Å, indicating that the C-terminal domain is not part of the dimerization interface (Sidote *et al*., 2008).

Future experiments would ideally include the structure of the full-length LytR which would be more beneficial into detailing the dimerization mode of the protein, and the effect on the C-terminal domain. Also, higher resolution of $LytR^{N}$ crystal structures could provide more detailed information about the solvent involved in the protein stabilization. Design of a new protein construct that removes residues 119-134, which could not be mapped due to the missing electron density and add disorder to the crystal packing, could lead to a higher resolution crystal structure. Lastly, crystallization of the protein without the sulfate ion would possibly result in a more representative apo form of the protein allowing for comparison with the phosphorylated LytR.

5.2 TrbC and TraW from T4SSs

The emergence of the antibiotic resistance genes and the evolutionary spread of such genes are due to the bacterial T4SSs. Conjugation, as the most important process for bacterial DNA transfer, requires independently replicating conjugative plasmids that encode and facilitate their own transfer as well as the transfer of other cellular DNA between two cells (Lanka & Wilkins, 1995; Frost *et al*., 2005). Genes necessary for conjugative single-stranded DNA transfer of the *Escherichia coli* K-12 fertility factor, F, are encoded within the 33.3kb transfer region of this 99.159 kb plasmid, assemble into a T4SS machinery. This system uses an appendage also known as the F pilus for DNA transfer into the recipient cells.

While homologues exist between different plasmid encoded T4SS, TrbC and TraW remain unique to the F plasmid of *E. coli*. Both proteins are localized in the periplasm and are both required for the F pilus assembly and DNA transfer proficiency (Anthony *et al*., 1999; Arutyunov *et al*., 2010). In the absence of structural and functional information on the proteins, and the low number of published articles on these proteins, we endeavoured to study these proteins in detail and determine their structures crystallographically. Both proteins are predicted to have thioredoxin-like properties according to their predicted structures. In addition, both TrbC and TraW are both cysteine-containing proteins, which are known to promote proper protein folding and assembly of pore complex proteins in the periplasm (Hemmis & Schildbach, 2013), much like the F pilus assembly.

Our crystallization attempts of both proteins was not entirely successful, possibly due to the unstable redox nature of the proteins. While TrbC exist as a mixture of oligomers in solution (which was confirmed by MS), homogenous protein samples were obtained only in the presence of GSH. However due to the low solubility of the protein, crystallization of the protein was not acheived. On the other hand, TraW is crystallizable, but the unstable nature of the protein either caused the crystals to dissolve or precipitate. This is possibly due to the lone thiol group of the protein from the single unpaired cysteine. For the future, in order to obtain thermally stable TraW the mutation of the lone cysteine residue could potentially help. On the other hand the addition of another cysteine residue could have the same positive effect, as seen on oxidized cross-linked lysozyme upon addition of double cysteine resides (Matsumura *et al*.,1989). This however, would be a trial and error method in order to get the right proximity of the two cysteine residues. Since TraW has been shown to interact with TrbB, whose function as a disulfide isomerase is also linked with maintaining of free thiols (Harris & Silverman, 2004; Hemmis & Schildbach, 2013), co-crystallization of the proteins could increase the stability of TraW.

Interaction of TrbC with TraW has been speculated but never experimentally shown. Using SEC and affinity chromatography, we show that the two proteins interact. In the absence of a crystallographically determined structure of the TraW-TrbC complex, the proteins were found to interact with both free or tagged N-terminal and C-terminal ends, our results indicate that these proteins interact *in vitro*. However, since the N-terminally 6-His tagged TraW is co-eluted from the SEC with the untagged TrbC, versus

the N-terminally 6-His tagged TrbC, we suggest that TraW prefers interaction with the N-terminal end of TrbC. We have attempted to analyze the kinetic properties of the interaction between the two proteins using Isothermal Titration Calorimetry, however due to the precipitation of the proteins in the cell we were not able to optimize parameters to characterize this interaction. In the future, it would be valuable to determine the redox properties of the proteins, this could lead to useful insight on dealing with sensitive proteins and help in their crystallization.

Chapter 6

References

Abdallah, A.M., Gey van Pittius, N.C., DiGiuseppe Champion, P.A., Cox, J., Luirink ,J., Vandenbrouke-Grauls, C.M.J.E., Appelmelk, B.J., Bitter, W. (2007). Type Vii secretion-mycobacteria show the way. *Nat. Rev. Microbiol.* **5,**883-891**.**,

Abrusci, P., McDowell, M.A., Lea, S.M., Johnson, S. (2014). Building a secreting nanomachine: a structural overview of T3SS. *Curr. Opin. Struct. Biol.* **25**, 111-117.

Alex, L.A., Simon, M.I. (1994). Protein histidine kinases and signal transduction in prokaryotes and eukaryotes. *Trends Genet.* **10**, 133–38.

Alvarez-Martinez, C. E., & Christie, P.J. (2009). Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev.* **73**, 775-808.

Anderson, M., Aly, K.A., Chen, Y. Missiakas, D. (2013). Secretion of atypical protein substrates by the ESAT-6 Secretion system of *Staphyococcus aureus*. *Mol. Microbiol*. **90**, 734-743.

Anthony, K.G., Klimke, W.A., Manchak, J., Frost, L.S. (1999). Comparison of proteins involved in pilus synthesis and mating pair stabilization from the related plasmids F and R100-1: insights into the mechanism of onjugation. *J. Bacteriol*. **181**, 5149-5159.

Arutyunov, D., Arenson, B., Manchak, J., Frost, L.S. (2010). F plasmid TraF and TraH are components of an outer membrane complex involved in conjugation. *J. Bacteriol*. **192**, 1730-1734.

Arutyunov, D., Frost, L.S. (2013). F conjugation: back to the beginning. *Plasmid*. **70**, 18-32.

Atkinson, M.R., Ninfa, A.J. (1992). Characterization of *Escherichia coli* glnL mutations affecting nitrogen regulation. *J. Bacteriol.* **174**, 4538–48.

Babić, A., Lindner, A.B., Vulić, M, Stewart, E.J., Radman, M. (2008). Direct visualization of horizontal gene transfer. *Science*. **319**, 1533-1536.

Bachhawat, P., Swapna, G.V., Montelione, G.T., Stock, A.M. (2005). Mechanism of activation for transcription factor PhoB suggested by different modes of dimerization in the inactive and active states. *Structure*. **13**, 1353-1363.

Bachhawat, P, and Stock, A.M. (2007). Crystal structures of the receiver domain of the response regulator PhoP from *Escherichia coli* in the absence and presence of the phosphoryl analog beryllofluoride. *J. Bacteriol.,* **189**, 5987-5995.

Ballister, E.R., Lai, A.H., Zuckermann, R.N., Cheng, Y, Mougous, J.D. (2008). In vitro self-assembly of tailorable nanotubes from a simple protein building block. *PNAS USA*. **105**, 3733-3738**.**

Baker-Austin, C., Wright, M.S., Stepanauskas, R., McArthur, J.V. (2006). Co-selection of antibiotic and metal resistance. *TRENDS Microbiol.* **14**, 176-182.

Barbieri, C.M., Mack, T.R., Robinson, V.L., Miller, M.T., and Stock, A.M. (2010). Regulation of response regulator autophosphorylation through interdomain contacts. *J. Biol. Chem.*, **285**, 32325-32335.

Baron, C., and Coombes, B. (2007). Targeting Bacterial Secretion Systems: Benefits of Disarmament in the Microcosm, *Infect. Disord. Drug Targets.* **7**, 19-27.

Beaber, J.W., Hochhut, B., Waldor, M.K. (2002). Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. *J. Bacteriol*. **184**, 4259-4269.

Bhaya, D., Takahashi, A., Grossman, A.R. (2001). Light regulation of type IV pilus-dependant motility by chemosensor-like elements in *Synechocystis* PCC6803. *PNAS USA*. **98**, 7540-7545.

Bingle, L.E., Bailey, C.M., Pallen, M.J. (2008). Type VI secretion: a beginner's guide. *Curr. Opin. Microbiol.* **11**, 3-8.

Birck, C., Mourey, L., Gouet, P., Fabry, B., Schumacher, J., Rousseau, P., Kahn, D., Samama, J.P. (1999). Conformational changes induced by phosphorylation of the FixJ receiver domain. *Structure*. **7**, 1505-1515.

Bitter, W., Koster, M., Latijnhouwers, M., de Cock, H.,Tommassen, J. (1998). Formation of oligomeric rings by XcpQ and PilQ, which are involved in protein transport across the outer membrane of *Pseudomonas aureginosa*. *Mol. Microbiol.* **27**, 209-219.

Böltner, D., MacMahon, C., Pembroke, J.T., Strike, P., Osborn, A.M. (2002). R391: a conjugative integrating mosaic comprised of phage, plasmid, and transposons elements. *J. Bacteriol*. **184**, 5158-5169.

Brunskill, E. W., and Bayles, K. W. (1996). Identifiaction of LytSR-regulated genes from *Staphylococcus aureus*. *J. Bacteriol.* **178**, 5810-5812.

Buckler, D. R., Zhou, Y., and Stock, A.M. (2002). Evidence in itradomain and intrdomain flexibility in an OmpR/PhoB homolog from *Thermotoga maritima. Strucutre* (Camb.), **10**, 153-164.

