STRUCTURAL AND FUNCTIONAL STUDIES OF TRAG FROM THE F PLASMID AND THE MAJOR PILIN FROM *COXIELLA BURNETII*

FETTAH ERDOGAN

A THESIS SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

GRADUATE PROGRAM IN CHEMISTRY
YORK UNIVERSITY
TORONTO, ONTARIO

August 2016

© Fettah Erdogan, 2016

Abstract

The *Escherichia coli* F-plasmid houses all genes necessary for the assembly of its conjugative machinery to facilitate successful DNA transfer. TraG is one of the largest proteins (102 kDa) encoded in the transfer (tra) region of the F-plasmid that is known to facilitate conjugative transfer. The N-terminal region of TraG is known to be involved in pilus synthesis and assembly, while the C-terminal periplasmic domain is required for plasmid-specific entry exclusion and mating pair stabilization. In order to provide structural insights into TraG's function, we have expressed, purified and initiated crystallization studies of the C-terminal region of TraG (designated TraG*, 52.7 kDa) and have obtained initial crystals of TraG*. Our experiments on functional and dynamic characterization of TraG* using liquid chromatography coupled mass spectrometry (LC-MS), size exclusion chromatography coupled multi-angle light scattering (SEC-MALS) and protein gel electrophoresis has given significant insight into TraG* crystallization strategies as well as potential TraG mechanism of action during conjugation.

Acknowledgements

I would like to thank the following people who have made this work possible: Dr. Gerald F. Audette for his insight, mentorship, and support as my graduate supervisor; Dr. Derek J. Wilson for allowing me to use their Mass Spectrometry facility; Dr. Mark A. Bayfield for being part of my supervisory committee; Members of the Audette Lab, both past and present, who have shared much of their knowledge and made my experience an exception: Cristina Lento, Ayat Yaseen, Alex Kheyson, Roksana Nowrouzi, Irina Oganesyan, Fatima A. Portillo, Daniel Tsoy, and Gerald F. Audette, Members of the Wilson Lab who have collaborated with care and respect; Shaolong Zhu, Kerene Brown, and Bin Deng. And my whole family, with special thanks to my parents who have been a constant source of encouragement, love, and support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AIEX – anion exchange
Amp – Ampicillin
°C – degrees Celsius
Cm – Chloramphenicol
DNA – Deoxyribonucleic Acid
E. coli – *Escherichia coli*
ESI – electrospray ionization
g – grams
HGT – Horizontal Gene Transfer
His – Histidine
hr(s) – hour(s)
IMAC – Immobilized metal affinity chromatography
IPTG – Isopropyl-β-D-thiogalactopyranoside
Km – Kanamycin
LB – Luria Bertani
M – molar
MALS – multi-angle light scattering
Mdr – Multi-drug Resistant
min – minutes
mL – milliliter
mM – millimolar
Mpf – Mating Pair Formation
Mps – Mating Pair Stabilization
MS – mass spectrometry
MW – Molecular Weight
MWCO – Molecular Weight Cut Off
N-terminal – amino terminal
PCR – Polymerase Chain Reaction
PMSF – phenylmethylsulfonyl fluoride
rpm – revolutions per minute
SDS- PAGE – Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC – size exclusion chromatography
T4SS –Type IV Secretion System
TAE – tris-acetic acid-EDTA
tra – Transfer
µg – micrograms
µL – microlitres
µM – micromolar

1. INTRODUCTION

1.1. Conjugation and Drug Resistance

The discovery of penicillin in 1928 by the Scottish scientist Alexander Fleming was thought to be the beginning of the end of bacterial infections. Within a year of its use as a treatment against bacterial infections, this hope was dashed by the identification of bacterial resistance towards the antimicrobial drug penicillin (Rammelkamp & Maxon, 1942). Antimicrobial resistance in bacteria is acquired via either chromosomal mutation in their genomes or horizontal gene transfer (HGT) that may take place amongst related or unrelated bacterial species. These events can take place in the presence or absence of antibiotics (Furuya & Lowy, 2006). As a result, bacterial infections due to antimicrobial-resistance have become one of the major focuses of research in the field of medicine.

HGT can be defined as the movement of genetic material between unicellular and/or multi-

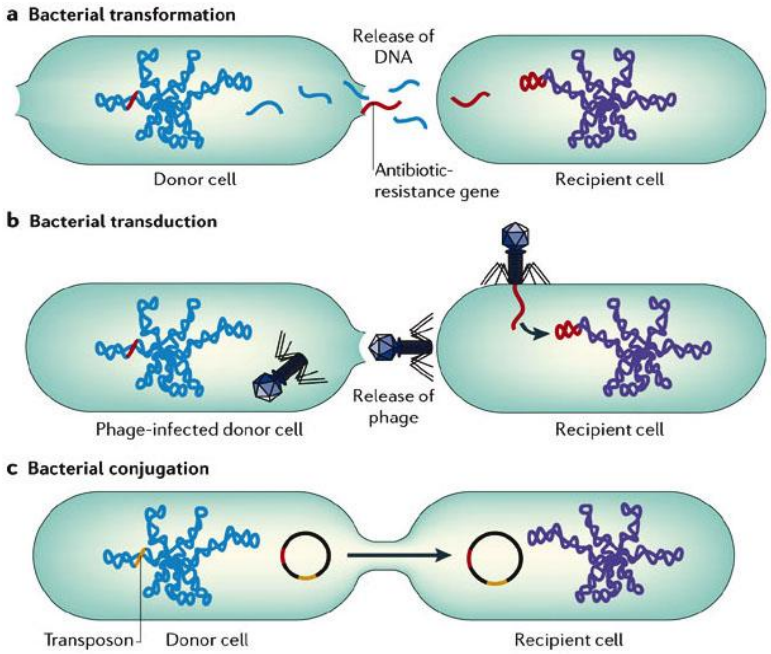


Figure 1: Pathways of bacterial gene transfer (adapted from Furuya & Lowy, 2006).

cellular organisms via any route outside of vertical transmission (i.e. without cell division) (Keeling & Palmer, 2008). This process often involves mediator bacteriophages and plasmids. In essence, genes possessing antibiotic resistance properties in one bacterium are passed onto another bacterium to equip the recipient bacterium with same antibiotic resistant genes that the donor possessed, allowing for the share of antibiotic resistance features. This exchange of gene fragments can take place via three different routes of transfer; cellular transformation, conjugation through cell-to-cell contact and transduction using a viral vector (**Figure 1**), (Dobrindt et al., 2004; Furuya & Lowy, 2006).

Transformation involves the transfer of cellular DNA via uptake of DNA molecules that are freely available in the surrounding environment of a cell (Frost et al., 2005). The process of conjugation is facilitated by the independently replicating genetic elements in the form of conjugative plasmids or chromosomally integrated conjugative elements (ICEs) which encode proteins that facilitate the transfer of DNA from a donor cell to a recipient cell lacking the plasmid or ICE (Burrus & Waldor, 2004). The process takes place in the following manner (**Figure 2**); a donor cell protrudes a conjugative pilus that identifies a recipient cell. The pilus then attaches and retracts to bring the two cells close together. After a series of steps, a mating bridge forms between the two cells and the DNA material (designated as transfer DNA or T-DNA) is transported into the recipient cell with an accompanying relaxase (a nucleoprotein complex). Inside the recipient cell, the T-DNA is re-circularized, replicated and established to give the recipient cell "a donor (F^+) cell status" (Arutyunov & Frost, 2013). DNA transfer via the route of transduction is mediated by bacterial viruses called bacteriophages. These viruses can package segments of host DNA into their capsids and then inject this DNA into a new host which may then be incorporated into a recipient cell's genome and be inherited (Frost et al., 2005).

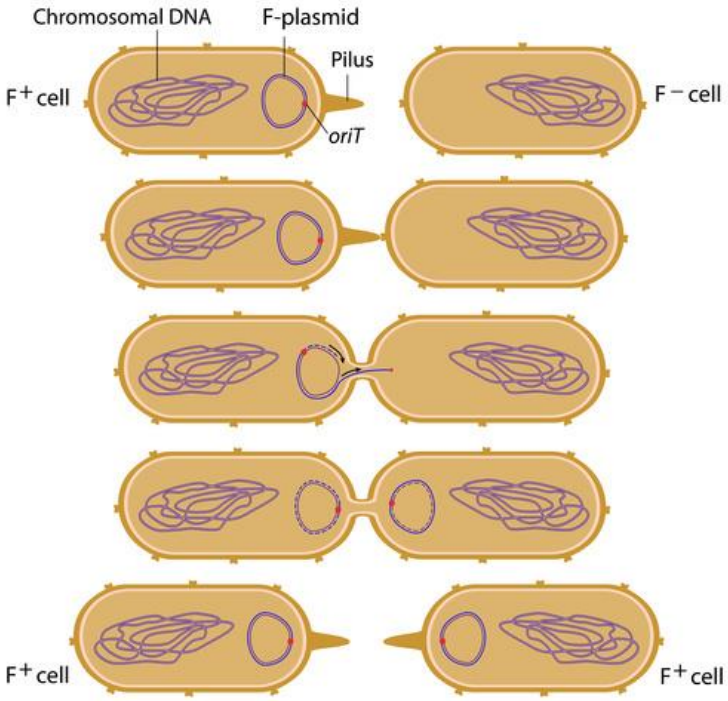


Figure 2: The process of conjugation in bacteria (adapted from (Le, 07-04-2016))**.**

Conjugation is a unique process compared to transformation and transduction because during conjugation, the transfer of DNA takes place in a donor-controlled fashion (in gram-negative bacteria), by complex systems that bridge and connect a donor bacterium to a host cell. This is the most direct way in which bacterial cells interact with host cells to inject genes, proteins or chemicals in to their systems (Griffiths et al., 2000; Smillie et al., 2010). Quite often, the transfer of such agents has remarkable effects on the host, ranging from pathogenesis and carcinogenesis to host evolution and adaptation. It has been shown that the rate of adaptation in bacteria increases 3-fold by genetic recombination (Cooper, 2007). Moreover, conjugation by far is the most common route through which antibiotic resistance genes in bacterial strains are spread (Lujan, et al., 2007). Although conjugation may seem to be a pro-infection concern, it can also offer benefits for the public as well. Using conjugation, the delivery of plasmids carrying antimicrobial agents

into a new host can be used as a treatment against multi-drug resistant pathogens, such as *Acinetobacter baumanii* (Arutyunov & Frost, 2013). Systems in which uncontrolled plasmid replication takes place (runaway plasmids) and/or those that deliver plasmid coupled toxins are lethal for recipient cells lacking the regulators of plasmid replication or anti-toxins (Filutowicz et al., 2008). In order to prevent the spread of antibiotic resistance genes, one can inhibit plasmid conjugation by developing drugs that target bacterial transfer complexes (Williams & Hergenrother, 2008).

1.2. Secretion Systems

Horizontal transfer is not limited to sharing of DNA material between cells. Micro-organisms are equipped with systems that can also facilitate the transfer of protein effectors and chemical messengers between bacteria. Many of these processes play critical roles in bacterial survival and host infection pathways. These transfer complexes that facilitate the transfer of subcellular materials, generally plasmids, proteins, enzymes or toxins, are named secretion systems. There are 9 types of secretion systems (TSS's); T1SS, T2SS, T3SS, T4SS, T5SS, T6SS, T7SS, the Sec (secretion) and Tat (two-arginine translocation) pathways (**Figure 3**, Tseng, et al., 2009). Each type of secretion system is further divided into different subtypes. The divisions and sub-divisions of secretion systems have become necessary due to diversity of proteins and the distinctiveness of pathways involved in different bacterial strains. For example, within the type IV secretion system (T4SS), the Ti and Cag systems facilitate effector transport whereas the F-plasmid, R27 and

pKM101 T4SS machineries facilitate plasmid transfer, using homologous proteins and altered pathways.

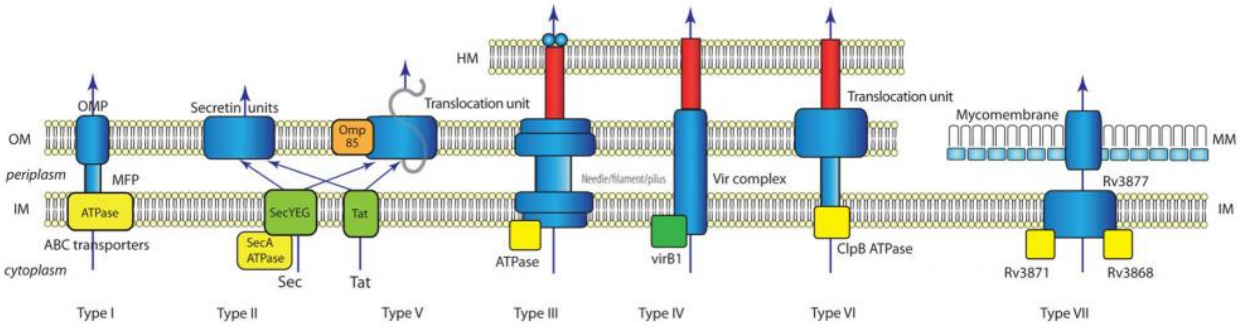


Figure 3: Bacterial Secretion Systems. A simplified diagram showing the basics of various secretion systems know up to date. The abbreviations stand as follows; HM: Host membrane; OM: outer membrane; IM: inner membrane; MM: mycomembrane; OMP: outer membrane protein; MFP: membrane fusion protein. ATPases and chaperones are shown in yellow (adapted from Tseng et al., 2009).

1.3. F-Plasmid Conjugative System

As is the case with all other secretion systems, the F-plasmid conjugation system possesses essential transfer elements that are required for building its transfer machinery to facilitate the process of conjugation. These elements (**Table 1** and **Figure 4**) encoded in the 33.3 kb transfer (*tra*) region of the 100 kb F-plasmid (**Figure 5**) in *Escherichia coli* (Willetts & Skurray, 1980) consist of the following; DNA transfer proteins (Dtr) made up of a relaxase and its assisting proteins and a core complex of type IV secretion system (T4SS) possessing a conjugative pilus for mating pair formation (Mpf), and a type 4 coupling protein (T4CP). The T4CP at the cytoplasmic end of the channel links the Dtr and Mpf to generate a pumping platform for the transfer of the F-plasmid relaxosome-nucleoprotein complex, generated by a relaxase and its auxiliary components (Arutyunov & Frost, 2013). The functions of the other proteins (**Table 1**) encoded in the *tra* region of F-plasmid include surface exclusion (Sfx) which prevents incorrect donor-donor contacts, entry

exclusion (Eex) which prevents repetitive nucleoprotein transport between donors, and mating pair stabilization (Mps) for efficient plasmid transfer (Audette et al., 2007; Frost et al., 1994; Lawley et al., 2003).

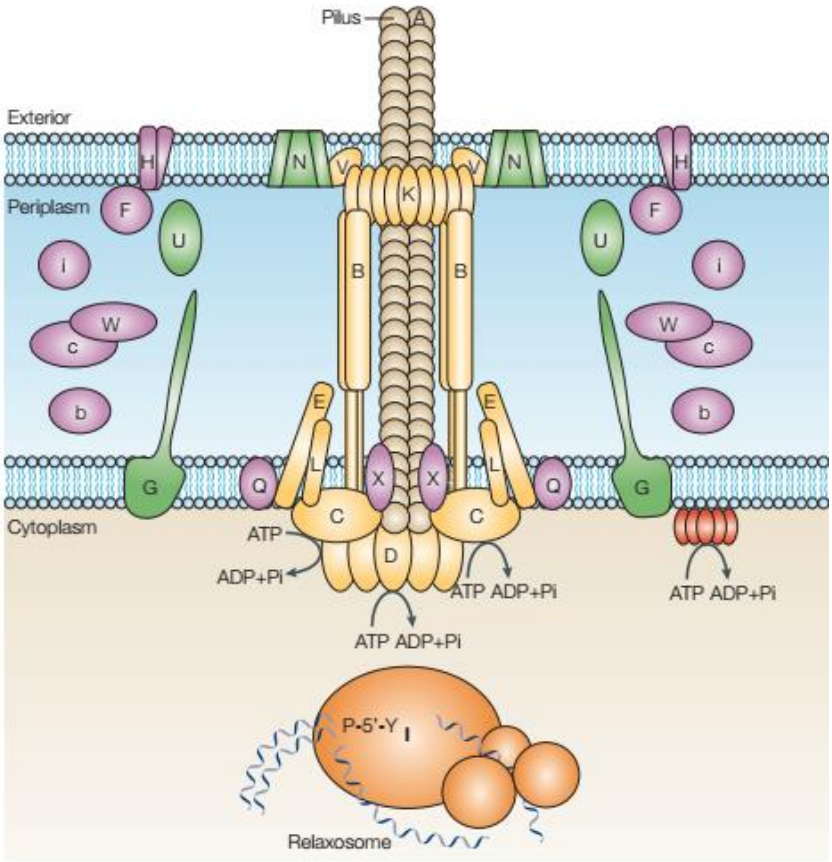


Figure 4: Signature proteins of the F-type conjugation system. In gram-negative bacteria, the F-type (F) conjugative system extends from the inner membrane, through the periplasm and onto the outer membrane. Tra proteins are labelled with capital letters and Trb proteins with lower-case letters. (adapted from Frost et al., 2005).

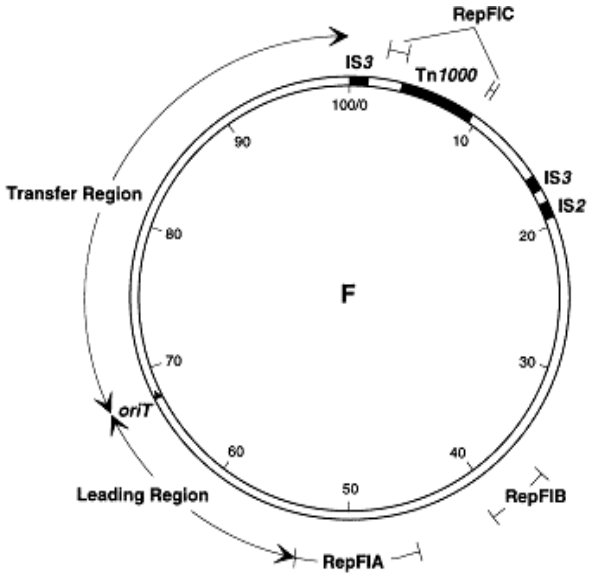


Figure 5: The F-plasmid. IS = Insertion sequences, Tn1000= transposon 1000, OriT = origin of conjugal transfer, Transfer region = conjugal transfer proteins, RepFIA, B, C = plasmid replicons (adapted from Neville Firth, Ippen-ihler, & Skurray, 1996)**.**

1.3.1. TraG

TraG, the main transfer protein of interest in this study, is one of the largest proteins (938 aa) encoded in the *tra* region of the F-plasmid. It is reported to be anchored in the inner membrane via the N-terminal domain and has its C-terminal domain extending into the periplasm (Firth & Skurray, 1992). The N-terminal of TraG is known to facilitate pilus synthesis, while the C-terminal periplasmic domain is required for Mps (Achtman et al., 1972; Audette et al., 2007). Some have suggested that the C-terminal end of TraG interacts with TraN protein which is located in the outer membrane of the cell to stabilize the pairing between the donor and the recipient cells (Klimke et al., 2005). This however, is yet to be proven. TraG interaction with other transfer proteins is also deemed necessary for Mps (Audette et al., 2007). Furthermore, TraG has been shown to interact with the entry exclusion (Eex) factor TraS of the recipient cell as well. Therefore, via its interaction

with TraS, TraG may potentially be involved in entry exclusion functions in preventing redundant conjugative DNA synthesis and transportation (Audette et al., 2007).

It was initally thought that, during Mps, the C-terminal regoin of TraG was cleaved from the intact protein, producing a 458aa periplasmic fragment designated TraG*; however Audette and collegues demonstrated that this was not the case and that full-length TraG is required for Mps (Audette et al., 2007). Others have suggested that TraG* may be translocated into the recipient cell in order to interact with TraS to control conjugative DNA synthesis (Arutyunov & Frost, 2013; Marrero & Waldor, 2007). The amount of knowledge about the pathway and the interactions that TraG is involved in during conjugation is limited and there still remain some functional aspects of the protein to be clarified. Moreover, there is a difference of results and opinion on the topology of TraG, with regards to number of α-helices, β-sheets and trans-membrane segments it possesses (Audette et al., 2007; Klimke et al., 2005; Kohler et al., 2013).