Burbulys, D., Trach, K.A., Hoch, J.A. (1991). Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay.*Cell.* **64**, 545–52.

Boucher, H., Miller, L. G. & Razonable, R. R. (2012). Serious Infections Caused by Methicillin-Resistant *Staphylococcus aureus*. *Clinical Infectious Diseases.* **57**, S183-S197.

Boudes, M., Sanchez, D., Graille, M., vanTilbeurgh, H., Durand, D., Quevillon-cheruel, S. (2014). Structural insight into the dimerization of the response regulator ComE from the *Streptococcus pneumoniae*. *Nuc. Aci. Res.* 1-12.

Campbell-Valois, F.-X. & Sansonetti, P.J. (2014). Tracking bacterial pathogens with genetically-encoded reporters. *FEBS letters*. In Press.

Cascales, E., and Christie, P.J. (2003). The versatile bacterial type IV secretion systems. *Nat. Rev. Microbiol.* **1**, 137-149.

Chabre, M. (1990). Aluminofluoride and beryllofluoride complexes: new phosphate analogs in enzymology. *Trends Biochem*. *Sci.*, **15**, 6-10.

Chalkley, R.J., Baker, P.R., Huang, L., Hansen, K.C., Allen, N.P., Rexach, M., Burlingame, A.L. (2005). Comprehensive analysis of a multidiomensional liquid chromatography mass psectrometry dataset acquired on a QqTOF mass spectrometer: 2. New developments in protein preospector allow for reiable and comprehensive automatic analysis of large datasets. *Mol. Cell Proteomics*. **4**, 1194-2004.

Chambers, H.F., and DeLeo, F. (2009). Waves of resistance: *Staphylococcus aureus* in the antibiotic era. *Nat. Rev. Microbiol.* **7**, 629-641.

Chapman, M.R., Robinson, L.S., Pinkner, J.S., Roth, R., Heuser, J., Hammar, M., Normark, S., Hultgren, S.J. (2002). Role of *Escherichia coli* curli operons in directing amyloid fiber formation. *Science*. **295**, 851-855.

Chen, I., Christie, P.J., Dubnau, D. (2005). The ins and outs of DNA transfer in bacteria. *Science*. **310**, 1456-1460.

Choo, H., Wang, W., Kim, R., Yokota, H., Damo, S., Kim, S., Wemmer, D., Kustu, S., Yan, D. (2001). BeF_3^- acts as a pohosphate analog in proteins phosphorylated on aspartae: structure of a BeF_3^- complex with phosphate phosphatase. *PNAS*, **98**, 8525-8530.

Christie, P.J. (2001). Type IV secretiom: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. *Mol. Microbiol*. **40**, 294-305.

Christie, P.J., Atmakuri, K., Krishnamoorthy, V., Jakubowski, S., Cascales, E. (2005). Biogenesis, architechture and function of bacterial type IV secretion systems. *Annu. Rev.. Microbiol.* **59**, 451-485.

Christie, P.J., Cascales, E. (2005). Structural and dynamic properties of bacterila type IV secretion system. *Mol. Mem. Biol*. **22**, 51-61.

Christie, P.J., Whitaker, N., Gonzalez-Rivera, C. (2014). Mechanism and structure of the bacterial type IV secretion systems. *Biochim. Biophys. Acta* **1843(8)**: 1578-91.

Cohen, S.X., Ben Jelloul, M., Long, F. *et al*. (2008). ARP/wARP and molecular replacement: the next generation. *Acta Crystallogr D Biol Crystallogr*. **64**, 49-60.

Collinson, S.K., Emödy, L., Müller, K.H., Trust, T.J., Kay, W.W. (1991). Purification and characterization of thin, aggregative fimbriae from *Salmonella enteritidis. J. Bacteriol.* **173**, 4773-4781.

CTV News, Canadian Press. (2011). http://www.ctv.ca/CTVNews/Health/ 20110707/clostridium-difficile-outbreak-help-110707/, Accessed July 7, 2011

Davies J. 1994. Inactivation of antibiotics and the dissemination of resistance genes, *Science.* **264**, 375-382.

Dawid, S., Roche, A.M., and Weiser, J.N. (2007). The *blp* bacteriocins of *Streptococcus pneuumoniae* mediate intraspecies competition both *in vitro* and *in vivo*. *Infect. Immun.*, **75**, 443-451.

de la Cruz, F., Frost, L.S., Meyer, R.J., Zechner, E.L. (2010). Conjugative DNA metabolism in gram-negative bacteria. *FEMS Microbiol. Rev*. **34**, 18-40.

d'Enfert, C., Ryter, A., Pugsley, A.P. (1987). Cloning and expression on *escherichia coli* of the *Klebsiella pneumoniae* genes for production, surface localization and secretion of the lipoprotein pullulanase. *The EMBO Journal*. **6**, 3531-3538.

Desvaux, M., Hebraud, M., Talon, R., Henderson, I.R. (2009). Secretion and subcellular localization of bacterial proteins: a semantic awereness issue. *Trends Microbiol.* **17**, 139-145.

Djordjevic, S., Goudreau, P.N., Xu, Q., Stock, A.M., West, A.H. (1998). Structural basis for methylesterase CheB regulation by a phosphorylation-activated domain. *PNAS USA*. **95**, 1381-1386.

Donlan, R.M., Costerton, J.W. (2002). Biofilms: survval mechanisms of clinically relevant microorganisms. *Clinical Microbiol. Rev.* **15**, 167-193.

Drury, L.S., and Buxton, R.S. (1985). DNA sequence analysis of the dye gene of *Escherichia coli* reveals amino acid homology between the dye and OmpR proteins. *J. Biol. Chem.* **260**, 4236-4242.

Duong, F., Lazdunski, A., Murgier, M. (1996). Protein secretion by heterologous bacterial ABC-transporters: the C-terminus secretion signal of the secreted protein confers high recognition specificity. *Mol. Microbiol*. **21**, 459-470.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Cryst.,* D**60**, 2126-2132.

Evans, P.R. (2011). An introduction to data reduction:space-group determination, scaling and intensity statistics. *Acta Cryst.*, D**67**, 282-292.

Fiedler, U., and Weiss, V. (1995). A common switch in activation of the response regulators NtrC and PhoB: phosphorylation induces dimerization of the receiver modules. *EMBO J.* **14**, 3696-705.

Fischer, W. (2011). Assembly and molecular mode of action of the *Helicobacter pylori* Cag type IV secretion apparatus. *FEBS J.* **278**, 1203-1212.

Fleming, A. (1929). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*, *British J. Exptl Pathol*. **10**, 226-236.

Frost, L.S., Manchak, J. (1998). F-phenocopies: characterizaton of expression of the F transfer region in stationary phase. *Microbiol*. **144**, 2579-2587.

Frost, L.S., Ippen-Ihler, K., Skurray, R.A. (1994). Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev.* **58**, 162-210.

Frost, L. S., Leplae, R., Summers, A.O., Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* **3,** 722-732.

Frost, L.S., Koraimann, G. (2010). Regulation of bacterial conjugation: balancing opportunity with adversity. *Fut. Microbiol*. **5**, 1057-1071.

Galán, J.E. (1999). Type III secretion machines: bacterial devices for protein delivery into host cells. *Science*. **284**, 1322-1328.

Galperin, M. Y. (2006). Structural classification of bacterial response regulators: diversity of output domains and domain combinations. *J, Bacteriol*., **12**, 4169-4182.

Galperin, M. Y. (2008). Telliing bacteria: do not LytTR. *Structure,* **16**, 657-659.

Galperin, M. Y. (2010). Diversity of structure and function of response regulator output domains. *Curr. Opin. Microbiol*., **13,** 150-159.

Gao, R., and Stock, A. M. (2010). Molecular strategies for phosphorylation-mediated reglation of response regulator activity. *Curr. Opin. Microbiol.,* **13**, 160-167.

Gao, R., Mack, T. R., and Stock, A. M. (2007). Bacterial response regularos: versatile regulatory strategis from common domains. *Trends Biochem. Sci.,* **32**, 225-234.

Gardino, A.K., and Kern, D. (2007). Functinal dynamics of response regulators using NMR relaxation techniques. *Methods Enzymol*., **423**, 149-165.

Gawarzewski, I., Smits, S.H.J., Schmitt, L., Jose, J. (2013). Structural comparison of the transport units of type V secretion systems. *Biol. Chem*. **394**, 1385-1398.

Gentschev, I., Dietrichm G., Goebel, W. (2002). The *E. coli* alpha-hemolysin secretion system and its use in vaccine development. *Trends Microbiol*. **10**, 39-45.

Gophna, U. (2003). Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events. *Gene*. **312**, 151-163.

Gouet, P., Fabry, B., Guillet, V., Birck, C., Mourey, L., Kahn, D., Samama, J.P. (1999). Structural transitions on the FixJ receiver domain. *Strucure*, **7**, 1517-1526.

Goyal, P., Van Gerven, N., Jonckheere, W., Remaut, H. (2013). Crystallization and preliminary X-ray crystallographic analysis of the curli transporter CsgG. *Acta Cryst.* **F69**. 1349-1353.

Greub, G., Collyn, F., Guy, L., Roten, C.-A. (2004). A genomic island present along the bacterila chromosome of the *parachlamydiaceae* UWE25, an obligate amoebal endosymbiont, encodes a potentially functional F-like conjugative DNA transfer system. *BMC Microbiol*. **4**, 48.

Groicher, K. H., Firek, B. A., Fujimoto, D. F., and Bayles, K. W. (2000). The *Staphylococcus aureus lrgAB* operon modulates murein hydrlase activity and penicillin tolerance. *J. Bacteriol.* **182**, 1794-1801.

Gubbins, M.J., Arthur, D.C., Ghetu, A.F., Glover, J.N.M., Frost, L.S. (2003). Characterizing the structural features of RNA/RNA interactions of F-pasmid FinOP fertility inhibition system. *J. Biol. Chem*. **278**, 27663-27671.

Han, J.C., & Han, G.Y. (1994). A procedure for quantitative determination of tris(2-carboxyethyl)phosphine, an odorless reducing agent more stable and effective than dithiothreitol. *Anal. Biochem.* **220**, 5-10.

Harris, R.L., Hombs, V., Silverman, P. (2001). Evidence that F-plasmid proteins TraV, TraK and TraB assemble into an envelope-spanning structure in *Escherichia coli*. *Mol. Microbiol*. **42**, 757-766.

Harris, R.L., Silverman, P.M. (2004). Tra proteins characteristic of F-like type IV secretion systems constitute an interaction group by yeast two-hybrid analysis. *J. Bacteriol*. **186**, 5480-5485.

Hazes, B., and Frost, L. (2008). Towards a systems biology approach to study type II/IV secretion systems. *Biochim. Biophys. Acta.* **1778,** 1389-50.

Hemmis, C.W., & Schildbach, J.F. (2013). Thioredoxin-like proteins in F and other plasmid systems. *Plasmid*. **70**, 168-189.

Henderson, I.R., Navarro-Garcia, F., Desvaux, M., Fernandez, R.C. Aldeen, D.A. (2004). Type V protein secretion pathway: the autotransporter story. *Society*. **68**, 692-744.

Holland, I.B., Schmitt, L., Young, J. (2005). Type I protein secretion in bacteria, the ABC-trasnporter dependent pathway. *Mol. Membr. Biol.* **22**, 29-39.

Hueck, C.J. (1998). Type III protein secretion system in bacterial pathogesn of animal and plants – cellular and molecular impact of type III secretion in bacteria. *Microbiol*. **62**, 379-433.

Ippen-Ihler, K.A., Minkley, E.G. (1986). The conjugation system of F, the fertility factor of *Escherichia coli*. *Annu. Rev. Genet*. **20**, 593-624.

Ippolito, G., Leone, S., Lauria, F.N., Nicastri, E., and Wenzel, R.P. (2010). Methicillin-resistant *Staphylococcus aureus*: the superbug. *Int J Infect Dis.,* **14**, S7-11.

Jensen, S.O., Lyon, B.R. (2009) Genetics of antimicrobial resistance in *Staphylococcus aureus*. *Future Microbiol.* **4**, 565–582.

Jiang, M., Bourret, R.B., Simon, M.I., Volz, K. (1997). Uncoupled phosphorylation and activation in bacterial chemotaxis. The 2.3 A structure of an Aspartate to lysine mutant at position 13 of CheY. *J. Biol. Chem.* **272**, 11850-5.

Johnson, T.L., Abendroth, J., Hol, W.G.J., Sandkvist, M. (2006). Type II secretion: from structure to function. *FEMS Microbiol. Lett.* **255**, 175-86.

Kan, C.-C., Hambly, K. and Debe, D.A. (2005). *Molecular Analysis and Genome Discovery*, Rapley, R. & Harbron, S. eds., Chapter 14 "Structure-Based Drug Discovery," (John Wiley & Sons, Ltd, Chichester, UK.), pp. 295-322.

Keener, J., Kustu, S. (1988). Protein kinase and phosphoprotein phosphatases activities of nitrogen protein NTRB and NTRC of enteric bacteria: roles of the conserved amino terminal domain of NTRC. *Proc. Natl. Acad. Sci. USA* **85**, 4976–80.

Kelley, L.A., Sternberg, M.J.E. (2009). Protein structure prediction on the web: a case study using Phyre server. *Nature Protocols*. **4**, 363-371.

Kapitein, N. & Mogk, A. (2013). Deadly syringes: type VI secretion system activities in pathogenicity and interbacterial competition. *Curr. Opin. Microbiol.* **16**, 52-58.

Knight, G.M., Budd, E.L., Whitney, L., Thornley, A., Al-Ghusein, T., H., Planche, T., Lindsay, J.A. (2012). Shift in dominant hospital-associated methicillin-resistant

Staphylococcus aureus (HA-MRSA) clones over time. *J. Antimicrob. Chemother.* **67**, 2514-2522.

Koenig, R.L., Ray, J.L., Maleki, S.J., Smettzer, M.S., and Hurlburt, B.K. (2004). *Staphylococcus aureus* AgrA binding to the RNAIII-*agr* regulatory region. *J. Bacteriol*. **186**, 7549-7555.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774-797.

Kuhn, P., Wilson, K., Pach, M.G., and Stevens, R.C. (2002). The genesis of high-throughput structure-based drug discovery using protein crystallography, *Curr. Opin. Chem. Biol,*. **6**, pp. 704-710.

Lanka, E., Wilkins, B.M. (1995). DNA processing reaction in bactetial conjugation. *Anuu. Rev. Biochem*. **64**, 141-169.

Laskowski, R.A., McArthur, M.W., Moss, D.S., Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**, 282-291.

Lawley, T.D., Klimke, W.A., Gubbins, M.J., and Frost, L.S. (2003). F factor is a true type IV secretion system. *FEMS Microbiol. Lett.* **224**, 1-15.

Lederberg, J. and Tatum, E.L. (1946). Gene recombination in *Escherichia coli*. *Nature*. **158**, 558.

Lee, S.Y., Chi, H.S., Pelton, J.G., Yan, D., Berry, E.A., Wemmer, D.E. (2001). Crystal structure of activated CheY. Comparison with other activated receiver domains. *J. Biol. Chem.* **276**, 16425-16431.

Leonard, P.G., Bezar, I.F., Sidote, D.J., Stock, A.M. (2012). Identification of a hydrophobic cleft in the LytTR domain of AgrA as a locus for small molecule interactions that inhibit DNA binding. *Biochemistry*, **51**, 10035-10043.

Leonard, P.G., Golemi-Kotra, D., Stock, A.M. (2013). Phosphorylation-dependent conformational changes and domain rearrangements in *Staphylococcus aureus* VraR activation. *PNAS*. **110**, 8525-8530.

Lesley, S.A., Wilson, I.A. (2005). Protein production and crystallization at the joint center for structural genomics. *J. Struct. Funct. Genom.* **6**, 71-79.

Leslie, A.G.W, & Powell, H.R. (2007). *Evolving methods for macromolecular crystallography* (Read, R.J. & Sussman, J.L., ed), pp 41-51, Springer, Dordrecht.

Li, M., Lai, Y., Villaruz, A. E., Cha, D. J., Sturdevant, D. E., and Otto, M. (2007). Gram-positive three-component antimicrobial peptide-sensing system, *PNAS,* **104**, 9469-9474.

Li, W., Ancona, V., Zhao, Y. (2014). Co-regulation of polysaccharide production, motility, and expression of type III secretion genes by EnvZ/OmpR and GrrS/GrrA systems in *Erwinia amylovora*. *Mol. Genet. Genom.* **289**, 63-75.

Lindsay, J.A., 2010. Genomic variation and evolution of *Staphylococcus aureus*. *Int. J.Med. Microbiol*. **300**, 98–103.

Lindsay, J.A. (2014). *Staphylocossu aureus* genomics and the impact of horizontal gene transfer. *Int. J. Med. Microbiol.* **304**, 103-109.

Lizewski, S.E., Lundberg, D.S., Schurr, M.J. (2002). The trasmcriptional regulator AlgR is essential for *Pseudomonas aureginosa* pathogenesis. *Infect. Immun.*, **70**, 6083-6093.

Lu, J., Manchak, J., Klimke, W., Davidson, C., Nirth, N., Skurray, R.A., Frost, L.S. (2002). Analysis and characterization of the IncFV plasmid pED208 transfer region. *Plasmid*. **48**, 24-37.

Lu, J., Wong, J.J.W., Edwards, R.A., Manchak, J., Frost, L.S., Glover, J.N.M. (2008). Structural basis of specific TraD-TraM recognition during F plasmid-mediated bacterial conjugation. *Mol. Microbiol*. **70**, 89-99.

Lukat, G.S., McCleary ,W.R., Stock, A.M., Stock, J.B. (1992). Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors. *PNAS USA*. **89**, 718-22.

MacFarlane, S.A., Merrick, M. (1985). The nucleotide sequence of the nitrogen regulation gene ntrB and the glnA-ntrBC intergenic region of *Klebsiella pneumoniae*. *Nucleic Acids Res.* **13**, 7591–606.

Maiorov, V.N., & Crippen, G.M. (1994). Significance of root-mean-square deviation in comparing three-dimsnional structures of globular proteins. *J. Mol. Biol.* **235**, 625-634.

Maneewannakul, S., Kathir, P., Moore, D., Le, L.A., Wu, J.H., Ippen-Ihler, K. (1987). Location of F plasmid transfer opern genes traC and traW and identification of the traW product. *J. Bacteriol*. **169**, 5119-5124.