Table 1: F-plasmid conjugation proteins and their functions[¶]

F conjugation				Homology to Ti/R27 proteins
Protein	**Length [Mature] (aa)**	**Size [Mature] (kDa)**	**Function**	
Orf169	169	19.3	Lytic transglycosylase	VirB1/Orf130
TraA	121 [70]	12.8 [7.2]	Pilin	VirB2/TrhA
TraB	475	50.5	Pilus assembly	VirB10/TrhB
TraC	875	99.2	Pilus assembly	VirB4/TrhC
TraD	717	81.7	DNA transport	-/TraG
TraE	188	21.2	Pilus assembly	VirB5/TrhE
TraF	247 [228]	28.0 [25.9]	Pilus assembly	-/TrhF
TraG	938 [487]	102.5 [52.7]	Pilus assembly, stability	VirB6/TrhG
TraH	458 [434-435]	50.2 [~47.8]	Pilus assembly	-/TrhH
TraI/I*	1756/802*	192.0/87.9*	Relaxase	-/TraI
TraJ	229	27.1	Regulation	-
TraK	242 [221]	25.6 [23.3]	Pilus assembly	VirB9/TrhK
TraL	91	10.4	Pilus assembly	VirB3/TrhL
TraM	127	14.5	Mating Signal	-/-
TraN	602 [584]	65.7 [63.8]	Aggregate stability	-/TrhN
TraP	196	22.0	Minor role in stability	-/-
TraQ	94	10.9	Pilus synthesis	-/-
TraR	73	8.3	Regulation	-/-
TraS	149	16.9	Surface exclusion	-/-
TraT	244 [233]	26.0 [23.8]	Surface exclusion	-/-
TraU	330 [308]	36.8 [34.3]	Pilus assembly	-/TrhU
TraV	171 [153]	18.6 [16.6]	Pilus synthesis	VirB7/TrhV
TraW	210 [192-193]	23.6 [~21.7]	Pilus synthesis	-/TrhW
TraX	248	27.5	Pilin acetylation	-/-
TraY	131	15.2	oriT nicking	-/-
TrbC	212 [191]	23.4 [21.2]	Pilus assembly	-/TrhW

¶A list of F plasmid transfer proteins involved in F conjugation with their respective polypeptide lengths in amino acids (aa), their theoretical molecular weight in kilo Daltons (kDa) and their function(s). The far right column depicts the peptide homolog of each F protein in the Ti and R27 type IV secretion systems. (Blankschien et al., 2009; Shala et al., 2016)

1.4. Homology between Type IV Secretion Systems (T4SS's)

The diversity of pathways and elements involved in the secretion systems of bacterial species has led to further classification of T4SS's into three functional groups. The first group mediates the direct transfer of DNA from one bacterial cell to another as in the process of conjugation, while the second group mediates DNA release and uptake. The systems of *Helicobacter pylori* and *Neisseria gonorrhea* resemble the second group of T4SS (Lederberg & Tatum, 1953). The third group of T4SSs mediates translocation of protein complexes ranging from small protein effectors to large proteins. This pathway is used by pathogenic gram-negative bacteria including *H. pylori, Brucella suis* and *Legionella pneumophilaue* to inject virulence proteins into mammalian host cells (Waksman & Orlova, 2014). The T4SS of *Bordetella pertussisus* secretes pertussis toxin into its extracellular environment (O'Cellaghan et al., 1999). *Coxiella burnetii,* the bacteria responsible for the unusually lethal Q-fever, is known to possess a T4SS that it uses to share its intracellular material and begin proliferating within a phagolysosome in its host (Beeckman & Vanrompay, 2010). Even though the third group of T4SS seem to be most problematic, the first group of T4SS, due to its ability to rapidly disseminate antibiotic-resistance genes and other virulence traits among pathogens, poses perhaps the largest problem in public health today (Waksman & Orlova, 2014).

This diversity of T4SS's is subtle at the genetic level however the shared homologies and evolutionary relationships between most of the T4SSs are still poorly characterized (**Figure 6**) (Cascales & Christie, 2003; Guglielmini et al., 2014; Bart Hazes & Frost, 2008; Shala et al., 2016). One of the reasons for this poor characterization is the lack of a standard nomenclature. For example, within T4SSs, the genes that have similar names are not necessarily homologs and homologous transfer proteins do not necessarily carry the same names (Guglielmini et al., 2014).

The VirB proteins of the Ti system of *Agrobacterium tumefaciens* are an exception to this and are by far the best studied T4SS (**Table 1**). Hence, the Ti transfer system is the best homology model of a T4SS (**Figure 7a**) and an ideal model that we can use to study the "tra" proteins of the F system (Darbari & Waksman, 2015). The Ti plasmid's VirB system is composed of 11 genes, ranging from VirB1 to VirB11. The core of the secretion channel complex of the VirB systems is composed of VirB7, VirB9 and VirB10 proteins, spanning the periplasm and both cell membranes of the cell. VirB10 forms a bottom ring at the inner membrane layer by lining the inner surface of the core complex chambers and spanning the whole length. VirB9 forms an outer sheath around core complex and interacts with VirB7. The proteins VirB3, VirB6 and VirB8 are thought to join the core complex to produce the inner-membrane pore (Guglielmini et al., 2014). VirB6 is the homolog of TraG of the F system. It is a polytopic integral membrane protein that possesses a periplasmic N terminus loop (P1 loop), five transmembrane helices (TMDs) and a large central periplasmic loop (P2 loop) sandwiched by the TMDs. The C-terminus is located in the cytoplasm (Darbari & Waksman, 2015; Judd, Mahli, & Das, 2005). Similar to what is suggested with the periplasmic segment of TraG, the P2 periplasmic loop of VirB6 is known to interact with a substrate protein to carry out entry exclusion functions (**Figure 7b**). Furthermore, the N-terminal P1 loop and the C-terminal regions of Virb6 coordinate with the pilin VirB2 and VirB9 to also modulate pilus assembly (Jakubowski et al, 2004). The other proteins VirB4, VirB11 and the VirD4 operate to energize pilus biogenesis which is composed of a major (VirB2) and a minor pilin (VirB5). VirB1 is a transglycosylase that degrades peptidoglycan (Guglielmini et al., 2014).

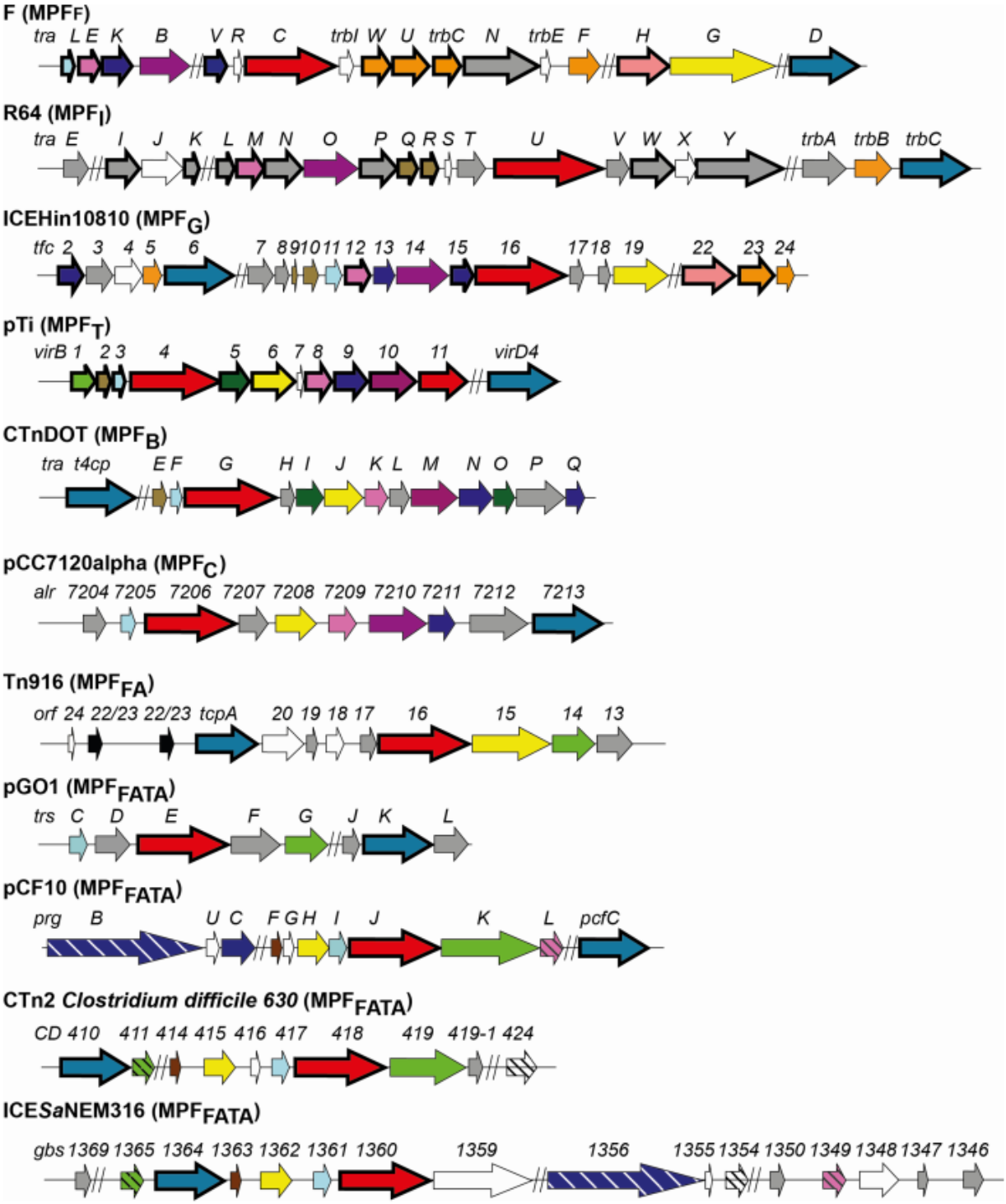


Figure 6: MPF classes within the family of T4SS and their homologies. Mating-pair formation (MPF) genes are labeled with the name of their mobile element (F for F-plasmid, Ti plasmid and so on). Arrow lengths are proportional to length of genes; white arrows represent genes whose profile is unknown; gray arrows represent genes lacking homologs in other classes. Other colors correspond to a single family of homologs (adapted from Guglielmini et al., 2014).

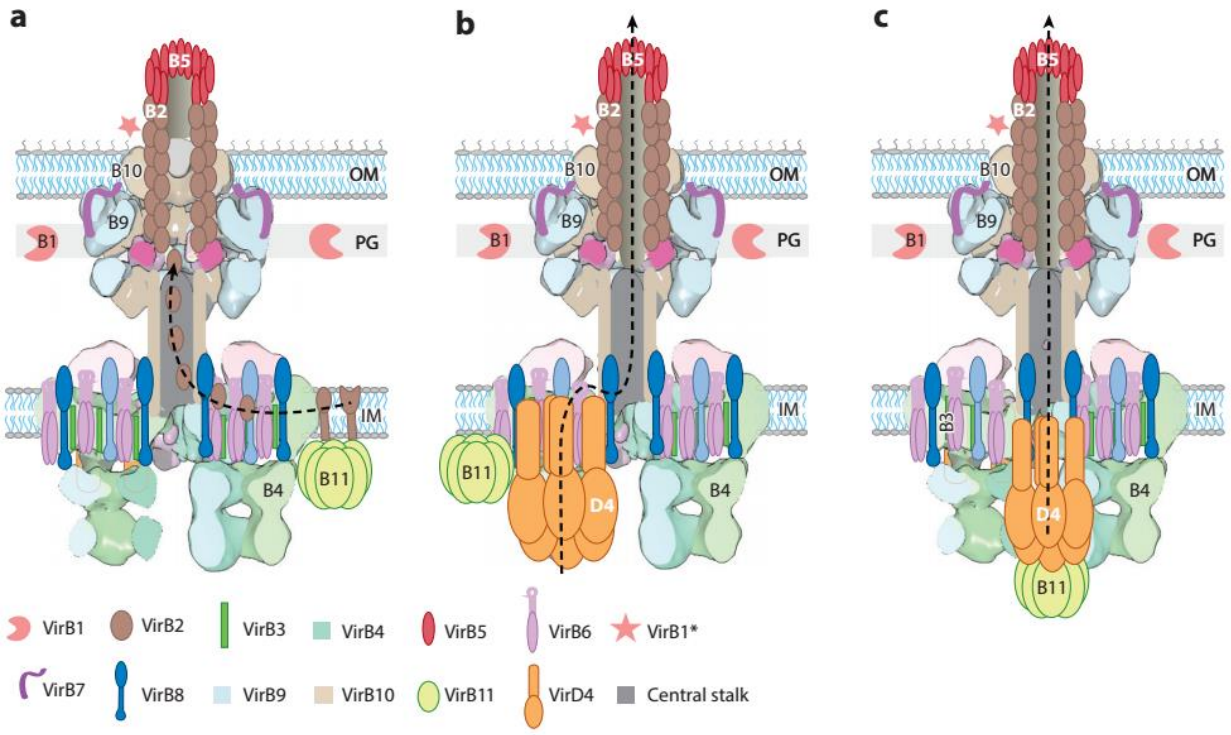


Figure 7a. An up to date model of the T4SS. The legend uses the standard VirB/D4 nomenclature. During pilus biogenesis (a) VirB2 propilins are recruited to the inner membrane (IM) and assembled into a pilus by the VirB4 and VirB11. The process of substrate translocation is predicted to happen in two different ways. The first (b) is via the coupling protein VirD4 with the aid of other cytoplasmic ATPases. In this pathway, one of the hexameric VirB4 barrels is replaced by a VirD4 hexamer and acts as a continuous platform for the substrate. The second possibility (c) is solely via VirD4 being located alongside the VirB4 hexameric barrels with VirB11 could stacked underneath or next to it. OM: outer membrane. (adapted from Darbari & Waksman, 2015).

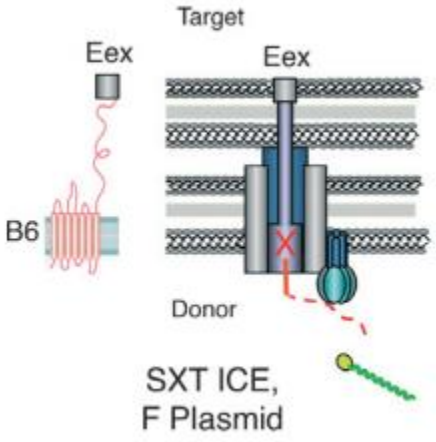


Figure 7b: A VirB6-like subunits possessing C-terminal extensions into the periplasm. The VirB6 is thought to extend either through the periplasm (shown on the left) or through the T4SS channel (on the right, without VirB6 being shown), into recipient cell's inner membrane where contact is established with an entry exclusion (Eex) protein to prevent redundant transfer to donor cells (adapted from Thanassi et al., 2012).

1.5. Research Objective I

The X-ray structure of TraG has not been determined yet. In this thesis project, our aim is to determine the crystal structure of TraG and bring further insight into the ambiguity of its potential interactions and the pathway that it takes in facilitating conjugation. The overall impression given from previous studies outlined above suggest that this protein may be quite flexible when performing its functions. Added to this difficulty the large size of the protein (102 kDa), we have decided to focus our study to crystallize only the C-terminal region of TraG (aa 452-938, designated as TraG*, with a size of 52.7 kDa).

The latest structural studies indicate that the TraG homolog VirB6 does not function as a monomer (**Figure 7b**) (Darbari & Waksman, 2015). In consideration of the potential of TraG* behaving in the similar manner as well as the intention to optimize sample preparation for crystallization, we perform light scattering, mass spectrometry and pseudo-native gel electrophoresis experiments to elucidate potential TraG*-TraG* protein-protein interactions.

1.6. Pili of Secretion Systems

Bacteria assemble a range of proteinaceous appendages on their surfaces that serve to either propel the cell through media (generally called flagella) or adhere to surrounding objects (called pili or fimbriae). Flagella are long and robust whereas pili are thin and more hair-like. The term "fimbriae" is suggested to be restricted to surface structure involving adherence and the term "pili" is used for sexual appendages (Ottow, 1975). The pili assemble and function as the part of secretion systems. Via the pilus, bacteria attach to specific host cells to initiate a process of infection or conjugation. Pili carry adhesive elements at their tips which behave as lectins in recognizing oligosaccharide residues of glycoprotein or glycolipid receptors on a target cell. These adhesins also bind to structural elements of membranes such as collagen and fibronectin (Proft & Baker, 2009). Moreover, the ability of some pilins to self-assemble makes them a good candidate for their use in bionanotechnology as protein nanotubes (PNTs) (Petrov & Audette, 2012; Petrov et al., 2013).

Amongst the secretion systems depicted in **Figure 3**, the only systems that have been reported to assemble pili are the T2SS, T3SS and T4SS (Thanassi et al., 2012; Tseng et al., 2009). The classification of pili is independent of the type of secretion system that they are expressed in because some pili that have been thought to belong to a particular secretion system have been observed in other secretion systems. For example; the *Pseudomonas aeruginosa* species has a T2SS that assembles pilus designated as type IV pilus (T4P) (Lu, et al., 1997) and the assembly of this T4P takes place via a mechanism that is closely related to the T2SS pathway (Thanassi et al., 2012). Pili (or fimbriae) have been classified based on many various features including: hemagglutination phenotype and mannose sensitivity (Duguid et al., 1966; Thanassi, 2007), the morphology of the pilus (Duguid et al., 1966; Lawn et al., 1967), assembly pathway of the pilus

(Proft & Baker, 2009), phylogeny and structural predictions of pilin (Girardeau et al., 2000; Wurpel et al., 2013) as well as function carried out by the pilus (Ottow, 1975).

When classified based on their pathway of assembly, one can deduce four main pili groups in gram negative bacteria: a) pili assembled by the chaperone-usher pathway (CU pathway); b) pili assembled by the extracellular nucleation/precipitation pathway (curli pili); c) pili assembled by the "alternative chaperone-usher pathway" (CS1 pilus family) and d) the Type 4 pili (T4P). Each pili group is also further subcategorized (Proft & Baker, 2009). In the CU pathway, the pilus subunits (pilins) bound to specific chaperones are folded and sorted in the periplasm and then transported to the outer membrane usher for pilus assembly (Thanassi et al., 1998). The best characterized pili of this group is the Type I pili of *Enterobacteriaceae* family, which are known for their notable contribution to bladder infections (cystitis) (Connell et al., 1996). The assembly of the curli pili takes place outside of the cell and is initiated by nucleation of soluble curlin subunits that have been secreted into the extracellular media by an outer membrane lipoprotein (Bian & Normark, 1997). Amongst the processes that curli pili are involved are biofilm formation, cell aggregation, cell adhesion and invasion, and inflammatory response induction (Proft & Baker, 2009). The CS1 pili are assembled by the alternative chaperone usher pathway in which specific chaperones, coupled separately to major and minor pilin subunits, transport the pilins to the outer membrane and displace each other at the base initiated fiber growth to feed the assembly with the pilins (Sakellaris & Scott, 1998). The CS1 pilus is known to be associated with the enterotoxigenic *E. coli* that is a major cause of human diarrhoeal disease and significant mortality among infants and young children in developing countries (Qadri et al., 2005).

1.6.1. Type IV Pili and Properties

The assembly mechanism of type IV pili (T4P) is known to be an energy dependent process that involves various T2SS related proteins (Audette & Hazes, 2007; Remaut & Waksman, 2004). T4P are anchored in the membrane and are known to contribute to multi-virulence properties of many gram-negative bacteria including *Neisseria gonorrhoeae* and *Neisseria meningitidis* (Merz & So, 2000) and pathogenesis in species such as *Vibrio cholera* (Taylo et al., 1987), *E. coli* (Girón et al., 1991), and *Salmonella typhi* (Zhang et al., 2000). The T4P also assemble in the potential bioterrorism agent *Francisella tularensis* (Gil et al., 2004) and the *C. burnetii* species which is known for both bioterrorism and Q-fever properties (Beeckman & Vanrompay, 2010; Moodie et al., 2008). Resolving the structure of the major pilin (PilA or "cbu0156") of the pilus system encoded by *C. burnetii* is the second objective in this research project. Other functions that T4P are astonishingly involved include bacterial motility, surface adhesion, microcolony formation, biofilm formation, natural transformation, and immune escape (Craig et al., 2004).