Maneewannakul, S., Maneewannakul, K, Ippen-Ihler, K. (1991). Characterization of trbC, a new F plasmid tra operon gene that is essential to conjugative transfer. *J. Bacteriol*. **173**, 3872-3878.

Maneewannakul, S., Maneewannakul, K., Ippen-Ihler, K. (1992). Characterization,

localization, and sequence of F transfer region products: the pilus assembly gene product TraW and a new product, TrbI. *J. Bacteriol*. **174**, 5567-5574.

Makino, K., Shinagawa, H., Amemura, M., Nakata, A. (1989). Signal transduction in the phosphate regulon of *Escherichia coli* involves phosphotransfer between PhoR and PhoB protein. *J. Mol. Biol.*, **210**, 551-559.

Malachowa, N., and DeLeo, F.R. (2010). Mobile genetic elements of *Staphylococcus aureus. Cell. Mol. Life. Sci.* **67**, 3057-3071.

Mangayarkarasi, N., and Francetic, O. (2014). Type II secretion system: A magic beanstalk or a protein escalator. *Biochim. Biophys. Acta.* In Press.

Martin, J.L. (1995). Thioredoxin- a fold for all reasons. *Structure*. **3**, 245-250.

Matsumura, M., Becktel, W.J., Levitt, M., Matthews, B.M. (1989). Stabilization of phage T4 lysozyme by engineered disulfide bonds. *PNAS USA*. **86**, 6562-6566.

McCleary, W.R. (1996). The activation of PhoB by acetylphosphate. *Mol. Microbiol.* **20**, 1155-63.

McPherson, A. *Crystallization of biological macromolecules.* New York: Cold Spring Harbor Laboratory Press. 1999.

Michiels, T., Wattiau, P., Brasseur, R., Ruysschaert, J-M., Cornelis, G. (1990). Secretion of Yop proteins by *Yersiniae*. *Infect. And Immun.* **58**, 2840-2849.

Milani, M., Leoni, L., Rampioni, G., Zennaro, E., Ascenzi, P., Bolognesi, M. (2005). An active-like structure in the unphosphorylated StyR response regulator suggests a phosphorylation-dependent allosteric activation mechanism. *Structure*, **13**, 1289-1297.

Moore, D., Maneewannakul, K., Maneewannakul, S., Wu, J.H., Ippen-Ihler, K., radley, D.E. (1990). Characterization of the F-plasmid conjugative transfer gene traU. *J. Bacteriol*. **172**, 4263-4270.

Morens, D.M., Folkers, G.K., and Fauci, A.S. (2004). The challenge of emerging and re-emerging infectious diseases. *Nature*, **430**, 242-249.

Murata, T., Ohnishi, M., Ara, T., Kaneko, J., Han, C-G., Li, Y.F., Takashima, K., Nojima, H., Nakayama, K., Kaji, A., Kamio, Y., Miki, T., Mori, H., Ohtsubo, E., Terawaki, Y., Hayashi, T. (2002). Complete nucleotide sequence of plasmid Rts1: implications for evolution of large plasmid genomes. *J. Bacteriol.* **184**, 3194-3202.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst.*, D**67**, 355-367.

Nathan, C. (2004). Antibiotics at the crossroads, *Nature,* **431**(7011), pp. 899-902.

Nikolskaya, A. N., and Galperin, M. Y. (2002). A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. *Nucleic Acids Res*., **30**, 2453-2459.

Noriega C.E., Lin, H.Y., Chen, L.L., Williams, S.B., Stewart, V. (2010). Asymmetric crossregulation between the nitrate-responsive NarX-NarL and NarQ-NarP two-component regulatory systems from *Escherichia coli* K-12. *Mol. Microbiol*., **75**, 394-412.

Novick, R.P. (2003). Autoinduction and signal transduction in the regulation of staphylococcal virulence. *Mol. Micorbiol.*, **48**, 1429-1449.

Ochman, H., Lawrence, J.G. & Groisman, E.A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature*, **405**, 299-304.

Olsen, J.E., Christensen, H., Aarestrup, F.M. (2006) Diversity and evolution of blaZ from Staphylococcus aureus and coagulasenegative staphylococci. J. Antimicrob. Chemother. **57**, 450–460

Oomen,, C.J., van Ulsen, P., van Gelder, P., Feijen, M., Tommassen, J., Gros, P. (2004). Structure of the translocator domain of a bacterial autotransfer. *The EMBO J.* **23**, 1257-1266.

Padilla, J. E. & Yeates, T.O. (2003). A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utilitu for detecting twinning. *Acta Cryst.,* D**59**, 1124-1130.

Pan, S.Q., Charles, T., Jin, S., Wu, Z-L., Nester, E.W. (1993). Preformed dimeric state of the sensor proteins VirA is involved in plant-Agrobacterium signal transduction. *Proc. Natl. Acad. Sci. USA* **90**, 9939–43.

Park, A.K., Moon, J.H., Lee, K.S., Chi, Y.M. (2012). Crystal structure of receiver domain of putative NarL family response regultor spr1814 from *Streptococcus pneumoniae* in the absence and presence of the phosphoryl analog beryllofluoride. *Biochem. Biophysical. Res. Comm.*, **421**, 403-407.

Parkinson, J.S., Kofoid, E.C. (1992). Communication modules in bacteria signaling proteins. *Annu. Rev. Genet.* **26**, 71–112.

Parkinson, J.S. (1993) Signal transduction schemes of bacteria. Cell. **73**, 857-871.

Patel, K. (2014) Biochemical characterization of the walker and the lytSR two component system from *Staphylococcus aureus*. MSc Thesis, Department of Chemistry, York University, Toronto, Canada.

Patton, T. G., Yang, S. J., and Bayles, K. W. (2006). The role of proton motive force in expression of the *Staphylococcus aureus cid* and *lrg* operons. *Mol Microbiol.,* **59**, 1395-1404.

PDB ID: 2QV0. Bonanno, J.B., Freeman, J., Bain, K.T., Iizuka, I., Ozyurt, S., Smith, D., Wasserman, S., Sauder, J.M., Burley, S.K., Almo, S.C., New York SGX Research Center for Structural Genomics (NYSGXRC). Crystal structure of the response regulatory domain of protein mrkE from *Klebsiella pneumoniae.*

Peschel, A., Jack, R. W., Otto, M., Collins, L. V., Staubitz, P., Nicholson, G., Kalbacher, H., Nieuwenhuizen, W. F., Jung, G., Tarkowski, A., van Kessel, K. P., and van Strijp, J. A. (2001). *Staphylococcus aureus* resistance to human defensins and evasion of neutrophil killing via the novel virulence factor Mprf in based on modification of membrane lipids with L-Lysine. *J Exp Med.,* **193**, 1067-1076.

Peschel, A. (2002). How do bacteria resist human antimicrobial peptides?. *Trends in Microbiol.,* **10**, 179-186.

PHAC-Public Health Agency of Canada (2013). The chief public health officer's report on the state of public health in Canada: Infectious disease-the never-ending threat. Retrieved on May 26, 2014, from http://www.phac-aspc.gc.ca/cphorsphc-respcacsp/2013/infections-eng.php.

Posas, F., Saito, H. (1998). Activation of the yeast SSK2 MAP kinase kinase kinase by the SSK1 two-component response regulator. *EMBO J.* **17**, 1385-94.

Pukatzki, S., Ma, A.T., Sturtevant, D., Krastins, B., Sarracino, D., Nelson, W.C., Heidelberg, J.F., Mekalanos, J.J. (2006). Identification of a conserved bacterial protein secretion system in *vibrio cholerae* using the dictyostelium host model system. *PNAS USA*. **103**, 1528-1533

Pukatzki, S., Ma, A.T., Revel, A.T., Sturtevant, D., Mekalanos, J.J. (2007). Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *PNAS USA*. **104**, 15508-15513.

Pukatzki, S., McAuley, S.B., Miyata, S.T. (2009). The type Vi secretion system: translocation of effectors and effector-domains. *Curr. Opin. Microbiol.* **12**, 11-17.

Québatte, M., Dick, M.S., Kaever, V., Schmidt, A., Dehio, C. (2013). Dual input control: activation of the *Bartonella henselae* VirB/D4 type IV secretion system by the stringent sigma factor RpoH1 and the BatR/BatS two-component system. *Mol. Microbiol*. **90**, 756-775.

Ramsey, M.E., Hackett, K.T., Kotha, C., Dillard, J.P. (2012). New complementation constructs for inducible and constitutive gene expression in *Neisseria gonorrhoeae* and *Neisseria meningitides. App. Environ. Microbiol*. **78**, 3068-3078.

Renshaw, P.S., Lightbody, K.L., Veverka, V., Muskett, F.W., Kelly, G., Frenkiel, T.A., Gordon, S.V., Hewinson, R.G., Burke, B., Norma, J., Williamson, R.A., Carr, M.D. (2005). Structure and function of the complex formed by the turbeculosis factors CFP-10 and ESAT-6. *EMBO J.* **24**, 2491-2498.

Richardson, J.S., and Richardson, D.C. (1990). Principles and patterns of protein conformation. In *prediction of protein structure and principles of protein conformation* (Fasman, G.D., ed.), pp 1-98, Plenum Press, New York.

Rice, K. C., Firek, B. A., Nelson, J. B., Yang, S. J., Patton, T. G., and Bayles, K. W. (2003). The *Staphylococcus aureus cidAB* operon: evaluation of its role in regulation of murein hydrolase activity and penicillin tolerance. *J. Bacteriol.* **185**, 2635-2643.

Rice, K.C., Nelson, J.B., Patton, T.G., Yang, S.J., Bayles, K.W. (2005). Acetic acid induces expression of the *Staphylococcus aureus cidABC* and *lgAB* murein hydrolase reglator operons. *J. Bacteriol*. **187**, 813-821.

Robinson, V.L., Buckler, D.R., Stock, A.M. (2000). A tale of two components: a novel kinase and a regulatory switch. *Nat. Ctruct. Biol.* **7**, 626-633.

Romine, M.F., Stillwell, L.C., Wong, K.-K., Thurston, S.J., Sisk, E.C., Sensen, C., Gaasterland, T., Fredrickson, J.K., Saffer, J.D. (1999). Complete sequence of a 184-kilobase catabolic plasmid from *Sphingomonas aromaticivorans* F199. *J. Bacteriol*. **181**, 1585-1602.

Rood, J.I. (1998). Virulence genes of *Clostridium perfingens*. *Annu. Rev. Microbiol.*, **52**, 333-360.

Russell, A.B., Singh, P., Brittnacher, M., Bui, N.K., Hood, R.D., Carl, M.A., Agnello, D.M., Schwarz, S., Goodlett, D.R., Vollmer, W. *et al*. (2012). A widespread bacterail type VI. Secretion effector superfamily identified using a heuristic approach. *Cell Host Microbe*. **11**, 538-549.

Sanders, D.A., Gillece-Castro, B.L., Stock, A.M., Burlingame, A.L., Koshland, D.E. jr. (1989). Identification of the site of phosphorylation of the chemotaxis response regulator protein, CheY. *J. Biol. Chem.* **264**, 21770-8.

Schröder, G., Lanka, E. (2005). The mating pair formation system of conjugative plasmids – a versatile secretion machinery for transfer of proteins and DNA. *Plasmid*. **54**, 1-25.

Schrodinger, L.L.C. (2010). The PyMol molecular graphics system, Version 1.3r1.

SERC (UK) collaborative Computational Project. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystal*., D**50**, 760-763.

Shala, A., Patel, K.H., Golemi-Kotra, D, Audette, G.F. (2013). Expression, purification, crystallization and preliminary X-ray analysis of the receiver domain of *Staphylococcus aureus* LytR protein. *Acta Cryst.*, F**69**, 1418-1421.

Shala, A., Ferraro, M., Audette, G. F. "Bacterial Secretion Systems: Nanomachines for Infection and Genetic Diversity", in R. Bawa, G.F. Audette & I. Rubenstein Eds. *Clinical Nanomedicine: From Bench to Bedside*, Pan Sanford Publishing, Singapore, pp 1-22 (2015). In Press.

Sharma-Kuinkel, B. K., Mann, E. E., Ahn, J. S., Kuechenmeister, L. J., Dunman, P. M., and Bayles, K. W. (2009). The *Staphylococcus aureus* LytSR two-component regulatory system affects biofilm formation. *J. Bacteriol.,* **191**, 4767-4775.

Sherburne, C.K., Lawley, T.D., Gilmour, M.W., Blattner, F.R., Burland, V., Grotbeck, E., Rose, D.J., Taylor, D.E. (2000). The complete DNA sequence and analysis of R27, a large incHI plasmid form *Salmonella typhi* that is temperature sensitive for transfer. *Nucleic Acids Res.* **28**, 2177-2187.

Shrivastava, R., and Miller, J.F. (2009). Virulence factor secretion and translocation by *Bordetella* species. *Curr. Opin. Microbiol*. **12**, 88-93.

Sidote, D. J., Barbieri, C. M., Wu, T., and Stock, A. M. (2008). Structure of the *Staphylococcus aureus* AgrA LytTR domain bound to DNA reveals a beta fold with an unusual mode of binding. *Structure*, **16**, 727-735.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Sörding, J., Thompson, J.D.,Higging, D.G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.

Simms, S.A., Keane, M.G., Stock, J. (1985). Multiple forms of the CheB methylesterase in bacterial chemosensing. *J. Biol. Chem.* **26**, 10161-68.

Solà, M., Gomis-Rüth, F.X., Serrano, L., Gonzáles, A., & Coll, M. (1999). Three-dimensional crystal structure of the transcription factor PhoB receiver domain. *J. Mol. Biol.,* **285**, 675-687.

Solà, M., Drew, D.L., Blanco, A.G., Gomis-Rüth, F.X., Coll, M. (2006). The cofactor-induced pre-active conforamtion on PhoB. *Acta Cryst.*, D**62**, 1046-1057.

Stock, A., Chen, T.,Welsh, D., Stock, J. (1988). CheA protein, a central regulator of bacterial chemotaxis, belongs to a family of proteins that control gene expression in response to changing environmental conditions. *Proc. Natl. Acad. Sci. USA* **85**, 1403–7.

Stock, J.B., and Da Re, S. (2000). Signal transduction: response regulators on and off. *Curr. Biol.*, **10**, R420-424.

Stock, J.B., Ninfa, A.J., Stock, A.M. (1989). Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiol. Rev.* **53**, 450–90.

Stock, J.B., Stock, A.M., Mottonen, J.M. (1990). Signal transduction in bacteria. *Nature*. **344**, 395-400.

Stock, J. (1999). Signal transduction: gyrating protein kinases. *Curr. Biol.* **9**, R364–67.

Stock, A.M., Robinson, V.L., Goudreau, P.N. (2000) Two-component signal transduction. *Annu Rev Biochem.* **69**; 183-215.

Storoni, L.C., McCoy, A.J., Read, R.J. (2004). Likelihood-enhanced fast rotation functions. *Acta Cryst.*, D**60**, 432-438.

Sun, G.W., and Gan, Y-H. (2010). Unraveling type III secretion systems in the highly versatile *Burkholderia pseudomallei*. *Trends in Micro*. **18**, 561-568.

Surette, M.G., Levit, M., Liu, Y., Lukat, G., Ninfa, E.G., Ninfa, A., Stock, J.B. (1996). Dimerization is required for the activity of the protein histidine kinase CheA that mediates signal transduction in bacterial chemotaxis. *J. Biol. Chem.* **271**, 939–945.

Sysoeva, T.A., Zepeda-Rivera, M.A., Huppert, L.A., Burton, B.M. (2014). Dimer recognition and secretion by the ESX secretion system in *Bacillus subtilis*. *PNAS*. **111**, 7653-7658.

Toro-Roman, A., Wu, T., and Stock, A.M. (2005). A common dimerization interfacein bacterial response regulators KdpE and TorR. *Protein Science*, **14,** 3077-3088.

Trivedi, M.V., Laurence, J.S., Siahaan, T.J. (2009). The role of thiols and disulfides in protein chemical and physical stability. *Curr Protein Pept. Sci*. **10**, 614-625.

Trokter, M., Felisberto-Rodrigues, C., Christie, P.J., Waksman, G. (2014). Recent advances in the structural and molecular biology of type IV secretion systems. *Curr Opin Struc. Biol*. **27**, 16-23.

Vaguine, A.A., Richelle, J., Wodak, S.J. (1999). SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Cryst.*, D**55**, 191-205.

Valdez, Y., Ferreira, R.B., Finlay, B.B. (2009). Molecular mechanism of *Salmonella* virulence and host resistance. *Curr. Top. Microbiol. Immunol*. **337**, 93-127.

vanWely, K.H., Swaving, J., Freundl, R., Driessen, A.J. (2001). Translocation of proteins across the cell envelope of gram-positive bacteria. *FEMS Microbiol. Rev*. **25**, 437-54.

Volkman, B.F., Lipson, D., Wemmer, D.E., Kern, D. (2001). Two-state allosteric behavior in a single-domain signaling protein. *Science*, **291**, 2429-2433.

Volz, K., and Matsamura, P. (1991). Crystal structure of *Escheiricha coli* CheY refined at 1.7-Å resolution. *J. Biol. Chem*. **266**, 15511-15519.

Voyich, J. M., Braughton, K. R., Sturdevant, D. E., Whitney, A. R., Said-Salim, B., Porcella, S. F.,Long, R. D., Dorward, D. W., Gardner, D. J., Kreiswirth, B. N., Musser, J. M., and DeLeo, F.R. (2005). Insights into mechanism used by *Staphylococcus aureus* to avoid destruction by human neutrophils. *J. Immunol.* **175**, 3907-3919.

Wallace, B.N., & Janes, R.W. *Modern techniques for circular Dichroism and synchrotron radiation circular Dichroism spectroscopy*. Advances in biomedical spectroscopy, Vol1. Washington, DC: IOS Press. 2009. 9-18.

Walsh C. (2003). Where will new antibiotics come from? *Nat. Rev. Microbiol,*. **1**, pp. 65-70.

Wang, X-Y., Meng, F-G., Zhou, H-M. (2004). The role of disulfide bonds in the conformational stability and catalytic activity of phytase. *Biochem. Cell Biol.* **82**, 329-334.