T4P are polymers of a solely or predominantly single monomeric subunit that is called the type IV pilin (Craig et al., 2003). T4P have flexible rope-like structures with a diameter of about 6 nm and a length of several micrometers (Audette & Hazes, 2007). Studies up to date have resolved the X-ray structure of several models of pilins including PilE of *N. gonorrhoeae* (Craig et al., 2003), PAK of *P. aeruginosa* strain K (Hazes et al., 2000), the toxin co-regulated pilin (TcpA) of *V. cholera* (Craig et al., 2003), the minor PilX of *N. meningitidis* (Helaine et al., 2007), and the

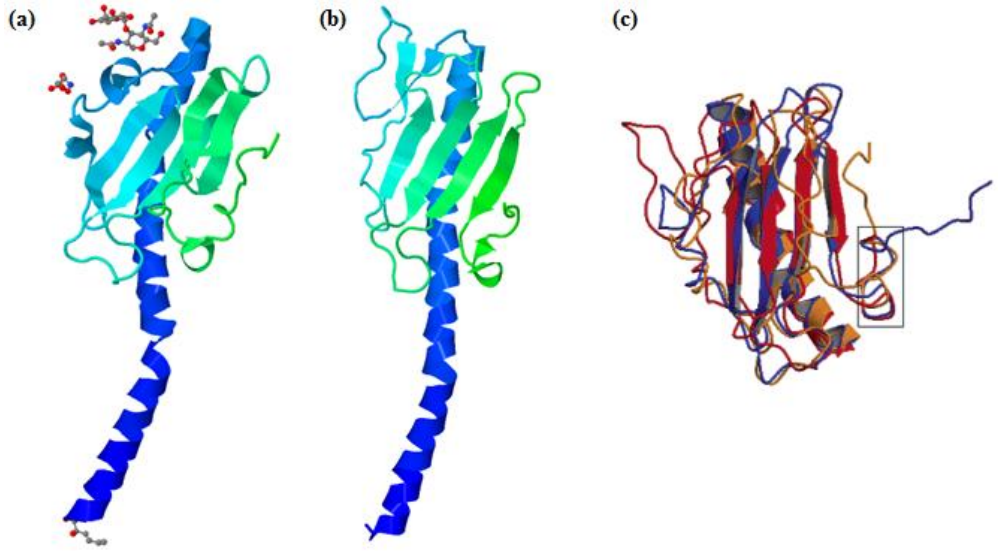


Figure 8: Schematic representation of PilE and PAK. The common structural features of (a) PilE (PDB ID 2HI2) and (b) PAK (PDB ID 1OQW) are the N-terminal α-helix (blue), the antiparallel β-sheets and the connecting loop region (green-turquoise). The image in (c) highlights the conserved structural architecture of the type IVa pilins by superimposing the truncated ΔK122-4 (blue), ΔPAK (red), and PilE (orange) pilins (Audette et al., 2004).

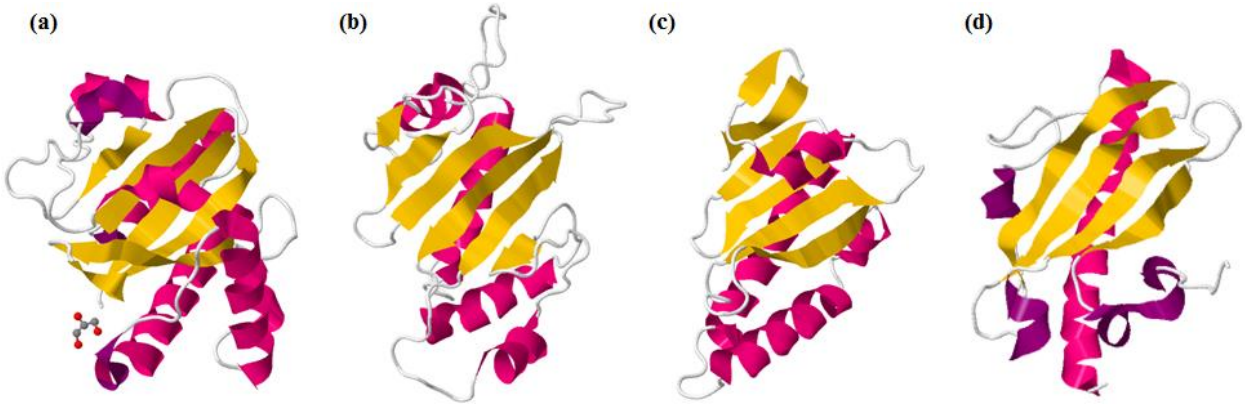


Figure 9: Comparison of pilin structures. Depicted in the above figure are (a) the TcpA pilin (PDB ID 1OQV), (b) truncated PilS (PDB ID 1Q5F), (c) PilX (PDB ID 2OPE) and (d) the truncated ΔK122-4 pilin (PDB ID 1QVE) (retrieved from PDB using JSmol).

PilS of *S. typhi* (Xu et al., 2004) (**Figure 8** and **Figure 9**). The pilins of *V. cholera*, *S. typhi*, and *N. meningitidis* are a sub-type b of T4P (Type IVb) whereas the *P. aeruginosa* and *N. gonorrhoeae*

pilins are sub-type a (Type IVa) due their variation in size and sequence (Craig et al., 2004). Furthermore, during X-ray crystallography and nanotechnology studies, Audette and colleagues designed a truncated version of pilin K122-4 from *P. aeruginosa* strain K122-4 (ΔK122-4) (**Figure 9** and **Figure 10**), by deleting the α1-N region of the N-terminal α-helix (Audette et al., 2004). This increases the solubility of the pilin and makes it easier to crystallize. The superimposition of the truncated PAK pilin (ΔPAK) with a truncated version of pilin K122-4 (ΔK122-4) as well as the PilE (MS11) of *N. gonorrhoeae* (**Figure 9**) gives notable evidence for the conservation of sequence and structure T4P between different organisms (Audette et al., 2004), whether in their truncated forms or native. Based on the model of the PAK polymer, an assembly of the ΔK122-4 pilin into a T4P has been proposed (**Figure 10**) (Audette & Hazes, 2007; Petrov et al., 2013) which was supported by monomer-dimer equilibrium studies (Lento et al., 2016). Audette and colleagues used this assembly model of ΔK122-4 fiber-like structure to develop protein-based nanotubes (PNTs) that can be used in production of nanodevices (Audette & Hazes, 2007). PNTs can be produced under much more feasible conditions that carbon nanotubes (CNTs), which usually require extreme conditions during production (Zhu et al., 2002). Compared to CNTs, the PNT production also provides richer structural and functional architectures that can be created (Audette et al., 2004).

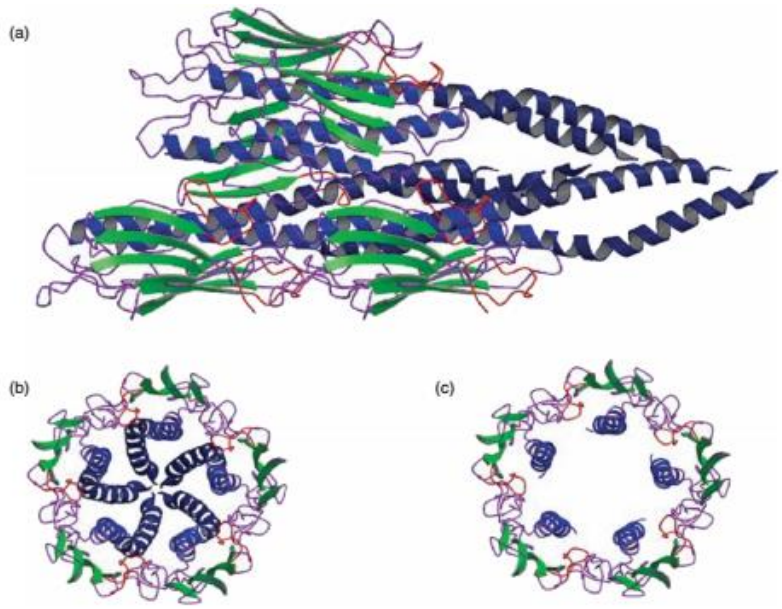


Figure 10: PAK pilin based PNT assembly model. Image (b) is the axial view of the T4P from (a). Image (c) is the same axial view of T4P but based on the assembly of the truncated PAK pilin ΔK122-4 monomer (adapted from Audette & Hazes, 2007).

1.6.2. *Coxiella burnetii* pilin and Research Objective II

As discussed previously, the role of T4P in the virulence of various bacteria as well as its use in biotechnology has been studied for many years and significant progress has been made. Excluded from this progress is the understanding of the role of T4P of *C. burnetii* which requires more attention for treatment of Q fever. Q fever is often caused by the transmission of *C. burnetii* bacterium via aerosolized soil and animal products, a process called zoonosis. The infection process begins with the docking of the bacterium onto the host adhesion receptor $\alpha_v\beta_3$ integrin inducing phagocytosis of the *C. burnetii* into the macrophage. The nascent *Coxiella*-containing vacuole (CCV) acidifies to pH 5.4 and bacteria begins to replicate and express its proteins within the phagolysosome. The CCV then starts to enlarge by fusing heterotypically with autophagic,

endocytic and lysosome vesicles and eventually takes up most of the space in the macrophage within 8 days (van Schaik et al., 2013).

C. burnetii is also known to be internalized by non-phagocytic cells types (Beare et al., 2011) and the infectious property of *C. burnetii* within the host system is highly aided the ability of its T4SS to facilitate material exchange. However, an understanding of the initial host binding of *C. burnetii* followed by the phagocytic induction during pathogenesis in non-phagocytic cells is still incomplete (van Schaik et al., 2013). The role of the T4P in the adhesion and motility of the bacterium as well its interactions with the host membranes within the phagolysosome needs more exploration. As part of our research work aimed towards supporting advancements in the field of bionanotechnology as well as characterization of the T4SSs of some of the most virulent bacterial species, I dedicate a portion of my studies towards crystallization trials of the *C. burnetii* major pilin PilA (designated as cbu0156). Cbu0156 is 140 aa protein with a theoretical mass of 14.8 kDa (Beare et al., 2009). A truncated version of cbu0156 excluding the first 28 amino acids of the N-terminus has been cloned into a pETSUMO vector, giving a 114 aa polypeptide with a theoretical mass of 11.7 kDa. The expression of cbu0156 in pETSUMO vector produces a fusion protein linked to Small Ubiquitin-like modifier (SUMO) which increases the solubility of the protein and makes purification process easy due to its 6 Histidine (6xHis) tag. We furthermore attempted to engineer the SUMO-tag cleaving protease, Ulp1 from *Saccharomyces cerevisiae*, into a plasmid system with a 6xHis tag to produce and purify this enzyme in our lab and utilize it to cleave the SUMO tag from the cbu0156 pilin. Ulp1 is a 621 amino acid long polypeptide chain that possesses a SUMO-specific protease activity within the region between amino acids 403 to 621. This catalytic region of Ulp1 is more specifically mapped to amino acid position 489 (Mossessova & Lima, 2000).

2. EXPERIMENTAL PROCEDURES

2.1. Reagents and Equipment

All chemicals, bacterial antibiotics used in this project were either purchased from Sigma, Thermo Fisher or Bio Basics, unless otherwise stated. The restriction enzymes, PCR (Polymerase Chain Reaction) reagents and DNA ladders were from New England BioLabs (NEB). Primers used for PCR amplification of all genes were purchased from IDT (Integrated DNA Technologies). The *E. coli* BL21 (DE3), Rosetta and DH5α as well as plasmids pBAD24, pET28a and pET26b were from laboratory stocks. The BL21 (DE3) pETSUMO:cbu0156 cells were also obtained from laboratory stocks (originally graciously provided by Dr. Erin van Schaik). The plasmid pET26b was purchased from Novagen. A list of plasmids, bacterial strains, and primers used in this study can be found in **Table 2** and **Table 3**. All cloning experiments were confirmed using Sanger sequencing (BioBasic Inc.) for correct framing and ligates. Protein purification experiments were performed using either the Fast Protein Liquid Chromatography (FPLC) on an Äkta Purifier P10 (GE Healthcare) or the manual gravity column purification technique. Resins for affinity purifications were also from GE Healthcare. Determination of protein concentration was done using bicinchoninic acid assay (BCA) (Thermo Scientific; cat#:23225).

Table 2: A list of primers used in this study.

CONSTRUCT (Primer Names)	PRIMERS	RESTRICTION ENZYMES
pET28a::G*	*For: 5’ CCATGGCAGGCAGTGTGG 3’	NcoI &
(FEP001 & 002)	*Rev: 5’ *CCAAAACAGCC*AAGCTTTT 3’	HindIII
pET28a::G*His	*For: 5’ CCATGGCAGGCAGTGTGG 3’	NcoI &
(FEP001 & 003)	Rev: 5’ CGAAGCTTTTCTTTATGCTGGTAACTC 3’	HindIII
pET26b::G*His	*For: 5’ CCATGGCAGGCAGTGTGG 3’	NcoI &
(FEP001 & 003)	Rev: 5’ CGAAGCTTTTCTTTATGCTGGTAACTC 3’	HindIII
pET28a::Ulp1His	For: 5’ GCGTCCATGGCCTCAGTTGAAGTAGATAAGC 3’	NcoI &
(IOP003 & 004)	Rev: 5’ GCTTCTCGAGTTTTAAAGCGTCGGTTAAAATC 3’	XhoI

*This reverse primer was designed such that it anneals to the pBAD24 plasmid sequence (*italicized*) that is flanking TraG* from its 3' end. The forward primer was same as that used in (Audette et al., 2007)

Table 3: Cell strains and plasmid constructs used in this study.

BACTERIAL STRAIN/ PLASMID	CHARACTERISTICS /DESCRIPTION	SELECTIVE MARKER^	REFERENCE
Bacterial Strains			
DH5α	F⁻ Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1	-	(Woodcock et al., 1989)
BL21 (DE3)	F⁻ *ompT hsdS$_B$ (r_B^- m_B^-) gal dcm* (DE3)	-	(Miroux & Walker, 1996)
Rosetta (DE3) pLysS	F⁻ *ompT hsdS$_B$ (r_B^- m_B^-) gal dcm, lacY1* (DE3) pLysSRARE[6]	Cm	(Baca & Hol, 2000)
Vectors and Constructs			
pET28a	5.4 cloning vector derived from pBR322 containing a N and C-terminal His Tag	Km	Novagen
pET26b	5.4 cloning vector derived from pBR322 possessing an N pelB leader sequence and C-terminal His Tag	Km	Novagen
pBAD24::G*	C-terminal half of TraG* in pBAD24	Amp	(Audette et al., 2007)
pET28a::G*	Untagged TraG* protein	Km	This study
pET28a::G*His	C-terminal 6xHis tagged TraG* protein	Km	This study
pET26b::G*His	N-terminal pelB (periplasmic localization) and C-terminal 6xHis tagged TraG* protein	Km	This study
pET28a::Ulp1His	C-terminal 6xHis tagged Ulp1 protease	Km	This study

^Cm, chloramphenicol; Km, kanamycin; Amp, ampicillin

2.2. Bacterial Growth Conditions and Plasmid Constructs

2.2.1. Bacterial Culturing and Protein Expression

All overnight (O/N) bacterial cultures were grown in an Erlenmeyer flask containing 20 mL of sterile LB broth, 10 mM Glucose and appropriate amount of antibiotics where needed. The final concentration of antibiotics Ampicillin (Amp) and Kanamycin used for growth in liquid

culture was 50 µg/mL and for Chloramphenicol (Cm) was 20 µg/mL. The concentration of Km and Cm used on agar plates was the same, 50 µg/mL and 20 µg/mL respectively, and for Amp was 100 µg/mL. Bacteria were obtained from either glycerol stocks or colonies on agar plates. Cultures were incubated with shaking at 200 rpm overnight at 37°C (~16-18 hours). Glycerol stocks of bacterial strains were prepared by mixing 500 µL of O/N culture with 500 µL of 100% glycerol in cryo-vial tubes and stored at -80°C.

2.2.2. Cloning of TraG* into pET Plasmids

2.2.2.1. PCR amplification and Restriction Enzyme Digestion

Three different constructs of TraG* were generated: a.) TraG* fused to a periplasmic localization sequence (pelB) at the N-terminal and 6xHis tag C-terminal (designated as pelBTraG*-His) using the pET26b vector; b.) TraG* fused only to a C-terminal 6xHis tag (designated as TraG*-His) using the pET28a vector and; c.) TraG* cloned into pET28a vector with a stop codon to prevent tagging at the end of its sequence (designated as TraG*). The cloning of TraG* for all the three constructs was performed by amplifying *traG** from its template nucleotide sequence in pBAD24::G* plasmid (recombinant pBAD24 with TraG* gene generated by (Audette et al., 2007)) along with their corresponding primers (**Table 2**) using Phusion polymerase (cat#E0553S). Correct amplicon sizes were confirmed using DNA gel electrophoresis as per Section 2.2.2.2. The amplicons were then purified using the PCR purification kit. The purified amplicons (1µg each) carrying restriction site overhangs along with their corresponding vectors (1µg) were double digested for 1hr at 37 °C using restriction enzymes (REs) HindIII (1 µL; cat# R0104L) and NcoI (0.5µL; cat# R3193S) with NEBuffer 2.1 in a 50µL total reaction volume. A

single digest with each enzyme was also performed in a similar fashion using the plasmid vectors only. The RE digests were heat inactivated for 20 mins at 80 °C. Samples can be stored at -20 °C.

2.2.2.2. Agarose Gel Electrophoresis

Agarose gels were made by heating 1.2 g of agarose in 100 mL 1X TAE in a 250 mL Erlenmeyer flask until the agarose dissolves and boils. The solution was cooled for 3 mins at room temperature and swirled after adding 2 µL of 10 mg/mL Ethidium Bromide. Each of the digestion reaction from Section 2.2.2.1 was mixed with 0.2 volumes of loading dye and run on agarose gel electrophoresis barely submerged in 1X TAE at 45-50V (4.5-5V/cm) until dye front is ¾ of the gel (65 mins). Bands corresponding to the inserts and the pET vectors were cut-out under UV and extracted using gel extraction kit (Thermo Scientific; cat#K0691). Samples were stored at -20°C.

2.2.2.3. Preparation of Chemically Competent Cells

Cells were grown O/N as in Section 2.2.1. From the O/N culture, 1 mL was aliquoted into 100 mL of fresh sterile LB broth, with corresponding antibiotic if applicable. The cells were shaken at 37 °C until the optical density reading at 600 nm wavelength (OD_{600}) reached a mid-log absorbency of 0.4-0.7 range. The cells were placed on ice for 10 mins and all relevant material such as pipettes tips and microfuge tubes to be used were placed at 4 °C or on ice. The cell culture was the transferred into two pre-chilled falcon tubes and centrifuged at 2700 xg for 10 mins at 4 °C. The supernatant was discarded and the pellet was gently resuspended with 1.6 mL of ice cold 100 mM $CaCl_2$ while on ice for 30 minutes. The resuspension was centrifuged again at 2700 xg

for 10 mins at 4°C and the supernatant was discarded. The cell pellet was again gently resuspended with 1.6 mL of ice cold 100 mM $CaCl_2$ while on ice for 20 minutes. The cell resuspensions were combined into one tube and mixed with 0.5 mL ice-cold 80% v/v glycerol. The chemically competent cells were stored in 220 µL aliquots in microfuge tubes at -80 °C.

2.2.2.4. Ligation and Heat Shock Transformation

A total of 100 ng volume of DNA from gel extraction (3:1 insert:vector ratio (using NEB ligation calculator)) was added to 2 µL of 10X T4 Ligase Buffer (brought to room temperature) and double distilled water (ddH_2O) to a final 20 µL reaction volume. After adding 1 µL of T4 DNA Ligase (NEB M0202S), the samples were gently mixed and incubated for 30 mins at room temperature. A negative control was also set up for each reaction using the vectors to test for self-ligation. Samples were heat inactivated at 65°C for 10 mins and then placed on ice.

Chemically competent DH5α cells prepared in Section 2.2.2.3 were thawed on ice. To 100 µL of chemically competent cells, 15 µL of the ligation reaction was added and incubated on ice for 20 mins. The cells were directly transferred to a 42 °C water bath and heat shocked for 90 seconds. Subsequently, the sample was placed back on ice for 5 min and then diluted with 900 µL of sterile LB broth. The negative control is treated the same way. The samples were then incubated for 1.5 hrs at 37 °C in a 125 rpm shaker. After incubation, 100 µL of each sample of transformed cells was antiseptically spread on an LB agar plate containing selective antibiotic(s). The plates were placed upside-down in a 37 °C incubator O/N. A single colony of transformed DH5α cells from the experimental plate was grown in LB broth containing the appropriate antibiotic(s), as per Section 2.2.1 and glycerol stocks from the O/N culture of the new recombinant cells were made.