Weichsel, A., Gasdaska, J.R., Powis, G., Montfort, W.R. (1996). Crystal structures of reduced, oxidized, and mutated human thioredoxins: evidence for a regulatory homodimer. *Structure*. **4**, 735-751.

Weigel L.M., Clewell, D.B., Gill, S.R., Clark, N.C., McDougal, L.K., Flannagan, S.E., Kolonay, J.F., Shetty, J., Killgore, G.E., Tenover, F.C. (2003) Genetic analysis of a high-level vancomycin resistant isolate of Staphylococcus aureus. Science. **302**, 1569–1571.

Weiss, V., Claverie-Martin, F., Magasanik, B. (1992). Phosphorylation of nitrogen regulator I on *Escherichia coli* induced stron cooperative binding to DNA essential for activation of transcription. *PNAS USA*. **89**, 5088-92.

Welch, M., Oosawa, K., Aizawa, S., Eisencbach, M. (1993). Phosphorylation-dependent binding of a signal molecule to the flagellar switch of bacteria. *PNAS USA*. **90**, 8787-91.

West, A.H., and Stock, A.M. (2001). Histidine kinase and response regulator proteins in two-component signaling systems. *Trends Biochem. Sci.*, **26**, 369-376.

Wilkins, B.M., and Frost, L.S. (2001). *Molecular Medical Microbiology,* eds. Sussman, M., Chapter 17 "Mechanisms of gene exchange between bacteria," (Academic Press, London) pp. 355-400.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Reas, R.J., Vagin, A.,

Wilson, K.S. (2011). Overview of the CCP4 suite and current developments. *Acta Cryst.,* D**67**, 235-242.

World Health Organization, WHO. (2011) http://www.euro.who.int/en/what-we-do/health-topics/emergencies/international-health-regulations/news/news/2011/07/outbreaks-of-e.-coli-o104h4-infection-update-29, Accessed July 7, 2011.

Wolfe, A.J., Stewart, R.C. (1993). The short form of the CheA protein restores kinase activity and chemotactit ability to kinase-deficient mutants. *Proc. Natl. Acad. Sci. USA* **90**, 1518–22.

Yagi, T., Kurokawa, H., Senda, K., Ichiyama, S., Ito, H., Ohsuka, S., Shibayama, K., Shimokata, K., Kato, N., Ohta, M., Arakawa, Y. (1997). Nosocomial spread of cephem-resistant *Escherichia coli* carrying multiple toho-1-like b-lactamase genes. *Antimicrob. Agen. Chem*., **41**, 2606-2611.

Yan, D., Cho, H.S., Hastings, C.A., Igo, M.M., Lee, S.-Y., Pelton, J.G., Stewart, V., Wemmer, D.E., Kustu, S. (1999). Beryllofluoride mimcs phosphorylation of NtrC and other bacterial response regulators. *PNAS*, **96**, 14789-14794.

Yang, Y., Park, H., Inouye, M. (1993). Ligand binding induces an asymmetrical transmembrane signal through a receptor dimer. *J. Mol. Biol.* **232**, 493–98.

Zhang, R., Scott, D.L., Westbrook, M.L., Nance, S., Spangler, B.D., Shipley, G.G., Westbook, E.M. (1995). The three-dimensional crystal structure of cholera toxin. *JMB*. **251**, 563-573.

Zheng, J., Leung, K.Y. (2007). Dissection of a type VI secretion system in *Edwardsiella tarda*. *Mol. Microbiol.* **66**, 1192-1206.

Zhu, T., Lou, Q., Wu, Y., Hu, J., Yu, F., Qu, D. (2010). Impact of the *Staphylococcus epidermidis* LytSR two-component regulatory system on murein hydrolase activity, pyruvate utilization and global transcriptional profile. *BMC Microbiol.* **10**, 1-16.

Zhu, X., Rebello, J., Matsumura, P., Volz, K. (1997). Crystal structure of CheY mutants Y106W and T87I/Y106W. ChY activation correlates with movement of residue 106. *J. Biol. Chem.* **272**, 5000-6.

Zoutman, D. E., Ford, B. D., Bryce, E., Gourdeau, M. *et al*. (2003). The state of infection surveillance and control in Canadian acute care hospitals. *Am. J. Infect. Cont.,* **31**, 266-273.

APPENDIX A

A.1 Protein X-ray Crystallography

This brief description of the method of X-ray crystallography was adapted from the following excellent reference textbooks:

1) Crystallization of Biological Macromolecules by Alexander McPherson (1999)
2) Introduction to Macromolecular Crystallography by Alexander McPherson (2003)
3) Principles of Protein X-ray Crystallography by Jan Drenth (2007)

X-ray crystallography is a powerful technique used for direct visualization of atomic and molecular structure solution of macromolecules. Protein structure information explains how are the atoms of a molecule arranged relative to one-another in three-dimensional space. Considering that the function of proteins is closely related to its structure, X-ray crystallography is a valuable tool for elucidating or confirming functional properties of unknown proteins. The X-ray diffraction of a pure protein crystal is collected and analyzed, leading to the information of atomic arrangement of the protein structure. While other methods exist for obtaining macromolecular structure information, protein X-ray crystallography is preferred as it in not limited by the size of the macromolecule. The biggest challenge in X-ray crystallography, however, is obtaining protein crystals in addition to them being of sufficient quality for structure solution. In

this thesis structures of $LytR^N$ and BeF-$LytR^N$ complex were determined, while crystallization of TrbC and good quality crystals of TraW were more challenging tasks.

Structures of $LytR^N$ were determined by molecular replacement (MR), for this reason this method will be explained in more detail. In general there are four main steps in protein X-ray crystallography: a) protein crystallization, b) data collection and processing, c) phase determination and structure solution, and d) structure refinement.

A.2 Protein Crystallization

Protein crystals, like all other crystals, are precisely ordered three-dimensional repetitions of single building block, also known as the unit cell. The smallest unit of the unit cell that cannot be explained by crystallographic symmetry operations, is called the asymmetric unit. On the other hand, perfect packing and arrangement of unit cells is termed the symmetric unit and is used to describe crystals via symmetry properties, repetitive features, and the distribution of atoms in the unit cell. The unit cell is defined by three repeating vectors: a, b, and c with three angles: α, β, and γ, respectively, between them. According to these parameters crystals may belong to one of the seven crystal systems as shown in table A.1, and defined by 32 allowed space groups, containing only rotational and/or translational symmetry elements.

Crystallization of proteins occurs when proteins are precipitated from a solution, and its molecules in an attempt to reach the lowest free-energy state, they pack and arrange perfectly, forming crystals. This principle is followed by all materials not only

proteins; however, there are cases when some proteins do not crystallize at all, and therefore cannot be used for crystallography. For perfect molecule packing, the protein sample to be crystallized must be of high purity. The next step involves dissolving the proteins in solvent from which they would crystallize; the protein solution is then brought to supersaturation, by the addition of a precipitant solution, to induce the crystallization process. The precipitant may be salt, or any organic molecule that will draw water away from the protein. In addition the solubility of the protein is affected by the pH, and temperature and for this reason protein crystallization can become complex. There are many methods used to crystallize proteins, while they all follow the general principle to supersaturate protein molecules, they vary by how the protein solution and the precipitant solutions are mixed. The most popular method is vapor diffusion, where the protein is equilibrated against the precipitant solution (mother liquor) through the vapor phase; this can be achieved using either the hanging drop or the sitting drop vapor diffusion, where the water as it leaves the protein/precipitant mixture drop it results in protein supersaturation and crystallization. Figure A.2, shows the sitting drop vapor equilibration. Once crystals are obtained, they have to be prepared for crystal mounting and X-ray diffraction analysis. Usually most of crystals are diffracted at 100 K, in order to reduce damage to the protein caused by X-rays. For this reason they have to be flash frozen or vitrified in liquid nitrogen prior to being subjected to X-rays. When frozen in liquid nitrogen the crystal is cooled rapidly so that the mother liquor forms an amorphous glass surface around the crystal, protecting it, while a continuous stream of liquid nitrogen is used to keep the crystals at ~ 100K during data collection.

Crystal System	*Types of Lattices*	*Minimum Symmetry of Unit Cell*	*Unit Cell Edges and Angles*	*Diffraction Symmetry*	*Permissible Space Groups*
Triclinic	P	None	$a \neq b \neq c$ $\alpha \neq \beta \neq \gamma$	$\bar{1}$	P1
Monoclinic	P	A single twofold axis	$a \neq b \neq c$ $\alpha = \gamma = 90°$ $\beta \neq 90°$	2/m	P2, $P2_1$ C2
Orthorhombic	P C I F	Three mutually perpendicular twofold axes	$a \neq b \neq c$ $\alpha = \beta = \gamma = 90°$	mmm	P222, $P2_12_12_1$, $P222_1$, $P2_12_12$ C222, $C222_1$ I222, $I2_12_12_1$ F222
Tetragonal	P I	A single fourfold axis	$a = b \neq c$ $\alpha = b = \gamma = 90°$	4/m 4/mmm	P4, $P4_1$, $P4_3$, $P4_2$ I4, $I4_1$ P422, $P4_122$, $P4_322$, $P4_222$, $P42_12$, $P4_12_12$, $P4_32_12$, $P4_22_12$ I422, $I4_122$
Trigonal/ rhombohedral	R P	A single threefold axis	$a = b = c$ $\alpha = \beta = \gamma \neq 90°$	$\bar{3}$ $\bar{3}$ m	R3 P3, $P\bar{3}_1$, $P3_2$ R32 P321, P312 $P3_121$, $P3_221$, $P3_112$, $P3_212$
Hexagonal	P	A single sixfold axis	$a = b \neq c$ $\alpha = \beta = 90°$ $\gamma = 120°$	6/m 6/mmm	P6, $P6_1$, $P6_5$, $P6_3$, $P6_2$, $P6_4$ P622, $P6_122$, $P6_522$, $P6_322$, $P6_222$, $P6_422$
Cubic	P I F	Threefold axes along cube diagonals	$a = b = c$ $\alpha = \beta = \gamma = 90°$	m3 m3m	P23, $P2_13$ I23, $I2_13$ F23 P432, $P4_132$, $P4_332$, $P4_232$ I432, $I4_132$ F432, $F4_132$

Figure A.1 The seven crystal systems and the allowed space groups for macromolecules. Copied from McPherson, 2003.

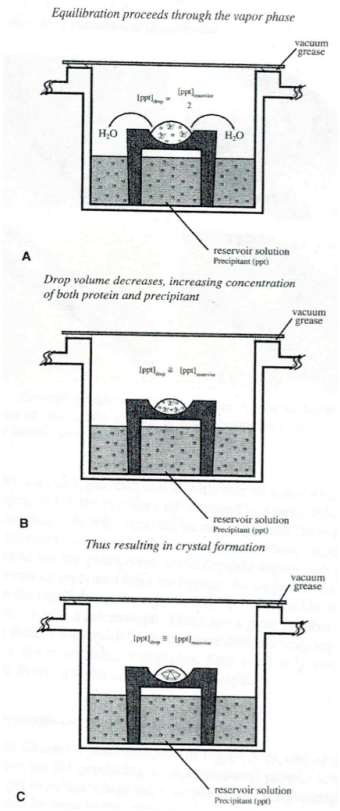


Figure A.2 The process of vapor equilibration through loss of water vapor to the larger volume of reservoir, inducing crystallization when the two are in equilibrium with the same concentration of the precipitant in the drop and in the reservoir. Copied from McPherson, 1999.

A.3 Data Collection of X-ray Diffraction Data

The wave nature of X-rays and the ability of crystals to diffract X-rays became clear in the early 1900s. Sir Lawrence Bragg came up with a mathematical equation (1) used to describe constructive interference of reflective waves by a series of lattice planes, which can also be used to describe reflection of incident X-rays by crystals.

$$n\lambda = 2d_{hkl}\sin\theta_{hkl} \quad (1)$$

Where n is an integer, λ is the wavelength of the incident radiation, d_{hkl} is the perpendicular distance between the adjacent reflecting planes, as defined by the Miller indices hkl (relative to unit cell), and θ_{hkl} is the angle between the incident/reflected radiation and the reflecting planes.

Electrons have the ability to diffract X-rays. When electrons are hit by electromagnetic radiation, they oscillate at the same frequency as the incident radiation; in addition they scatter the radiation with the same frequency as the incident wave forming a cloud around the atomic nuclei. This in turn identifies the location of the atoms where the electrons are oscillating. The *Structure factor*, (F(S)) (2) is used to describe the relation between the molecules and the scattered X-rays of electrons from each atom of the molecule.

$$F(S) = \sum_{j=1}^{n} f_j e^{(2\pi r_j \bullet S)} \quad (2)$$

where f is the scattering factor of an atom, r is the position of the atom, and |S| = (2sinθ)/λ. The magnitude of f_j is equal to the number of electrons in a certain atom j, the summation of all the scattering factors of atoms in a molecule constitutes to the overall

structure factor of the molecule, which is dependent of the types of atoms present as well as their position in the unit cell.

Since the position of electrons may explained in terms of vectors (which have both magnitude and direction) the structure factor may then be described in terms of vectors, A and B, $F(S) = A + iB$, where $A=|F(S)|\cos\alpha$, $B=|F(S)|\sin\alpha$, and $|F(S)|$ is the magnitude of the structure factor while α is the phase angle. Because of the possibility of the electron position anywhere surrounding the nucleus, the structure factor can be both above or below the horizontal axis, and for electron position below the horizontal axis they are represented by $F^*(S) = A - iB$.

The higher the number of electrons in a molecule the higher the intensity of the reflected X-rays, for this reason the intensity is proportional to the product of structure factor above and below the horizontal axis (3).

$$I \alpha F(S) \bullet F^*(S) \qquad (3)$$

Expanding the relations of structure factors in terms of the positions A and B, we get the expression for the intensity of the reflection as (4).

$$I \alpha |F(S)| \bullet e^{i\alpha} \bullet |F(S)| \bullet e^{-i\alpha} \qquad (4)$$

This expression then becomes I α $|F(S)|^2 \bullet e^{(i\alpha - i\alpha)}$, and it reduces to I α $|F(S)|^2$, where the information regarding the phase angle α, is lost. This is known as the phase problem, and for this reason in X-ray crystallography we use, multiple-isomorphous replacement (MIR), multiple-wavelength anomalous dispersion (MAD), or molecular replacement (MR) for obtaining phase information about the position of the electrons, hence atoms in the molecule.

In X-ray diffraction experiments the relationship between the intensity of the observed reflections and the structure factor is also dependent on other factors as shown in equation (5).

$$I(hkl) = \frac{\lambda^3}{\omega \cdot V^2}\left(\frac{e^2}{mc^2}\right)^2 \cdot V_{Cr} \cdot I_O \cdot L \cdot P \cdot A \cdot \left|F(hkl)\right|^2 \quad (5)$$

where λ is the wavelength; ω is the angular rotation of the crystal; V is the volume of the unit cell; e and m, are the charge and the mass of an electron, and c is the speed of light; V_{Cr} is the volume of the crystal; I_O is the intensity of the incident beam; L is the Lorentz factor, which depends on the data collection technique; P is the polarization factor of the incident and reflected intensities; and A is the absorption factor and it depends on crystal size and composition.

A.4 Structure Solution and Phase Determination

For determining the position of the actual electron density we need the amplitude as well as the phase information, which is given by α. The MIR method requires attachment of a heavy atom to the protein molecule in such a manner that the three-dimensional structure of the protein is not affected by it. Since macromolecular crystals are composed of at least 50% solvent, through diffusion, heavy atoms might reach specific amino acid side chains on the molecule and not affect the structure. The only difference between the intensities obtained from the two different forms of crystals would be due to the presence of the heavy atoms. The second method used is MAD, which required the incorporation of heavy atoms into the protein structure as well as an X-ray source with at least two different wavelength options. The best way commonly used for incorporating the heavy atom is the seleno-methionine (Se-Met), which replaces the Sulfur in the Methionine residues, but incorporation of Fe in heme groups, or Zn and Cu may work. This method requires the identification of the positions, the x, y, z coordinates of the anomalous scatters, which can be determined by the difference in Patterson maps, which use the difference between h,k,l and –h,-k.-l, to locate the atoms within the protein. The third method of choice is MR, but available only when similar structures exist. The two structures known and unknown must have sequence similarity and similar folding in order for the known structure to be used for the determination of the initial phases of the unknown. This method does not compare structure factors which require both magnitude and phase information, rather it uses the Patterson functions P(uvw) of both molecules for comparison (6).

$$P(uvw) = \frac{1}{V}\sum_{hkl}\left|F(hkl)\right|^2 \cos\left[2\pi(hu + kv + lw)\right] \qquad (6)$$

The Patterson function is the square of structure factors, with all the phase angles set to zero. Patterson peaks represent the vectors between two atoms, and are proportional to the number of electrons in the atoms between the vectors. For this reason, interatomic vectors (self-Patterson vectors) will be small, whereas vectors between two atoms (cross-Patterson vectors) will be large due to the combination of electrons. Self-Patterson vectors are used in rotation functions, since they provide information regarding the location of the atoms in the unit cell. Cross-Patterson vectors are used in translation functions, providing information on the molecule's position with respect to the other symmetry related molecules in the unit cell. There are various computer programs that successfully can determine the rotational and translational search between the known and unknown structures. The known structure is then placed in the unit cell of the unknown structure, and using the known structure coordinates and structure factors, the unknown molecule can be used to calculate the structure factors of the unknown crystal. While this is only approximate since the two molecules are not identical, the calculated structure factors (F_{Calc}) are compared to the observed structure factors (F_{Obs}) suffice in obtaining the correct structure solution, after cycles of improvement and refinement. R-value (7) and correlation coefficient (CC) (8) are good indicators of how correct are the structure

solution by comparing the differences between the observed and calculated structure factors.

$$R = \frac{\sum_{hkl} \left| |F_{Obs}| - |F_{Calc}| \right|}{\sum_{hkl} |F_{Obs}|} \qquad (7)$$

The closer F_{Calc} resembles F_{Obs}, the lower R-value becomes lower, values of less that 25% are normally accepted.

$$CC = \frac{\sum_{hkl} (|F_{Obs}|^2 - |\overline{F_{Obs}}|^2) \cdot (|F_{Calc}|^2 - |\overline{F_{Calc}}|^2)}{\left[\sum_{hkl} (|F_{Obs}|^2 - |\overline{F_{Obs}}|^2)^2 \sum_{hkl} (|F_{Calc}|^2 - |\overline{F_{Calc}}|^2)^2 \right]^{1/2}} \qquad (8)$$

On the other hand, the closer the CC value is to 1 the more closely does the F_{Cacl} represent $F_{Obs.}$

A.5 Structure Refinement

After obtaining the phase information of the structure must be refined. This is done by fitting the structure into the observed electron density, as well as the addition of other structural components such as ligands, waters and other solvent molecules. The model building can be done using a computer program such as COOT, and refinement with Refmac, which calculates how close are the calculated structure factors to those observed using R-values (7). To confirm correct structure solutions R_{free} was introduced which uses a randomly chosen 5-10% of the observed reflections and compares them to the calculated structure factors, and as such it is unbiased by the refinement process (Brünger, 1993a). In the final stages of the structure refinement there are various software packages such as SFCHECK, PROCHECK, and EBI-PISA that are used to validate the structure, including bond lengths, ligands, and the correct stereochemistry of the amino acids.