In situations where slight growth was observed on the negative control plates due to self-ligation of vector, 3-4 colonies from the experimental plates were grown O/N to increase the chance of selecting the desired transformant cell line.

From the O/N culture of the cloned DH5α cells, plasmid miniprep of the new the construct was done using a plasmid miniprep kit (Thermo Scientific; cat#K0502). The constructs were confirmed for correct ligation using restriction mapping with the same enzymes used for double digestion (Section 2.2.2.1) and were checked for correct translational framing using Sanger sequencing (at the BioBasic sequencing facility). Once confirmed, 1 µg the pET plasmid gene from the plasmid miniprep above was transformed into 50 µL of chemically competent BL21 (DE3) cells, using the heat shock protocol described above.

2.2.3. Cloning of Ulp1 into pET28a Plasmid

Ulp1 was cloned into the pET28a plasmid with a C-terminal 6xHis tag. Cloning was performed by amplifying the *ulp1* nucleotide sequence with corresponding primers (**Table 2**) and using the *S. cerevisiae* Δsrb10 strain genomic DNA template, obtained from the Rosonina Lab at York University. The parameters for PCR amplification, RE digestion, ligation and transformation were same as those described in Section 2.2.2. The restriction enzymes used for cloning were NcoI HF (1 µL) and XhoI (1 µL) with NEBuffer 2.1. Furthermore, because the Ulp1 is a eukaryotic protein, there was a chance that its mRNA transcript may possess the rare codons that the prokaryotic BL21 (DE3) cells may not be able to read. In order to account for this, I transformed the plasmid constructs carrying *ulp1* into chemically competent *E. coli* Rosetta cells (obtained from lab stocks), which are enhanced BL21 cells designed to express eukaryotic proteins.

2.3. Protein Expression and Purification

2.3.1. Protein Expression

Small scale expression of all proteins was performed to monitor and optimize at temperatures 25 °C, 30 °C, and 37 °C. To 100 mL of sterile LB broth in Erlenmeyer flasks (1 flask for each temperature), 1 mL Glucose (from 1M stock), an appropriate amount of antibiotics (Section 2.2.1) and 1 mL of an O/N bacterial culture of BL21 (DE3) or Rosetta cells possessing a plasmid construct were antiseptically added. The cultures were incubated 37 °C with shaking at 200 rpm until the OD_{600} absorbency reached the mid-log phase (0.5-0.7). The cultures were cooled to room temperature and induced for expression with 50 µL of 1 M isopropyl-β-D-1-thiogalactopyranoside (IPTG; to a final concentration of 0.5 mM). The cultures were then incubated for 4 hrs at their designated temperature conditions (25 °C, 30 °C, and 37 °C) with shaking at 200 rpm. Aliquots of 1 mL were collected from each culture at every hour beginning from 0 hrs. The aliquots were centrifuge in a microfuge tube for 1 min at 13000 rpm and stored at -20°C for analysis on SDS PAGE (Section 2.3.3). After 4 hrs, a total cell pellet was obtained from rest of the cells in the culture in the 100 mL expression volume by centrifuging for 20 min at 3300 xg in 4 °C and storing at -20 °C. These pellets were used for testing binding of the 6xHis-tagged proteins to the Nickel-Immobilized Metal Affinity Chromatography (Ni-IMAC) column.

The large scale expression of proteins was performed at their optimal temperature of expression deduced from the small scale expression. This was determined to be 37 °C for 3 hrs for all proteins. Large scale expressions were set up by adding 10 mL of Glucose (1M stock), appropriate amount of antibiotic (Section 2.2.1) and 10 mL of the O/N bacterial culture aseptically to 1 L of sterile LB broth in an Erlenmeyer flask. Cells were grown at 37 °C with shaking at 200 rpm until mid-log phase (0.5-0.7) and expression was induced with 500 µL of 1M IPTG.

Expression cultures were incubated for 3 hrs at 37 °C with shaking at 200 rpm. The 1L expression volumes were centrifuged for 20 min at 3300 xg at 4 °C and the pellets were stored at -20 °C.

2.3.2. Sonication, Osmotic Shock and Purification of Proteins

2.3.2.1. Cell Lysis by Sonication and Purification

Sonication of the small scale protein expression was done by dissolving the pellet in 10 mL of IMAC Loading Buffer (LB) (200 mM NaCl, 20 mM Tris, pH 7.5) and large scale expression were done in the same manner except the amount of lysis buffer used to resuspend and dissolve the cell pellet was 40 mL in a 50 mL conical tube. A volume of 100 mM phenylmethylsulfonyl fluoride (PMSF; dissolved in 100% Isopropanol) was added to the dissolved pellet to a final concentration of 1.25 mM PMSF and mixed. The tube was packed tightly inside an ice bucket and was positioned such that the tip of the sonicator was at the midway point of the pellet solution in all directions. Cells were sonicated at amplitude 30% with 15s ON and 15s OFF for 5 mins. The lysed cells were spun at 25 000 xg for 30 mins in 4 °C. The supernatant from the lysate was saved and mixed with an appropriate volume of 100% glycerol a final concentration of 5% v/v glycerol.

2.3.2.1.1. Purification of 6xHis Tagged Cytoplasmic Proteins Using FPLC

The purification of all the 6xHis tagged cytoplasmically expressed proteins, namely; TraG*-His, cbu0156-SUMO-His and Ulp1-His was performed on the Ni-IMAC column by using the fast protein liquid chromatography (FPLC) technique. The supernatant obtained from the last step of Section 2.3.2.1 was loaded directly on the Äkta P10 device for facilitated purification by a preset software program.

2.3.2.1.2. Purification of TraG* by Gravity Column Technique

The purification of the untagged TraG* protein was performed using gravity column technique. The Ni-IMAC resin was pre-equilibrated with 3 column volumes (CV) of IMAC-LB. The supernatant obtained from the last step of Section 2.3.2.1 was loaded onto the column containing resin and incubated on a rocking platform in a 4 °C room for 15-30 mins depending on amount of lysate. A sample of flow through fraction was collected while the is put to vertical position to flow and settle. Once the resin settles, 2 CV of IMAC-LB was used to gently wash the column and a sample of the wash fraction was collected. The resin was washed with IMAC-LB for an additional 7-9 CV or until OD_{280} reached a steady state near zero absorbency. Care was taken not to disturb the resin from its settled position at all times. The column was then treated with 3 CV of IMAC- EB (Elution buffer; 80 mM Imidazole, 200 mM NaCl, 20 mM Tris, 5% v/v glycerol, pH 7.5) and protein elution fractions of 4-5 mL were collected. Elution fractions were saved for SDS PAGE analysis (Section 2.3.3) and the resin was cleaned with 3 CV of buffer containing 1M Imidazole, 200 mM NaCl, 20 mM Tris, pH 7.5, followed by 4-5 CV of 20% ethanol. Resin was left stored in 20% ethanol in the column.

2.3.2.1.3. Dialysis and Anion Exchange Purification of TraG*

Only the elution fractions of the untagged TraG* from gravity purification protocol (Section 2.3.2.1.1) and the cbu0156-SUMO-His samples were run under a secondary purification protocol. The elution fractions of TraG* from Section 2.3.2.1.1 were analyzed on SDS PAGE (as per Section 2.3.3) and combined into a single tube. A strip of 10 kDa molecular weight cut-off (MWCO) dialysis membrane (Spectrum Labs) was placed in ddH_2O for 15-20 mins at 4 °C and

the tubing was gently opened while submerged in ddH_2O. The combined TraG* fractions were poured into the dialysis tube and the tubing was incubated in 2.5 L of dialysis buffer (DB) containing 5 mM NaCl, 20 mM Tris, pH 7.5 for 2 hrs at 4 °C. After 2 hrs, the dialysis tube was placed in fresh 2.5 L of DB and incubated O/N at 4°C. The dialyzed protein sample was then loaded directly on the Äkta P10 device for facilitated anion exchange purification by a preset protocol on the software program.

2.3.2.2. Osmotic Shock and Purification of Periplasmic TraG*-His

After expressing periplasmic TraG*-His (pelBTraG*-His) for 3 hrs as per Section 2.3.1, the 1 L large expression volume was divided up into two 0.5 L volumes and placed inside two 1 L centrifugation bottles. A 1 mL aliquot of the expression was saved for SDS PAGE analysis of expression based on total cell fraction (TCF). The cells in the 1 L bottles were centrifuged for 20 mins at 3300 xg in 4 °C, the supernatant was discarded and each pellet was gently resuspended with 80 mL of ice cold Tris Wash solution (TW; 20 mM Tris, 20% w/v Sucrose, 1 mM EDTA, pH 7.4), while on ice. The resuspensions were transferred to separate 250 mL pre-cooled centrifuge bottles and incubated on ice for 20 mins. The samples were spun at 7000 xg for 20 mins in 4°C and a 100 µL aliquot of supernatant fraction (TWF) was saved for SDS PAGE analysis. The rest of supernatant was discarded and each pellet was gently resusupended in 80 mL of ice cold Osmotic Shock (OS; 5 mM $MgSO_4$, pH 7.4) solution while on ice. The samples were incubated on ice for 20 mins and then centrifuged at 35000 xg for 20 mins in 4 °C. The supernatant which now constituted the periplasmic fraction (PPF) was saved and the pellet which constituted the cytoplasmic fraction (CPF) was discarded after saving a small portion of it for SDS PAGE analysis.

The PPF fraction (~160 mL) was then loaded directly onto the Äkta Purifier P10 device for FPLC Ni-IMAC purification under software control.

2.3.3. SDS PAGE Analysis

Analysis of protein expression, purity and concentration was performed using 12.5% SDS PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) by mixing 15 µL of protein samples with 10 µL of 2X SDS loading dye **(**100 mM Tris-Cl pH 6.8, 4% w/v electrophoresis grade SDS, 0.2% w/v bromophenol blue, 20% v/v glycerol, 200 mM β-mercaptoethanol) or 100 µL for 1 mL cell pellet and boiling the samples for 5 mins. The analysis of protein dimerization and oligomerization was performed using both 12.5% and 7.5% pseudo-native SDS PAGE simply by mixing 10 µL of protein samples with 10 µL of 2X Native loading dye (120 mM Tris-Cl pH 6.8, 0.2% w/v bromophenol blue, 20% v/v glycerol) and loading directly onto the gel. The recipe used for making 4 of SDS PAGE gels was as follows; 6.93 mL ddH_2O (8.93 mL for 7.5% gel), 5.2 mL of 40% w/v acrylamide (3.2 mL for 7.5% gel), 4.2 mL of 1.5 M Tris-Cl pH 8.8, 166.5 µL of 10% w/v SDS, 166.5 µL of 10% w/v ammonium persulfate (APS) and 16.7 µL of tetramethylethylenediamine (TEMED). The 4% gel stacking was made with the mixture of 3.16 mL of ddH_2O, 499.5 µL of 40% w/v acrylamide, 1.265 mL of 1 M Tris-Cl pH 6.8, 50 µL of 10% w/v SDS, 50 µL of 10% w/v APS and 5 µL of TEMED. The 12.5% SDS PAGE gels were run for 40 mins at 220V in Tris-Glycine while the 7.5% were run for 30mins under the same conditions. All gels were stained in gel staining solution (1 g of Coomassie Brilliant Blue dissolved in a 1 L solution containing 50% v/v methanol, 10% v/v glacial acetic acid and 40% v/v ddH_2O) for 20

mins, destained in gel destaining solution (40% v/v methanol, 10% v/v glacial acetic acid, 50% v/v ddH_2O) for 20 mins destain and then stored in water O/N for improved destaining.

2.4. Protein Concentration and Crystallization

Protein samples were concentrated using centrifugal concentrators at 4 °C. Crystallization screening experiments were done using Microlytics MCSG Core I-IV screening kit in 96 well plates. Drops contained 1-1.25 µL of purified and concentrated protein with equal amounts of reservoir solution equilibrated over 100 µL of reservoir. Optimization trials of the crystal hits on the initial screen were conducted on 24 well plates by optimizing the crystallization conditions in the initial hits. Crystallization experiments were observed using a Nikon SMZ1500-Fiber Lite MI-150 microscope; preliminary diffraction analysis of crystals was performed on a Rigaku 007HF MicroMax Rotating Anode diffractometer equipped with a Saturn 944+ CCD Detector.

2.5. Electrospray Ionization Mass Spectrometry Experiments

Protein samples of concentration 1-2 mg/mL were first dialyzed into 200 mM ammonium acetate buffer O/N at 4 °C using 20 kDa MWCO (for TraG*) and 10 kDa MWCO (for cbu0156) dialysis cassettes (Slide-A-Lyzer Dialysis Cassettes; Thermo Scientific). Samples were analyzed by mass spectrometry using both a batch-load mode and liquid chromatography coupled to MS (LC-MS) mode. LC-MS experiments were performed using the Agilent 1200 series quaternary pump and Phenomenex BioSep-SEC-s4000 column connected to the Synapt G2-Si High Definition Mass Spectrometer (Waters).

2.6. Dynamic Light Scattering Experiments

Size exclusion chromatography (SEC) coupled to Multi-Angle Light Scattering (MALS) experiments were performed on Äkta P10 and Wyatt WTC030S5 size exclusion column connected to Dawn Heleos II and Optilab T-rEX instruments (Wyatt Technologies). Protein samples were dialyzed O/N at 4 °C using 20 kDa cutoff dialysis cassettes and their loading concentrations ranged from 0.5 mg/mL to 4 mg/mL. Treatment of samples with glycerol was done by adding 100% glycerol to the protein sample and allowing it to sit at 4 °C for at least 1 hr.

3. RESULTS AND DISCUSSION

3.1. TraG* Experiments

3.1.1. Periplasmic Expression of TraG*

The C-terminal region of the TraG protein is known to reside in the periplasm while the N-terminal is anchored at the inner membrane (N Firth & Skurray, 1992). The ideal conditions under which a native structure of the C-terminal half of TraG (designated as TraG*) can be obtained is the periplasmic environment of the cell. Hence, an initial attempt was made to express TraG*, have it localize to the periplasm of the cell for native folding and perform crystallization and structural studies of the protein extracted from the periplasm.

3.1.1.1. Cloning of TraG* into pET26b using RE digestion and Expression of pelBTraG*-His

An initial attempt was made to generate a fusion of TraG* that would localize to the periplasm once expressed. This was done by cloning *traG** gene from pBAD24::G* (recombinant pBAD24 with TraG* gene generated by (Audette et al., 2007)) in *E. coli* XK1200 cells into the pET26b vector using restriction enzyme (RE) digestion with HindIII (cut site flanking 3' end of *traG**) and NcoI (cut site flanking 5' end of *traG**). The pET26b::G* construct was successfully generated and transformed into both *E. coli* DH5α and BL21 (DE3) cells (**Figure 11**). This construct expresses TraG* with an N-terminally fused pelB periplasmic localization signal (PLS) and a C-terminal His-tag (**Figure 12**). Expression trials of pelBTraG*-His were performed at various temperatures using IPTG as the inducer and ideal expression conditions were observed when inducing at 37°C for 3 hrs (**Figure 13**). Osmotic shock experiments were performed to

extract periplasmic fractions form BL21 (DE3) cells and the amount of TraG* localized in the periplasm (PPF) was observed to be very little, while most of it resided in the cytoplasmic fraction (CPF) (**Figure 13**).

3.1.1.2. Cloning of TraG* into pET26b using PCR and Subsequent Expression Trials

A closer investigation on the original pBAD24::G* construct from Audette et al., 2007 indicated that the construct had already possessed a pelB leader sequence and a detailed investigation tracing the NcoI enzymatic cut sites revealed that the pBAD24::G* construct actually contained two NcoI cut sites: one at the 5' end of *traG** that has been introduced with the forward primer that was used by the Audette group and a second NcoI site that was carried along when they were cloning the *pelB-traG** from pET26b into the pBAD24 vector, (see (Audette et al., 2007)). Hence in theory, our RE based recombinants pET26b::G* in Section 3.1.1.1 were generated using mixed inserts of *traG** and *pelB-traG** which will then result in a double pelB leader sequence that is fused to TraG* in the pET26b::G* clones. Under the skepticism that the double pelB leader sequence may have hindered the localization of pelBTraG*-His into the periplasm and that the extra PLS sequence may cause bias in the native crystal structure, I decided to PCR amplify *traG** gene from the pBAD24::G* plasmid and clone the amplicon into pET26b using the amplicon. Primers marking the target *traG** sequence in pBAD24::G* were designed (**Table 2**) and used in PCR. Thus, the amplified the *traG** from pBAD24::G* would lack the *pelB* sequence and its ligation between the HindIII and NcoI sites on pET26b gives a single pelB leader sequence. Expression trials of pelBTraG*-His were performed using the newly generated pET26b::G* and expression profiles were analyzed on SDS PAGE. The obtained results from the

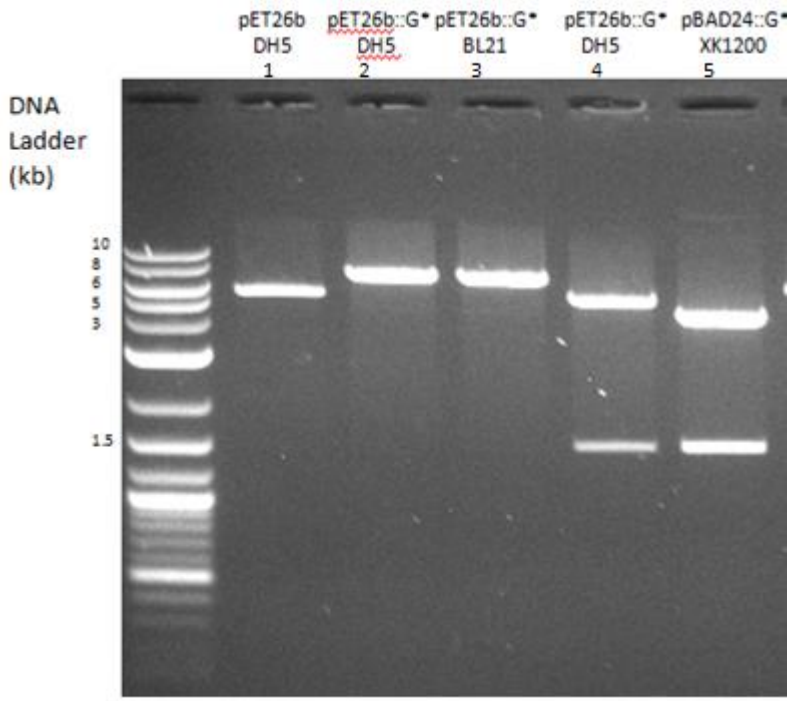


Figure 11: Evidence of successful cloning and transformation of pET26b::G* into DH5α and BL21 cells. All single bands are a product of the corresponding plasmid with single HindIII digestion. pET26b::G* from newly transformed DH5α cells and pBAD24::G* from XK1200 were double digested with HindIII+NcoI to confirm successful *traG** cloning from pBAD24::G* into pET26b::G* (double band wells #4 and #5).

Figure 12: Depicted above is a map of the pET26b system showing RE sites (red boxes) and pelB leader sequence flanking the target cloning site with the 6xHis tag at the 3' end (green arrows).

expression and purification of the periplasmic TraG*-His were not much distinct from those presented in **Figure 13**. The low profile expression of periplasmic TraG*-His made it quite difficult to perform crystallographic experiments which often require protein samples at high concentrations in large amounts.

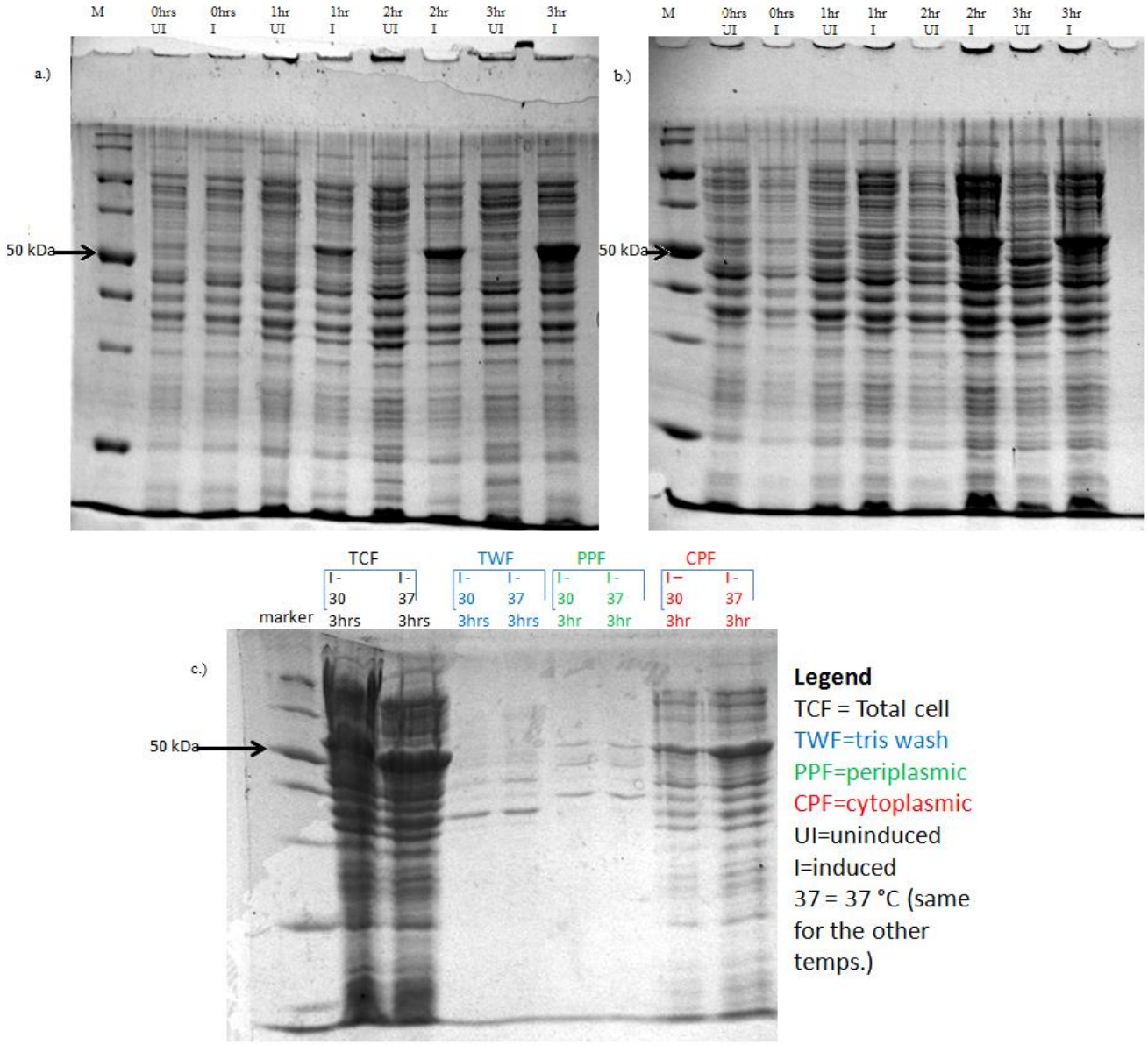


Figure 13: Induction of pET26b::G* in BL21 cells with 0.5 mM final IPTG concentration. (a) Inductions for 3hrs at 37 °C, and (b) for 3hrs at 30 °C. The various cell fractions collected during osmotic shock for both inductions are depicted in (c). The TraG* is a 52.7 kDa protein with a 6xHis fusion tag which adds another 1kDa, making the protein 53.7 kDa in weight.

3.1.2. Cytoplasmic Expression of TraG*

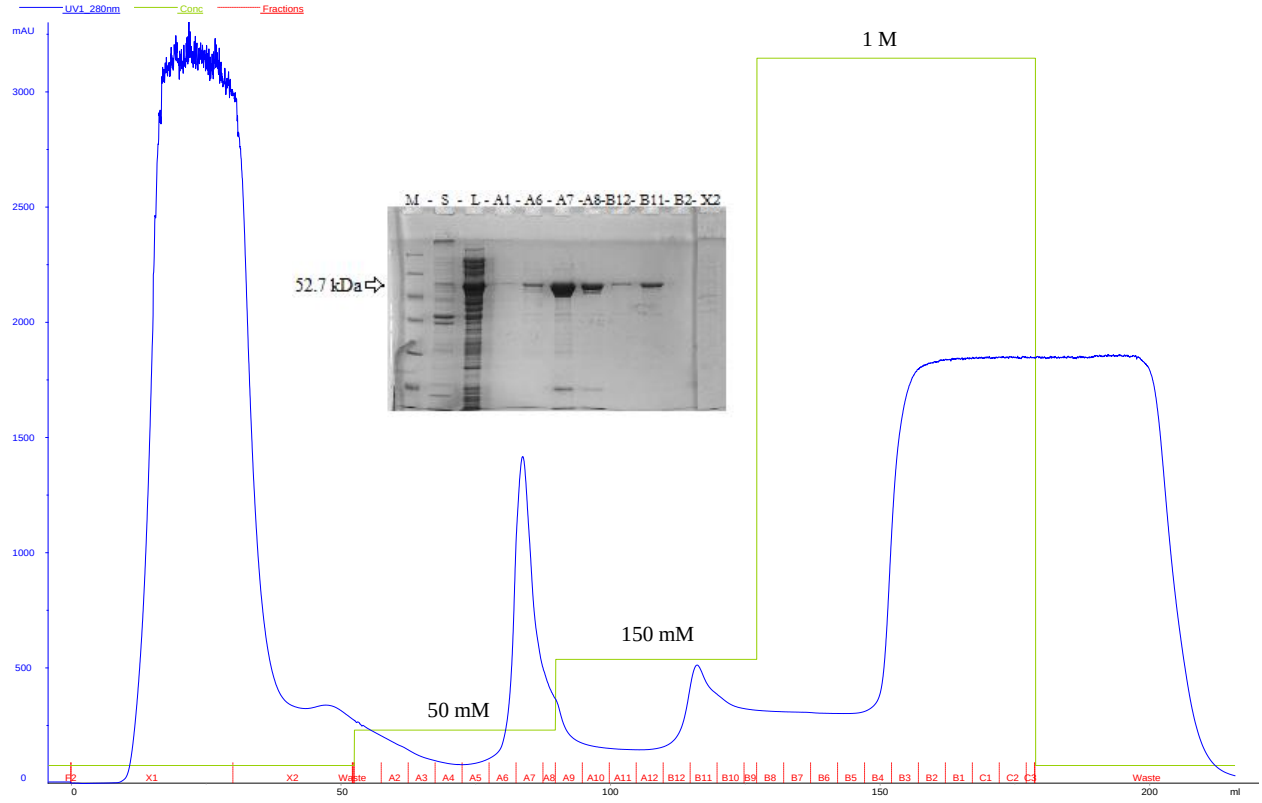
3.1.2.1. TraG* purification and crystallization

In order to increase protein yield during purification, TraG* was extracted from the whole cell rather than the periplasmic fraction. The cytoplasmic expression of TraG* was achieved using the construct generated by cloning its PCR amplicon from pBAD24::G* into the pET28a plasmid. The pET28a possess the same qualities as the pET26b except that it lacks the *pelB* leader sequence. Hence, the expressed TraG* polypeptides in the pET28a transformed BL21(DE3) cells were devoid of the PLS at their N-terminus. Using primers FEP001, FEP002 and FEP003 (**Table 2**), *traG** was cloned to generate both a 6xHis-tag (TraG*-His) fusion and a non-tagged version of the protein (designated as TraG* from this point).

3.1.2.1.1. Discovery of untagged TraG* binding on to nickel metal ions

Initially, both the untagged (TraG*) and the 6xHis tagged clones (TraG*-His) were tested for binding to the nickel metal ions using FPLC-IMAC purification. Unsurprisingly, only TraG*-His was observed to bind to the nickel column while TraG* would flow-through with minimal or no binding under liquid pressure (**Figure 14**). However, when purified using a gravity column (GC) IMAC purification technique, the untagged protein showed significant binding (**Figure 15a**). This is owing to the difference in the experimental steps of the two techniques. In FPLC facilitated IMAC purification, the protein sample is simply "run" through the column using liquid pressure whereas in the GC IMAC purification procedures, the protein sample is mixed with the nickel resin inside a column and then incubated with shaking for 30 minutes at 4 °C for increased binding. The resin is allowed to settle and the liquid is let loose to flow under gravitational pull, exclusively.

a.)



b.)

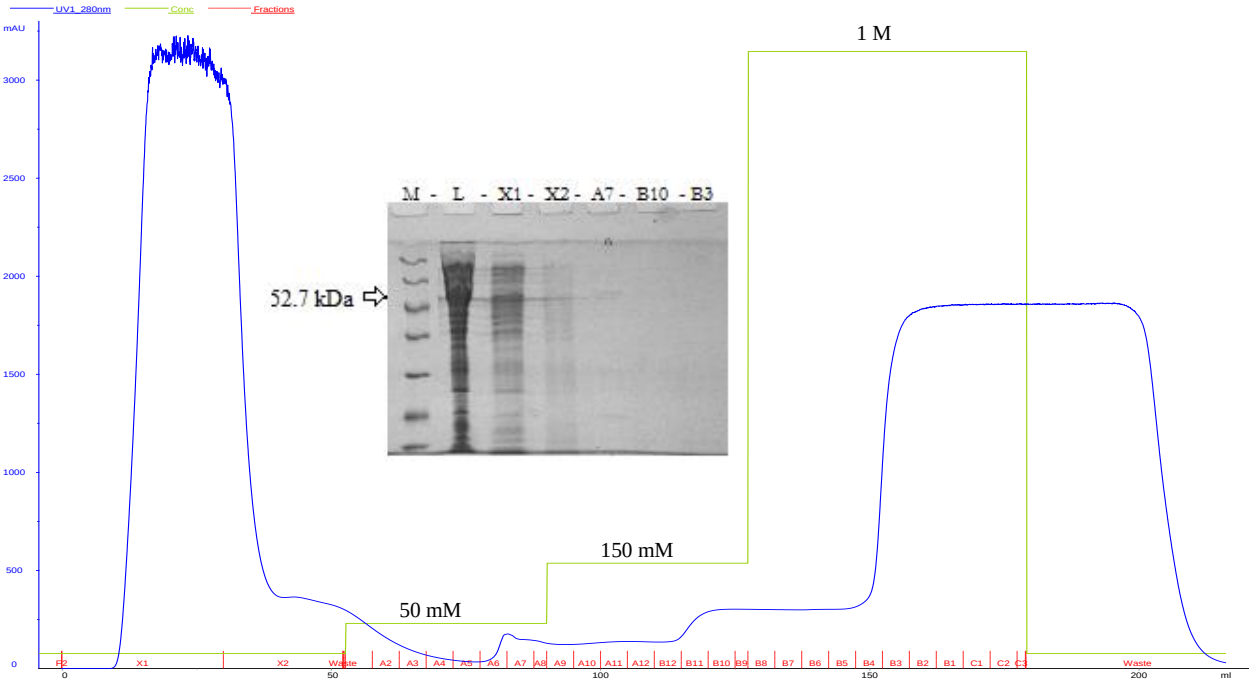


Figure 14: FPLC mediated IMAC purification of TraG*-His in (a) and TraG* in (b). The protocol used a step gradient elution (green line) with buffer containing 200 mM NaCl, 20 mM Tris and Imidazole at pH 7.5. UV280 is shown as the blue line. An SDS-PAGE of some fractions is also shown with corresponding label.

Furthermore, the binding of the untagged TraG* to nickel resin is not surprising given the histidine residue based analysis of its sequence. TraG* (486 aa) contains a total of fourteen (14) histidine residues with two (2) of these residues being adjacent to each other in the sequence. When compared to the polypeptide sequences of other F-conjugation proteins such as TraN (602 aa with 7 histidine residues), TraS (149 aa with 1 histidine residue) and TraF (247 aa with 3 histidine residues) the ratio of histidine residues to total number of amino acids in the polypeptide chain is significantly higher in TraG* (Frost et al., 1994). This potentially increases the number of histidine residues exposed and available to interact with the nickel ions. The binding is relatively weak; hence the protein loses interaction with the nickel resin with buffer containing imidazole concentrations that are as low as 20 mM. This indirectly affects the purity of TraG* elutions from GC IMAC (**Figure 15a**). TraG* was then purified using anion exchange column and then screen for crystallization (**Figure 15b**). TraG*-His did not require a secondary purification since the elution fractions were relatively pure (**Figure 14a**).

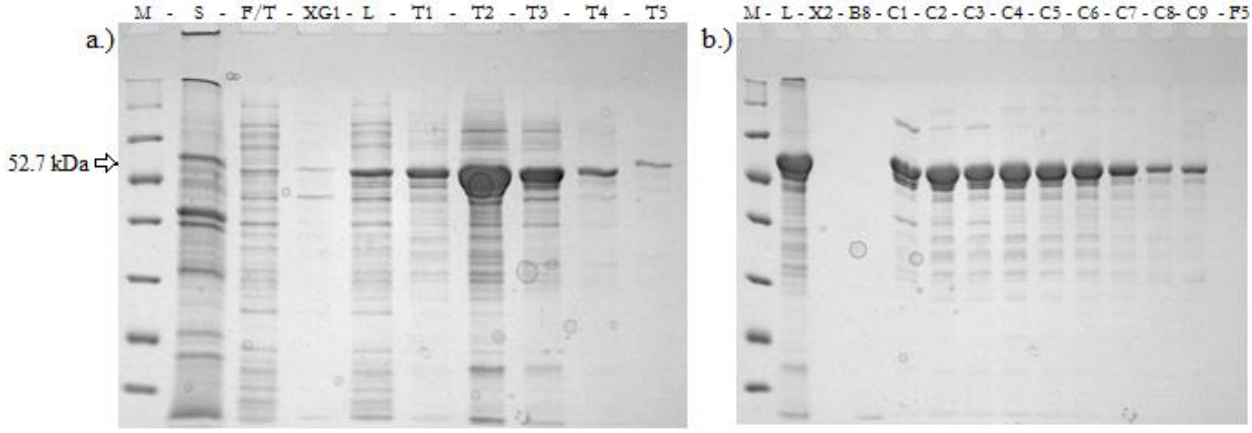


Figure 15: SDS PAGE post-purification analysis of TraG* on (a) gravity column (GC) IMAC and (b) anion exchange (AIEX) column. GC IMAC elutions were obtained with buffer containing 80mM Imidazole, 200 mM NaCl, 20 mM Tris, 5% Glycerol at pH 7.5. Elutions for the AIEX purification were obtained using a linearly increasing gradient of NaCl concentration with 20 mM Tris at pH 9.0. The abbreviations for the fractions are as follows; M = molecular weight marker, S = the post-sonication pellet, F/T = flow through and X/XG = washing steps fraction, L = protein load before purification and B/C/F or T = are the corresponding elution fractions.

3.1.2.1.2. Initial crystal hits with TraG*

Following concentration, TraG* and TraG*-His were screened for crystallization using the MCSG suites I-IV (Microlytic) and various protein concentrations (5-20mg/mL). The protein samples used were further dialyzed into buffers containing variations in native salt and pH levels to increase screening capacity. TraG* crystals formed immediately after setting a plate using the MCSGII screen, in particular under condition "D4" (0.2 M Calcium Acetate, 0.1 M Sodium Acetate: Acetic Acid pH 4.5, 30% (w/v) PEG 400). However, the crystals were unstable and did not last for long (**Figure 16**). The TraG*-His crystal screens also produced a number of crystals under various conditions that were analyzed using the in-house X-ray diffractometer. The diffraction data from many of these crystals were either poor or the data was an indicative of salt crystals.

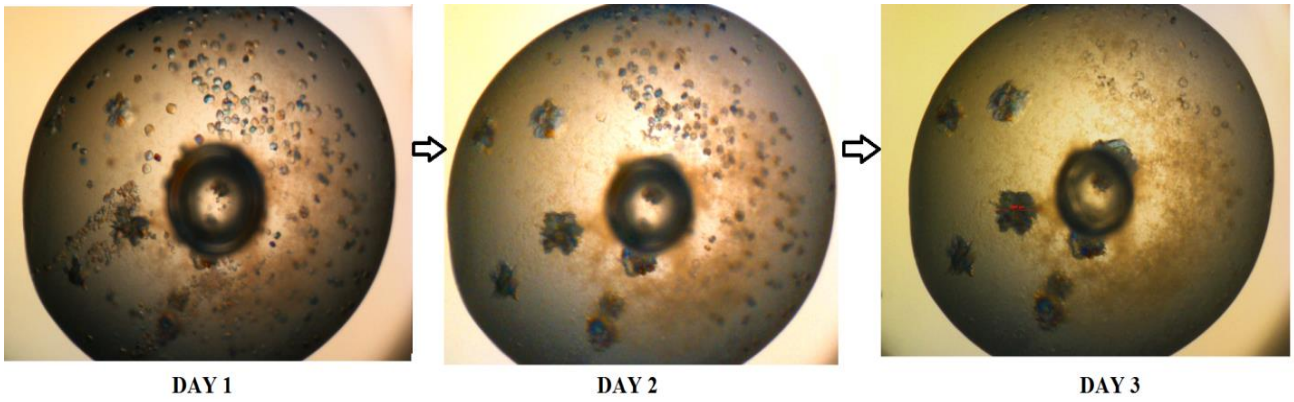


Figure 16: Initial TraG* crystal hits. TraG* crystals in condition D4 (0.2 M Calcium Acetate, 0.1 M Sodium Acetate: Acetic Acid pH 4.5, 30% (w/v) PEG 400) on the MCSGII screening plate. The TraG* samples were prepared by combining fractions C4 to C9 (from AIEX elutions similar to that in **Figure 15b**) and concentrated to ~10 mg/mL. Plates were set up on ice and incubated at room temperature with the 100 µL reservoir solution. Each drop contained 1.2 µL protein sample

(inside buffer with ~275 mM NaCl, 20 mM Tris, 5% Glycerol, at pH 9.0) and 1.2 µL reservoir solution.

3.1.2.1.3. Analysis and Optimization of TraG* Crystals

Despite their insignificant appearance in their final form (**Figure 16**, DAY 3), the TraG* crystals were mounted on the diffractometer and observed for diffraction. Unsurprisingly, the diffraction pattern was not ideal (**Figure 17**). These crystals were however used as micro-seed stocks to help nucleate TraG* crystallization in optimization experiments. In optimization experiments conducted by altering salt, precipitant, temperature and pH condition, showers of crystals were obtained with micro-seeding experiments (**Figure 18**). Many of these crystals are too small to diffract so they are being incubated as they are slowly growing.

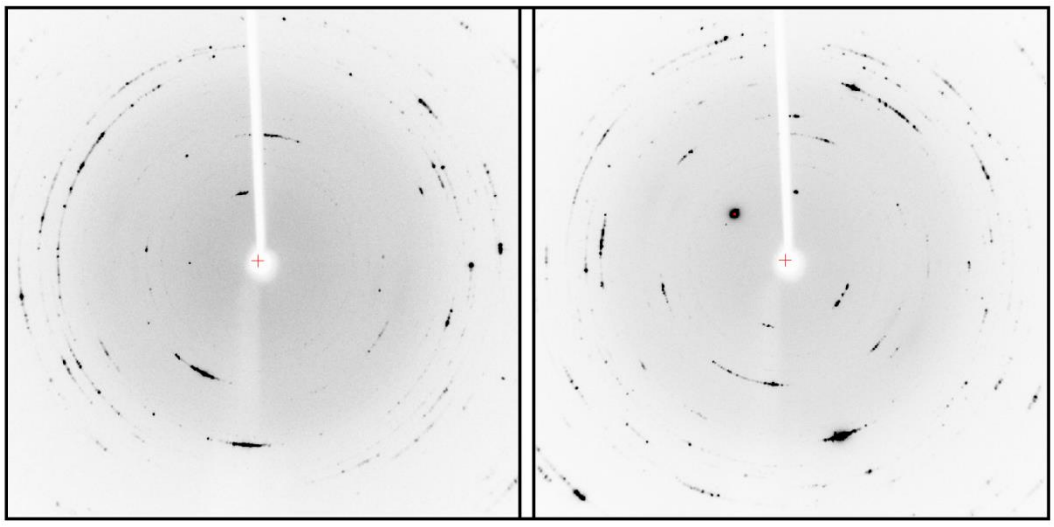
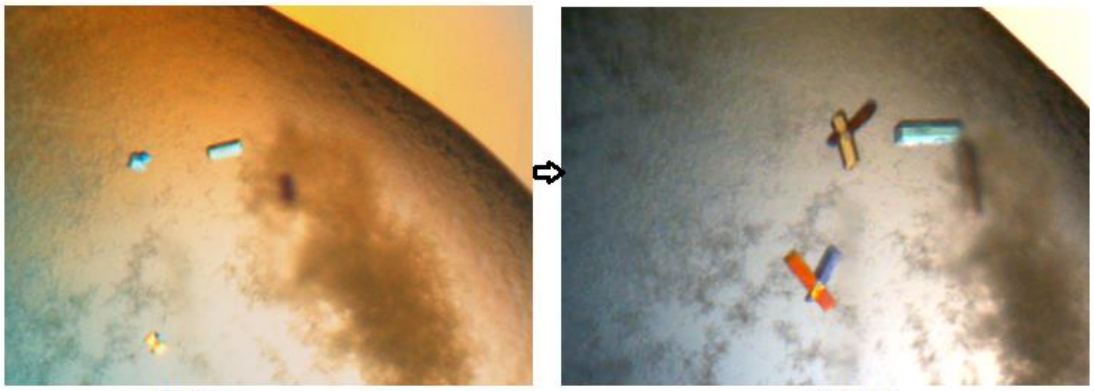


Figure 17: Two diffraction pattern images of TraG* crystal from condition D4 of MCSG suite II (Figure 16, DAY 4).