References:

Drenth, J. (2007). Principles of Protein X-ray Crystallography. Springer Science+Business Media, LLC, New York.

Brünger, A.T. (1993a). Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Crystallog. Sect. D* **49**, 24-36.

McPherson, A. (2003). Introduction to Macromolecular Crystallography. John Wiley & Sons, Inc., New Jersey.

McPherson, A. (1999). Crystallization of Biological Macromolecules. Cold Spring Harbor Laboratory Press, New York.

APPENDIX B

Protein sequence and relevant statistics for the proteins studied in this Dissertation

TrbC-His$_7$ (Full length TrbC, with a C terminal 7His tag)

```
        10          20          30          40          50          60
MASENVNTPE NRQFLKQQEN LSRQLREKPD HQLKAWAEKQ VLENPLQRSD NHFLDELVRK

        70          80          90         100         110         120
QQASQDGKPR QGALYFVSFS IPEEGLKRML GETRHFGIPA TLRGMVNNDL KTTAEAVLSL

       130         140         150         160         170         180
VKDGATDGVQ IDPTLFSQYG IRTVPALVVF CSQGYDIIRG NLRVGQALEK VAATGDCRQV

       190         200
AHDLLAGKGD SGKGSHHHHH HH
```

Number of amino acids	202
Molecular weight	22532.3 Da
Theoretical pI	7.88
Abs (1 g/l)	0.448

His$_6$-TrbC (Full length TrbC, with a cleavable N terminal 6His tag)

```
        10          20          30          40          50          60
MGSSHHHHHH SSGLVPR||GSH MSENVNTPEN RQFLKQQENL SRQLREKPDH QLKAWAEKQV

        70          80          90         100         110         120
LENPLQRSDN HFLDELVRKQ QASQDGKPRQ GALYFVSFSI PEEGLKRMLG ETRHFGIPAT

       130         140         150         160         170         180
LRGMVNNDLK TTAEAVLSLV KDGATDGVQI DPTLFSQYGI RTVPALVVFC SQGYDIIRGN

       190         200         210
LRVGQALEKV AATGDCRQVA HDLLAGKGDS GK
```

Number of amino acids	212	
Molecular weight	23520.4 Da	
Theoretical pI	8.56	
Abs (1 g/l)	0.429	|| TEV protease cleavage site

TrbC (Full length TrbC, with a cleaved tag)

```
        10          20          30          40          50          60
GSHMSENVNT PENRQFLKQQ ENLSRQLREK PDHQLKAWAE KQVLENPLQR SDNHFLDELV

        70          80          90         100         110         120
RKQQASQDGK PRQGALYFVS FSIPEEGLKR MLGETRHFGI PATLRGMVNN DLKTTAEAVL

       130         140         150         160         170         180
SLVKDGATDG VQIDPTLFSQ YGIRTVPALV VFCSQGYDII RGNLRVGQAL EKVAATGDCR

       190
QVAHDLLAGK GDSGK
```

Number of amino acids	195
Molecular weight	21638.4 Da
Theoretical pI	7.97
Abs (1 g/l)	0.467

TrbC^{C}-His$_6$ (C terminal domain TrbC, with a C terminal 6His tag)

```
        10          20          30          40          50          60
MGEEGLKRML GETRHFGIPA TLRGMVNNDL KTTAEAVLSL VKDGATDGVQ IDPTLFSQYG

        70          80          90         100         110         120
IIRTVPALVV FCSQGYDIIR GNLRVGQALE KVAATGDCRQ VAHDLLAGKG DSGKLEHHHH

HH
```

Number of amino acids	122
Molecular weight	1318.0 Da
Theoretical pI	6.70
Abs (1 g/l)	0.236

TraW-His$_7$ (Full length TraW, with a C terminal 7His tag)

```
        10          20          30          40          50          60
MADLGTWGDL WPVKEPDMLT VIMQRLTALE QSGEMGRKMD AFKERVIRNS LRPPAVPGIG

        70          80          90         100         110         120
RTEKYGSRLF DPSVRLAADI RDNEGRVFAR QGEVMNPLQY VPFNQTLYFI NGDDPAQVAW

       130         140         150         160         170         180
MKRQTPPTLE SKIILVQGSI PEMQKSLDSR VYFDQNGVLC QRLGIDQVPA RVSAVPGDRF

       190         200
LKVEFIPAEE GRKGSHHHHH HH
```

Number of amino acids	202
Molecular weight	22925.2 Da
Theoretical pI	7.93
Abs (1 g/l)	0.980

His$_6$-TraW (Full length TraW, with a cleavable N terminal 6His tag)

```
        10          20          30          40          50          60
MGSSHHHHHH SSGLVPR||GSH MADLGTWGDL WPVKEPDMLT VILQRLTALE QSGEMGRKMD

        70          80          90         100         110         120
AFKERVIRYS LRPPAVPGIG RTEKYGSRLF DPSVRLAADI RNNEGRVFAR QGEVMNPPQY

       130         140         150         160         170         180
VPFNQTLYFI NGDDPAQVAW MKRQTPPTLE SKIILVQGSI PEMQKSLDSR VYFDQNGVLC

       190         200         210
QRLGIDQVPA RVSAVPGDRF LKVEFIPAEE GRK
```

Number of amino acids	213	
Molecular weight	23998.4 Da	
Theoretical pI	9.06	
Abs (1 g/l)	0.998	|| TEV protease cleavage site

TraW (Full length TraW, with a cleaved tag)

```
        10          20          30          40          50          60
GSHMADLGTW GDLWPVKEPD MLTVILQRLT ALEQSGEMGR KMDAFKERVI RYSLRPPAVP

        70          80          90         100         110         120
GIGRTEKYGS RLFDPSVRLA ADIRNNEGRV FARQGEVMNP PQYVPFNQTL YFINGDDPAQ

       130         140         150         160         170         180
VAWMKRQTPP TLESKIILVQ GSIPEMQKSL DSRVYFDQNG VLCQRLGIDQ VPARVSAVPG

       190
DRFLKVEFIP AEEGRK
```

Number of amino acids	196
Molecular weight	22116.4 Da
Theoretical pI	8.72
Abs (1 g/l)	1.083

MBP-TrbC (Full length TrbC, with an N terminal MBP tag)

```
        10          20          30          40          50          60
MKIKTGARIL ALSALTTMMF SASALAKIEE GKLVIWINGD KGYNGLAEVG KKFEKDTGIK

        70          80          90         100         110         120
VTVEHPDKLE EKFPQVAATG DGPDIIFWAH DRFGGYAQSG LLAEITPDKA FQDKLYPFTW

       130         140         150         160         170         180
DAVRYNGKLI AYPIAVEALS LIYNKDLLPN PPKTWEEIPA LDKELKAKGK SALMFNLQEP

       190         200         210         220         230         240
YFTWPLIAAD GGYAFKYENG KYDIKDVGVD NAGAKAGLTF LVDLIKNKHM NADTDYSIAE

       250         260         270         280         290         300
AAFNKGETAM TINGPWAWSN IDTSKVNYGV TVLPTFKGQP SKPFVGVLSA GINAASPNKE

       310         320         330         340         350         360
LAKEFLENYL LTDEGLEAVN KDKPLGAVAL KSYEEELAKD PRIAATMENA QKGEIMPNIP

       370         380         390         400         410          420
QMSAFWYAVR TAVINAASGR QTVDEALKDA QTNSSSNNNN NNNNNNLGIE GR||ISEFASEN

       430         440         450         460         470         480
VNTPENRQFL KQQENLSRQL REKPDHQLKA WAEKQVLENP LQRSDNHFLD ELVRKQQASQ

       490         500         510         520         530         540
DGKPRQGALY FVSFSIPEEG LKRMLGETRH FGIPATLRGM VNNDLKTTAE AVLSLVKDGA

       550         560         570         580         590         600
TDGVQIDPTL FSQYGIRTVP ALVVFCSQGY DIIRGNLRVG QALEKVAATG DCRQVAHDLL

       610
AGKGDSGKGS
```

Number of amino acids	610	
Molecular weight	66930.8 Da	
Theoretical pI	5.75	
Abs (1 g/l)	1.142	|| Xa Protease cleavage site

References:

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. *Protein Identification and Analysis Tools on the ExPASy Server.* (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005), pp. 571-607