Figure 18: TraG* crystal showers from optimization of condition D4 of MCSGII. Instead of Calcium acetate, the salts in the conditions included Ammonium sulfate and Lithium sulfate. The pH was altered between 4.0 and 5.0 while the PEG400 concentrations were screened from 18% to 24%. Crystals formed on all plates whether incubated at 4°C or room temperature.

Crystals were also obtained (**Figure 19a**) in other MCSG conditions when existing TraG* crystals (from **Figure 16**, DAY 4) were provided as micro-seeds to nucleate TraG* crystallization. Under condition C8 of MCSGII (0.1 M Bis-Tris Propane pH 7.0, 1.8 M Magnesium Sulfate), crystals appeared a few days after screen set up however, it took more than 2 months for them to grow. These crystals also diffracted poorly (**Figure 19b**). However, condition C8 does not contain any cryo-protectant and hence the lattice structure of the crystal will likely have been significantly distorted under cryo-conditions without any cryo-protection, resulting in poor diffraction.

a)



b)

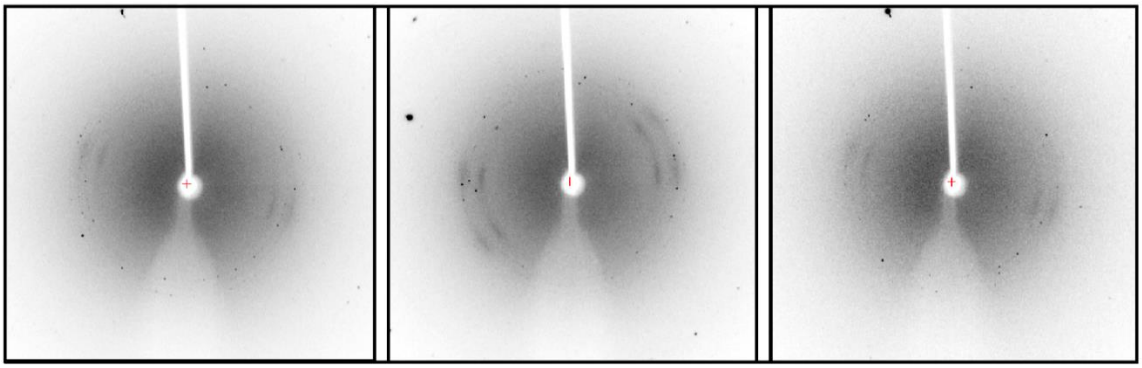


Figure 19: TraG* crystals hits from microseeding. TraG* crystals from condition C8 (0.1 M Bis-Tris Propane pH 7.0, 1.8 M Magnesium Sulfate) on the MCSGII screen. The TraG* samples were prepared by combining fractions C4 to C9 (from AIEX elutions similar to that in Figure 15b) and concentrating them to 13.85 mg/mL. Plates were both set up and incubated at room temperature with the 100 µL reservoir solution. Drops contained 1.25 µL protein (in ~275mM NaCl, 20mM Tris, 5% Glycerol, at pH 9.0), 1.2 µL reservoir solution and micro-seeds of TraG* crystals from condition D4 MCSGII (Figure 16, DAY 4) by dipping a cat whisker carrying seeds into the drop.

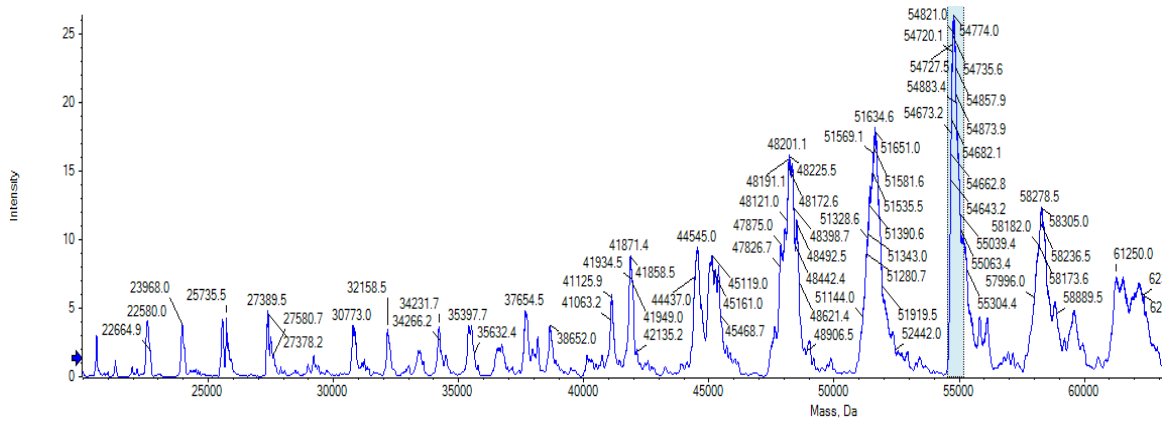
3.1.3. TraG* Protein-Protein Interactions

3.1.3.1. TraG* Mass Spectrometry Data

While the electrospray ionization mass spectrometry (ESI-MS) experiments performed on TraG* and TraG*-His confirmed the correct size of the proteins (**Figure 20**), the obtained data also indicated existence of TraG* dimers (**Figure 21**). The MS data shows that a molecule with a

size in the range of 100-110 kDa exists in the solution. The intensity of this peak which is interpreted to be a dimer is less than 30% of the intensity monomer peak.

a)



b)

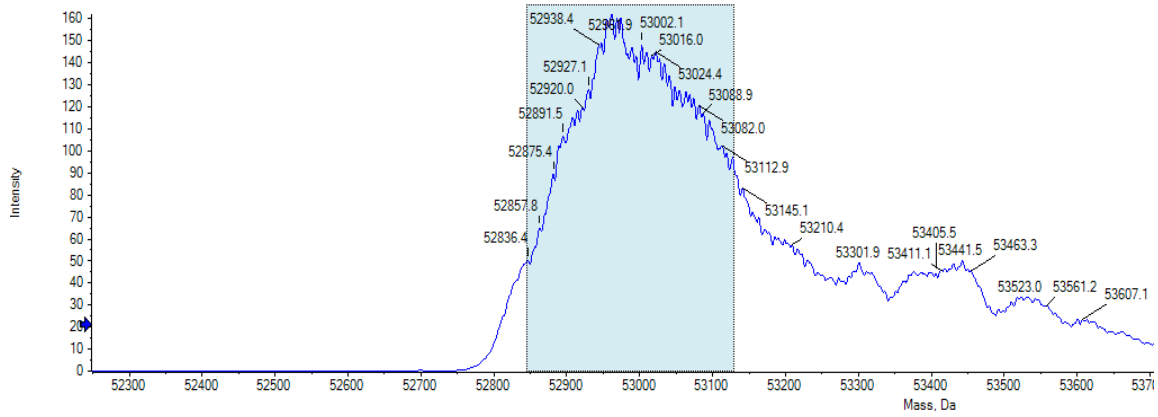
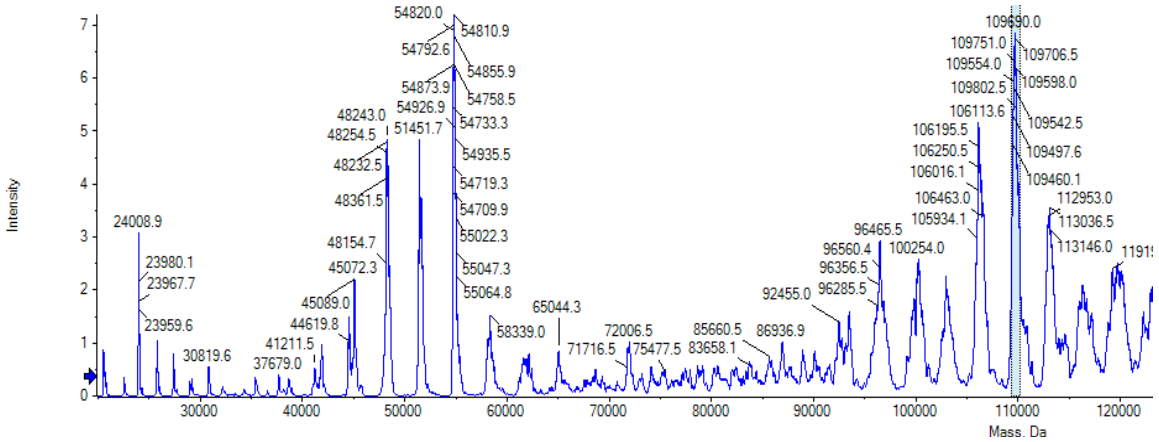


Figure 20: Mass spectrometry results of monomers. a.) TraG*-His and b.) TraG* samples (each approximately 1 mg/mL) were dialyzed into 200 mM Ammonium acetate overnight. The samples were analyzed on SDS PAGE gel for purity and then run on MS. The peaks corresponding to the monomers are highlighted in blue. TraG*-His has a slightly larger size than TraG* due to the 6xHis tag and the additional 6-7 amino acids residing between the 6xHis tag and the HindIII cut site on the pET28a plasmid, where the 3' of gene ligates (similar to **Figure 12**).

a)



b)

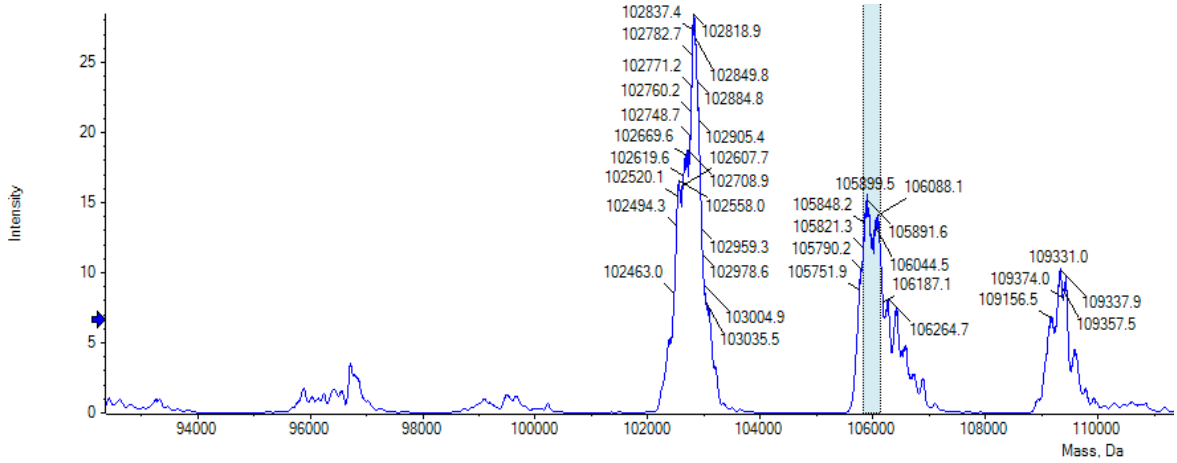


Figure 21: Mass spectrometry results of oligomers. a.) TraG*-His and b.) TraG* dimers observed in their respective batch mode MS experiments presented in **Figure 20**. The peaks corresponding to the dimers are highlighted in blue.

The LC-MS data on the other hand did not give conclusive results to elucidate the actual masses of species coming off as part of the various TraG* peaks from the size exclusion liquid chromatography column.

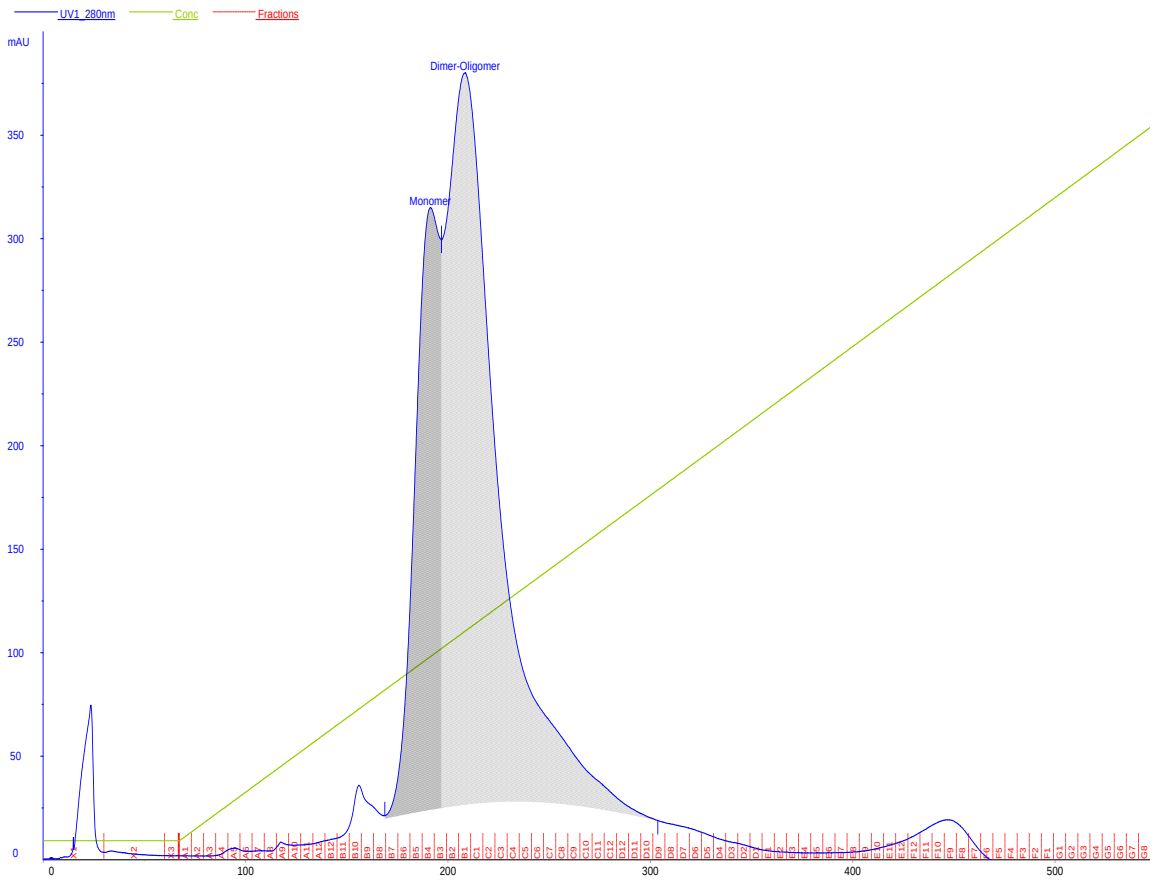


Figure 22: AIEX chromatogram of TraG*. The shaded regions of the bimodal peak correspond to the monomer peak (shaded with argyle) which makes up 27.8% of the total peak area while the dimer/oligomer fractions (shaded with stripes) make-up 72.2%. The green line corresponds to NaCl concentration increasing from 0 to 1M.

3.1.3.2. TraG* Pseudo-Native SDS PAGE Results

TraG* was observed to form oligomers in elution fractions of various purification methods. In AIEX experiments performed with an elongated elution step protocol (20 column volume elution instead of the routine 10 column volume elution step, **Figure 22**), TraG* was observed to exist in monomer, dimer as well as higher oligomer forms (**Figure 23**). While most of the elution fractions contained all three forms of TraG*, the bimodal peak in the elution profile indicates the elution of two distinct species. A closer inspection of the pseudo-native gel shows that the first peak collected

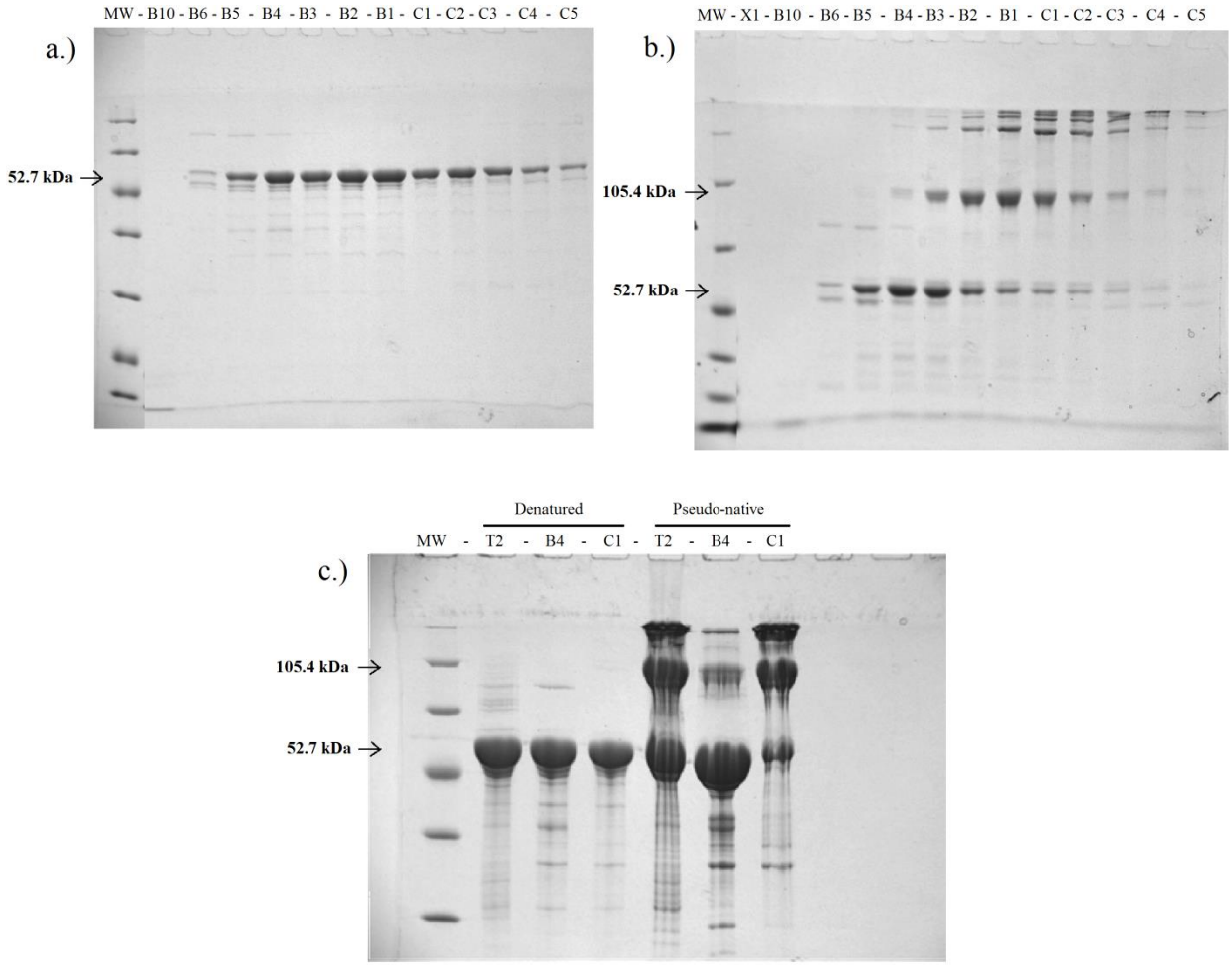


Figure 23: SDS PAGE of TraG* AIEX fractions. a.) TraG* fractions from the elution peaks presented in **Figure 22** were denatured and run on 12.5% SDS PAGE gel. The gel image in b.) shows the same samples run on 7.5% pseudo-native SDS PAGE gel. In c.), the AIEX fractions B4 and C1 were concentrated to 14.87 mg/mL and 8.2 mg/mL respectively, and run on a gel using both denaturing and pseudo-native conditions. Sample T2 (15.2 mg/mL) is a TraG* the elution fraction from GC-IMAC before purification on AIEX chromatography. MW= molecular weight marker and X1 = wash step fraction.

by fractions B6-B4 contains the monomer while the larger second peak makes up dimer and oligomer forms. One fraction (B4) from the monomer peak and one from the dimer/oligomer peak was concentrated along with a TraG* fraction from GC IMAC to observe the effect of protein

concentration on protein-protein interactions. The gel depicted in **Figure 23c** shows that upon concentrating the B4 TraG* monomer fraction, the amount of dimer or oligomer forms resulting from TraG*-TraG* interactions does not significantly increase. Furthermore, the intensity of all the bands observed in fraction C1 increase proportionately upon sample concentration. This result makes crystallization of TraG* monomer easier since existence of dimer, trimer or tetrameric forms of TraG* in solution can act as "contaminants" to prevent crystal stacking of single protein species in solution (refer to APPENDIX for more details).

The existence of dimers and oligomers in some of the AIEX fractions and their absence in others is still concerning. If there exists a TraG* monomer-dimer equilibrium in solution, then this should be present across all fractions. It is possible that there are two TraG* monomeric species that exist in solution, one properly folded and another mis-folded. The folding of the polypeptide chain into a non-native tertiary structure can be a byproduct of in-vivo high expression levels. In this sense, the TraG*-TraG* interactions can either be native or an artifact caused by strong interaction between misfolded TraG* species.

However, given the fact that the TraG* homolog VirB6 oligomerizes while functioning as part of the outer membrane complex of the Ti secretion system, (Darbari & Waksman, 2015), there is a high chance that the observed behavior of TraG* in vitro is due to its native property. In order for TraG to reach and interact with the TraS protein located in the inner membrane of the recipient cell to perform its entry exclusion functions (Audette et al., 2007), it is likely that this protein has two states. The first state would be a relaxed state whereby the C-terminal domain of TraG resides in the periplasm in a folded way and potentially allows the protein to oligomerize for stabilization and pilus assembly functions. The second state would be a flexed state of the C-terminal domain of TraG, by unfolding and stretching itself from the periplasm of

the donor cell onto the inner membrane of the recipient cell. In this sense, the two monomeric species of TraG* observed in vitro maybe representing the relaxed and flexed states of TraG C-terminal domain, which could be crucial for its multiple functions.

3.1.3.3. Size Exclusion Chromatography coupled Multi-Angle Light Scattering

In order to investigate effect of salt, glycerol and protein concentration on TraG* protein-protein interaction, SEC coupled MALS (SEC-MALS) experiments were performed. This method is fairly sensitive towards extrinsic molecular transformations. Furthermore, the characterization of TraG* behavior under various conditions is a critical step towards optimizing crystallization conditions. Although coupled, the analysis of SEC and MALS is presented separately below. TraG* samples were dialyzed into native buffer containing either 5 mM, 100 mM or 250 mM NaCl along with 20 mM Tris at pH 7.5. Each dialyzed TraG* sample had a concentration of approximately 4 mg/mL which was then diluted to 1:2 and 1:6 fold. Moreover, an aliquot of each sample varying in salt and protein concentration was mixed with 100% glycerol to final 7.5% glycerol and stored at 4 °C for at least 1hr before loading onto SEC-MALS. The effect of glycerol on TraG*-TraG* interactions is relevant since glycerol was used in protein samples (5% final) to prevent protein degradation during viva spin concentration protocols. It should be noted that when bringing the % v/v glycerol content to 7.5% in the experimental samples, the original 5% added for viva spin concentration was not considered due to the assumption that this amount was lost or brought to negligible levels after dialysis.

3.1.3.3.1. Size Exclusion Chromatography (SEC) Data

The analysis of TraG* protein-protein interaction profiles using SEC is based solely on the 280 nm UV absorbance peaks. The Wyatt HPLC column WTC-030S5 used to perform these experiments was effective only at resolving the TraG* oligomer species from other species in the solution. Hence, the TraG* oligomer UV peak would elute as a single peak whereas the dimer and monomer forms of TraG* would show up under a single peak that was either wide or slightly shouldered, depending on sample condition. The percentage of TraG* oligomer peak area within each run was compared the peak area of the dimer-monomer forms via integration and the results are presented in **Figure 24** to **Figure 27**.

The SEC data indicated that regardless of the amount of NaCl present, there was always an increase in TraG* oligomerization when sample protein concentration is decreased from 4 mg/mL to 0.7 mg/mL incrementally. An increase in the amount of TraG* oligomerization is also enhanced by glycerol content, which is indicated by the increase in the oligomer peak areas at constant protein concentrations. For example, in **Figure 24**, when going from 1 mg/mL to 1 mg/mL with 7.5% glycerol, the percentage of oligomer peak area increases from 2% to 8% while the area taken up by the monomer-dimer peak decrease from 96% to 86%.

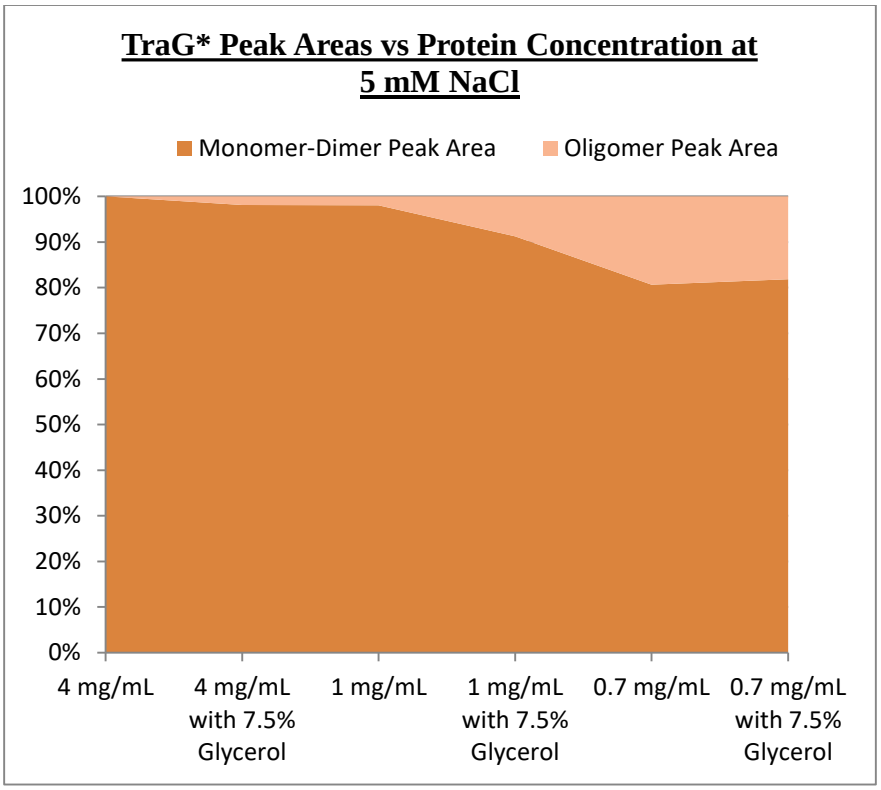


Figure 24: TraG* oligomerization based on protein concentration at 5 mM NaCl. TraG* sample was dialyzed from AIEX fractions into 5 mM NaCl, 20 mM Tris pH 7.5 O/N at 4 °C. The sample was then serially diluted and each sample was partly mixed with glycerol to run as separate experiments. The total peak area obtained in each run is sum of oligomer peak and the single monomer-dimer peak. The graph above figure illustrates the percent area covered by each peak within each respective run.

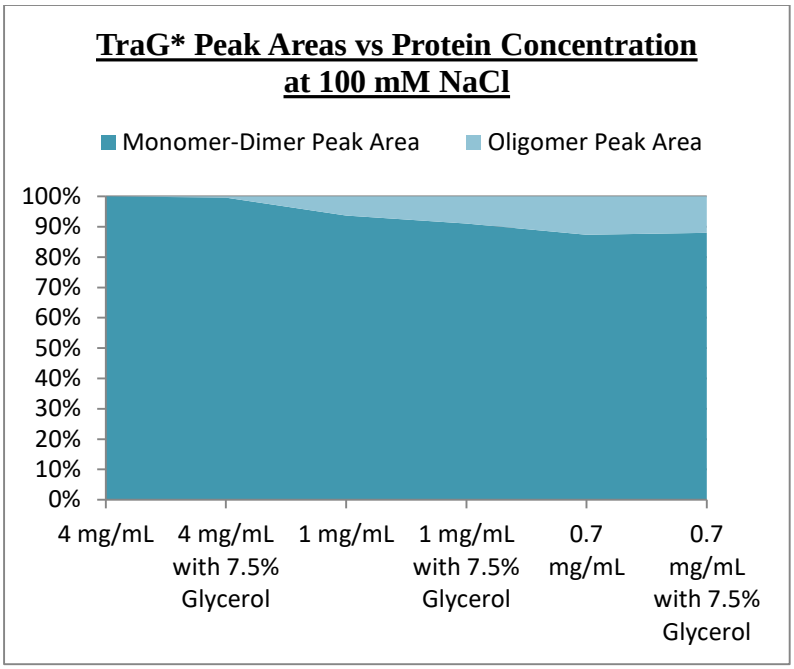


Figure 25: TraG* oligomerization based on protein concentration at 100 mM NaCl. TraG* sample was dialyzed from 5 mM NaCl, 20 mM Tris pH 7.5 into 100 mM NaCl, 20 mM Tris pH 7.5 O/N at 4 °C. See **Figure 24** caption for more details.

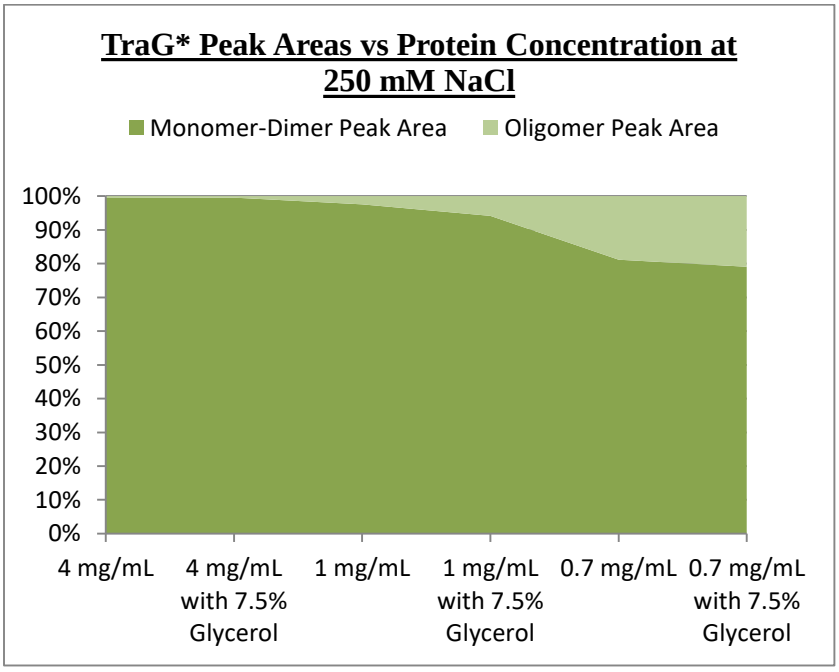


Figure 26: TraG* oligomerization based on protein concentration at 250 mM NaCl. TraG* sample was dialyzed from 100 mM NaCl, 20 mM Tris pH 7.5 into 250 mM NaCl, 20 mM Tris pH 7.5 O/N at 4 °C. See **Figure 24** caption for more details.

The analysis performed on the SEC data to see the effect of salt on TraG* oligomerization was not conclusive **Figure 27**. TraG* at 0.7 mg/mL across various salt concentrations (5 mM, 100 mM and 250 mM NaCl) seem to have oligomers formed at all times.

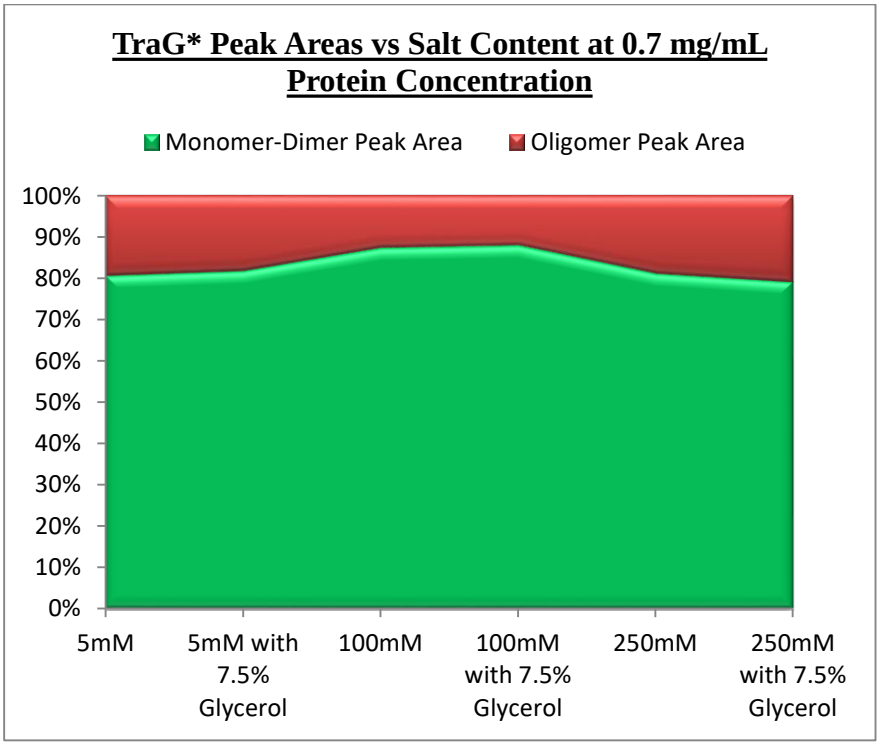


Figure 27: TraG* oligomerization based on salt content. The TraG* data from the 0.7 mg/mL samples across various salt concentrations was pooled and integrated on a single chromatogram. For details, refer to the caption under **Figure 24**.

3.1.3.3.2. Multi-Angle Light Scattering Data (MALS)

Dynamic light scattering (DLS) is known to be extremely sensitive to the presence of aggregates in solution (Bishop, Martin, & Rosenblum, 1991). In structural studies of biological compounds, DLS is used to identify solutions in which macromolecules remain monodisperse during crystal nucleation and growth (Borgstahl, 2007; Niesen et al., 2008). Due to its sensitive nature, the MALS instrument was able to better resolve the TraG* dimer peaks from the monomer as well the oligomer peak. These peaks were resolved due to distinct scattering of light beam and

the detection of particles of specific size eluting with each peak. The results from the MALS experiments showed that both the oligomerization and dimerization of TraG* increases distinctively as one decreases sample protein concentration (**Figure 30**), which support the SEC results discussed previously. This increase in formation of TraG* oligomer and dimer species at various sample protein concentrations does not seem to be effected by salt content (**Figure 28** and **Figure 29**).

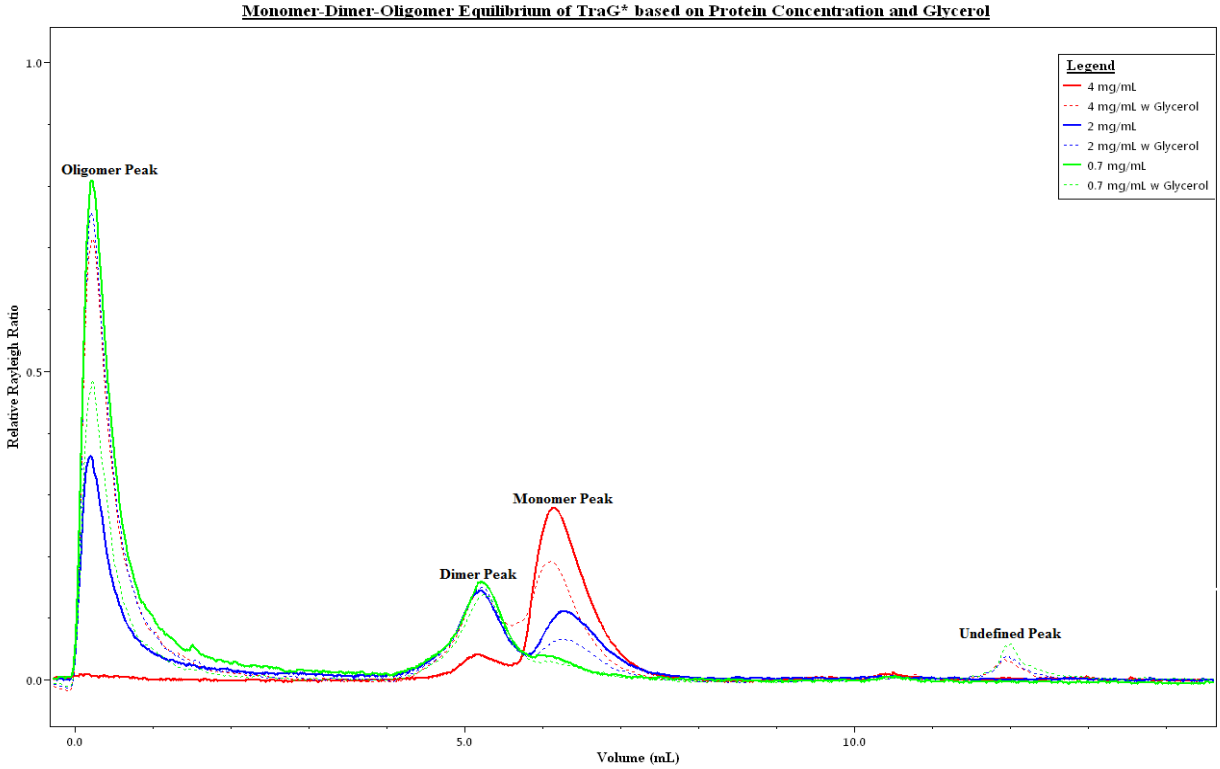


Figure 28: TraG* light scattering with varying protein concentrations at steady 5 mM NaCl. Sample were prepared as described in caption under **Figure 24** and protocols (Section 2.6). The Rayleigh ratio is defined as the ratio of intensities of incident and scattered light. The "undefined peak" seems to be a degradation product of TraG*.

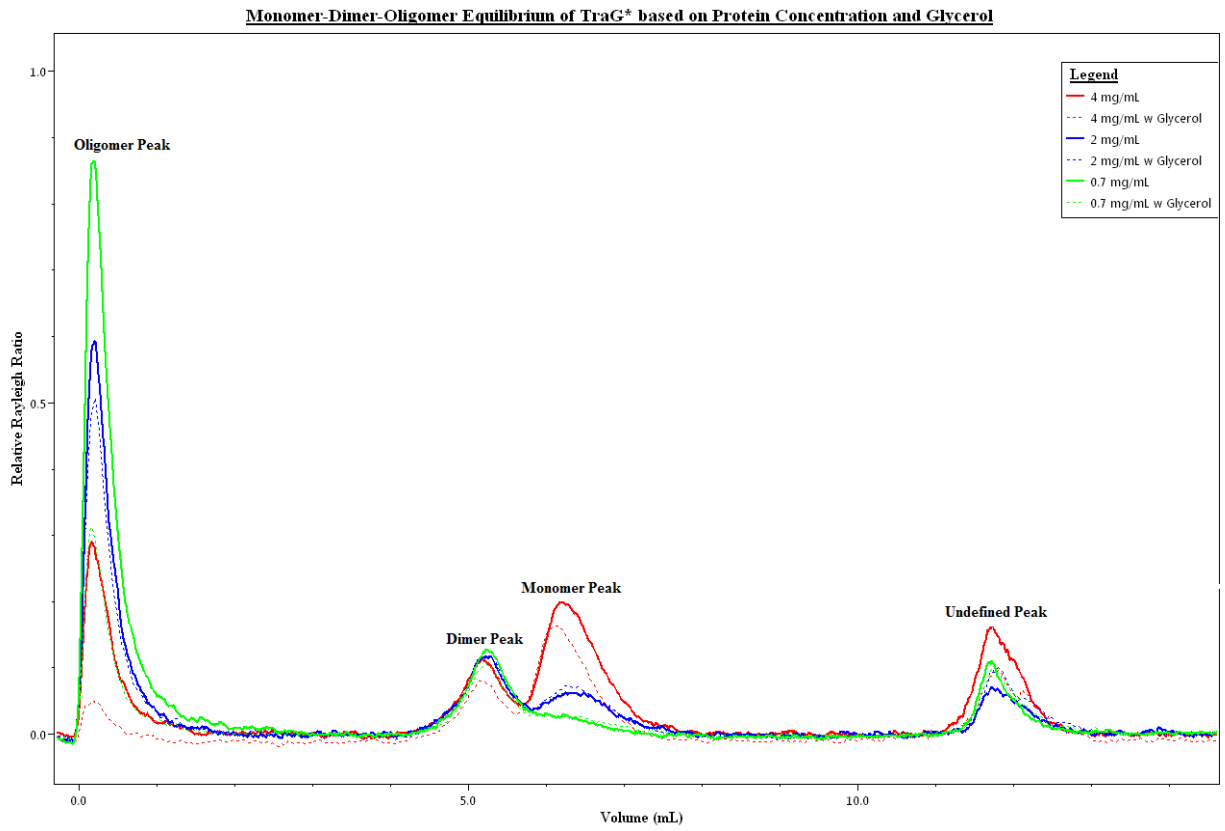


Figure 29: TraG* light scattering with varying protein concentrations at steady 250 mM NaCl. Samples were prepared as described in caption under **Figure 24** and protocols (Section 2.6). The Rayleigh ratio is defined as the ratio of intensities of incident and scattered light. The "undefined peak" seems to be a degradation product of TraG*.

While the effect of glycerol and salt content on the intensity each peak did not reveal a unanimously distinct pattern, presence of glycerol seems to induce the formation of a degradation product (**Figure 28**). Hence, the use of glycerol as a protecting agent during viva spin protein concentration should be reconsidered for optimal preparation of TraG* crystallization samples. This degradation product is also observed in samples with high salt content (250 mM, **Figure 29**) but lacking glycerol. Based on personal experience with TraG*, degradation in these samples without glycerol is likely due to the 4-5-day storage of the protein at 4 °C before running on SEC-MALS. The samples in **Figure 28** were relatively "fresh" and the SEC-MALS experiments on these samples were performed within a period of 24 hrs.

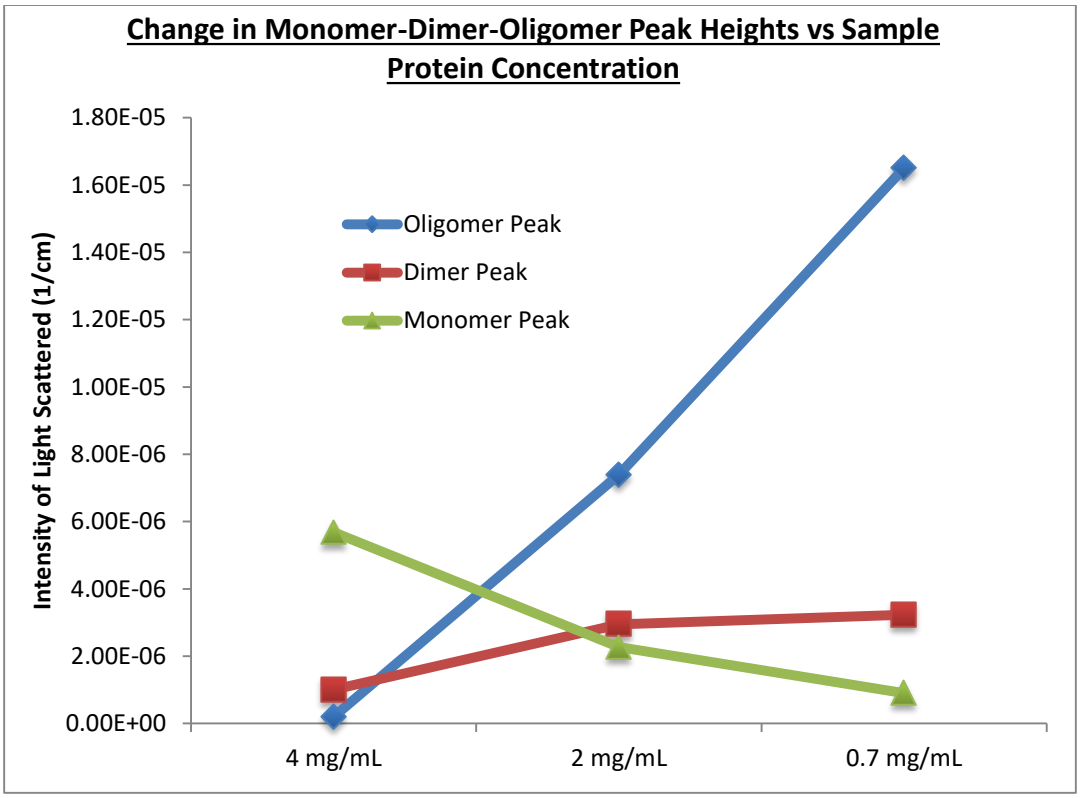


Figure 30: Change in intensities of TraG* peaks. The above data was plotted using the scattered light intensities of each peak at steady 5 mM NaCl with varying protein concentrations.

The average radius of hydration (Rh) data collected for each peak is also presented in **Figure 31**. The Rh value by MALS considers all protrusions from the core body of the protein and computes radius of the particle based on the outer most sphere that can form around a particle (see APPENDIX). According the results (**Figure 31**), the hydration sphere around the TraG* dimer is lower than that of its monomer which may be argued as contradictory. Presence of TraG* dimers even under denaturing SDS PAGE conditions (not shown) proves that the TraG* protein-protein interaction is quite strong. It is possible that the monomer species has more loose chains than dimer species of TraG* which may be more highly ordered and compact with less chains sticking out from the protein core. Since polypeptide sequence of the protein is devoid of any cysteine residues, TraG*-TraG* protein-protein interaction is thought to be resulting from non-covalent bonds such

as H-bonds, ionic and polar interactions. Moreover, this property of the protein may serve as an explanation for the inability of pelBTraG*His to localize to the periplasm due high levels of protein expression aiding protein-protein interactions (see Section 3.1.1.2.).

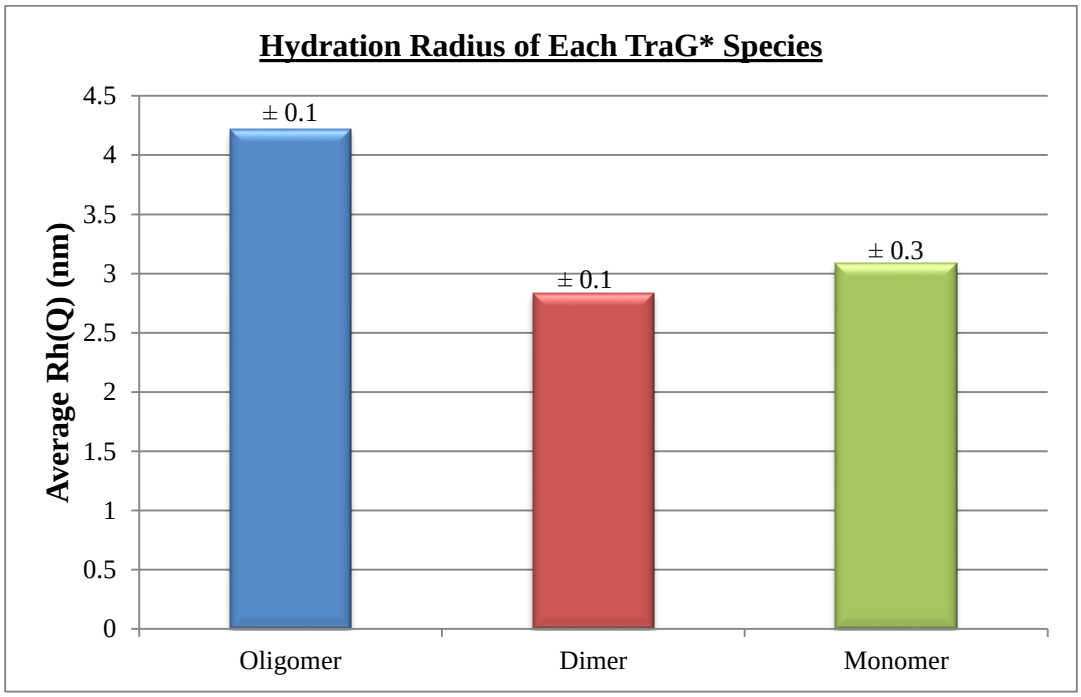


Figure 31: Radius of hydration around particles in TraG* peak. The Rh values from six consistent MALS experiments with least uncertainties were averaged and plotted. The standard deviation of values for each peak between all six experiments is labeled above each bar.

3.2. Ulp1 Experiments

The *ulp1* gene from the *S. cerevisiae* genome was amplified and cloned into the pET28a construct with a C-terminal 6xHis tag. We attempted to express Ulp1 in both BL21 (DE3) and Rosetta cells, however the protein did not express. Various growth and expression temperatures were also tested to expression this gene. The pET28a::Ulp1-His construct was further confirmed

by Sanger sequencing for correct reading frame. The next step is to clone the *ulp1* gene into different expression systems such as pMAL or pGEX and test its expression with more eukaryotic protein-friendly cell lines. Furthermore, attempts to clone a C-terminal truncate of it (Ulp1*; 25.3 kDa) covering the active catalytic site from amino acids 404 to 621 and testing for the expression of the truncate is currently being investigated. This idea has been inspired by that fact that the commercially available Ulp1 protease is also a truncate version of its full length (Mossessova & Lima, 2000).

3.3. Cbu0156 Experiments

3.3.1. Cbu0156 Purification Trials

The cbu0156-SUMO-His fusion protein was expressed and purified using IMAC and AIEX chromatography. The cleavage of the SUMO-His tag was done using the commercially available Ulp1 SUMO protease at two different temperatures (**Figure 32**). Cleavage of the tag was observed to be equally efficient at both 21 °C and 4 °C. The sequencing results of the pETSUMO::cbu0156 plasmid correspond to a theoretical molecular weight of 25.3 kDa for the cbu0156-SUMO-His fusion protein (13.6 kDa for the SUMO-His tag and 11.7 kDa for cbu0156). However, the fusion protein runs at around 30 kDa on 12.5% SDS PAGE (**Figure 32b**). Furthermore, the two lower bands that result from the cleavage of the fusion protein corresponded to ~18 kDa and ~13 kDa, instead of the expected theoretical masses of 13.6 kDa and 11.7 kDa.

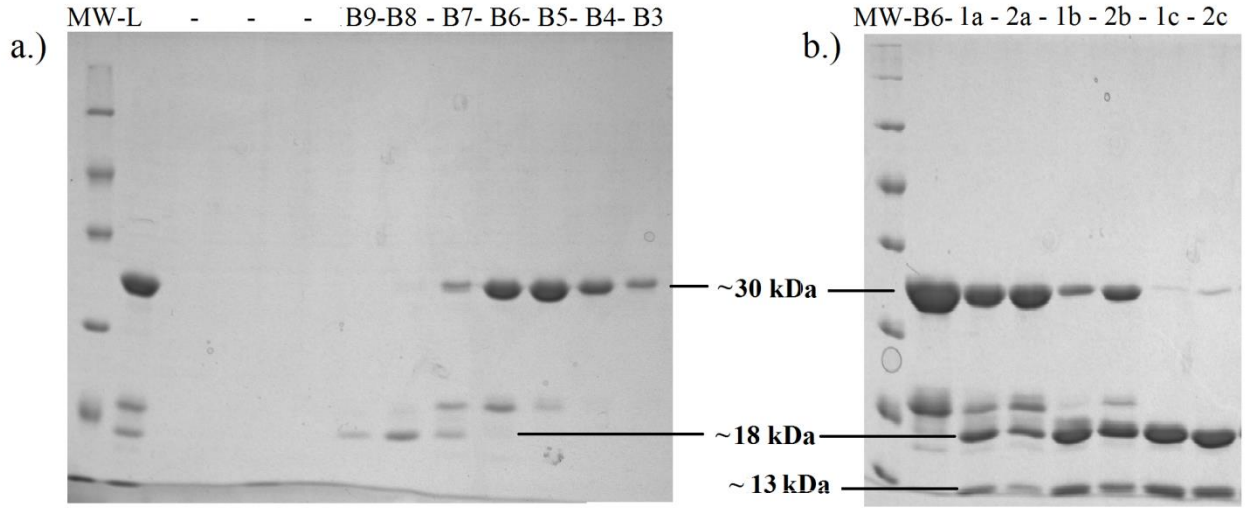


Figure 32: SDSPAGE of cbu0156-SUMO-His digestion. a.) AIEX elution fractions of cbu0156-SUMO-His. On the gel image depicted in b.), fraction B6 from AIEX fractions was digested at two different temperatures using protease buffer with salt and run on SDS PAGE. The samples 1a, 1b, 1c are digests at 21 °C with the numbering corresponding to time points of 1.5 hrs, 5 hrs and 24 hrs, respectively. The same designation is used for the 4 °C digests at the corresponding time points. Sample aliquots at the time points were quickly mixed with SDS dye and denatured. L= Load before purification and MW = molecular weight marker.

3.3.2. Cbu0156-SUMO-His Mass Spectrometry Data

In order to elucidate the discrepancy observed on the band sizes of the two cleavage products, the digestion products from sample 1c and 2c **Figure 32b** were run on ESI-MS. The MS data confirmed the presence of the two cleavage products at their expected theoretical size; cbu0156 at 11.928 kDa and the SUMO-His tag at 13.354 kDa (**Figure 33**). Since the 13.354 kDa peak matches very closely to that of the theoretically calculated mass of SUMO-His, the higher band that is observed around 18 kDa mark is the 11.928 kDa cbu0156 pilin.

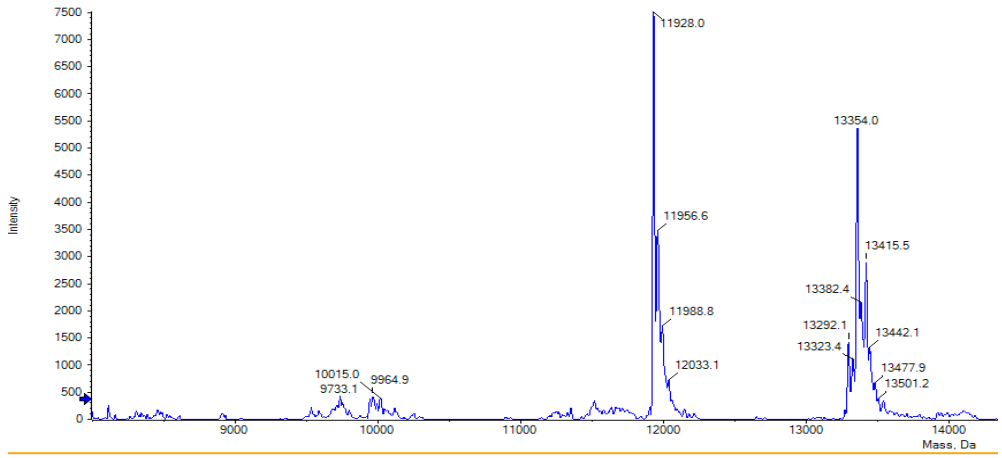


Figure 33: ESI-MS of cbu0156-SUMO-His cleavage products. The cleavage product of cbu0156-SUMO-His samples 1c and 2c shown in **Figure 32** were dialyzed into 200 mM Ammonium acetate buffer and run on MS using direct loading mode.

The property of slow running proteins on SDS PAGE is often associated with high proline content which causes kinks in the structure and makes migration more difficult on the gel (Hames, 1998). Given the sequence of cbu0156 obtained from Sanger sequencing, there are 7 proline (P) residues in the polypeptide chain of this expression product. Furthermore, the cbu0156 polypeptide sequence obtained from Sanger sequencing was aligned against its literature sequence (**Figure 34**). It was observed that the first 28 N-terminal amino acids of the pilin have been omitted during cloning, giving rise to a truncated version of the pilin, similar to those present in **Figure 10**, Section 1.6.1.

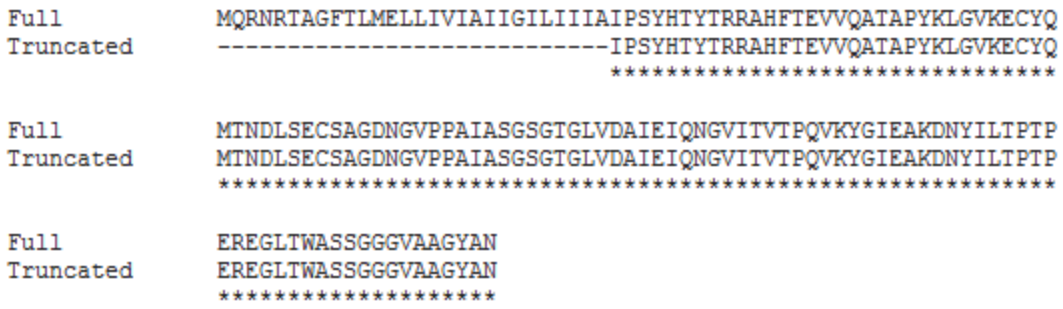


Figure 34: Clustal Omega sequence alignment of cbu0156. The "Full" was obtained from NCBI's protein bank (Accession#: NP_819206) and the "Truncated" sequence was derived from Sanger sequencing of the pETSUMO::cbu0156 plasmid.

3.3.3. Cbu0156 Crystallization Trials

The challenge with crystallizing the cleaved and purified cbu0156 protein is obtaining adequate protein amount and concentration. Hence, the screening of crystals on the MCSG I-IV buffers was performed using the full fusion construct cbu0156-SUMO-His, purified and concentrated after AIEX chromatography. Most of the crystals obtained from these trials showed diffraction patterns similar to that of salt crystals. Other conditions in which crystal-like structures formed were optimized using 24 plates.

4. CONCLUSIONS & FUTURE WORK

4.1. Conclusions

Attempts to express TraG* periplasmically were not successful. Due to large expression volumes, the protein may be aggregating in the cell cytoplasm and being devoid of the chance to localize to the periplasm. It is also possible that misfolding of the protein is interfering with its periplasmic localization. The cytoplasmic expression of TraG* was more successful and efficient at obtaining large amounts of pure protein. While I was able to acquire this amount for both the 6xHis-tagged and untagged TraG*, crystallization hits were observed only in screens performed on untagged TraG*. In reservoir conditions containing 0.2 M Calcium Acetate, 0.1 M Sodium Acetate: Acetic Acid pH 4.5, 30% (w/v) PEG 400, TraG* crystallized quite fast (within 10-15 minutes). This condition is also favorable since it contains the cryo-protectant PEG 400 which will act as a shield against hexagonal ice formation during cryo-cooling of crystals. Although a pH value of 4.5 may seem to be posing a harsh environment for the protein, this is likely brought to more favorable levels once the protein sample in native buffer pH 9.0 is added at a 1:1 ratio with the condition in a crystallization droplet. The crystals obtained from this condition did not last very long and turned into aggregates within a couple of days. These crystal aggregates were used as microseeds for further crystallization which resulted in better quality crystals in conditions of 0.1 M Bis-Tris Propane pH 7.0, 1.8 M Magnesium Sulfate (from the MCSGII screen). These crystals did not diffract well and we are suspecting the exposure of the crystals to freezing temperatures had caused the crystal lattices to deform due to inadequate cryo-protection. Optimization trials of both conditions were performed by altering pH levels, type of precipitant and their concentrations as well as by screening against different salts. More crystals were obtained in the optimization trials under fast crystallization conditions (0.2 M Calcium Acetate, 0.1 M Sodium Acetate: Acetic

Acid pH 4.5, 30% (w/v) PEG 400) while the optimization trials for 0.1 M Bis-Tris Propane pH 7.0, 1.8 M Magnesium Sulfate continue incubating. Crystals obtained in D4 optimization trials were very small and many of them have not given good quality X-ray diffraction.

The SEC-MALS experiments performed to monitor the effect of salt, glycerol and protein concentration on TraG* oligomerization revealed partially conclusive results. Experimental data showed that while an increase in sample protein concentration does not favor oligomerization of TraG*, the content of salt and glycerol in the protein sample did not alter the level oligomerization. The potential presence of different forms of TraG* however, adds a major complication to the above conclusion. More work is required to search for optimal conditions favoring monodispersity of TraG* sample as a monomer to be used in crystallization trials as well as in TraG*-TraG* protein-protein interaction studies.

4.2. Future Work

The reason for the separate elution of different forms of TraG* during AIEX chromatography (**Figure 22**) is yet to be clarified. While it is possible to claim that the increase in salt concentration during the linear elution gradient triggers formation of dimer and oligomer specie, our SEC-MALS data did not reveal clear results on the impact of salt concentration on TraG*-TraG* interactions. If TraG*-TraG* interaction is an experimental artifact caused by misfolded or partially folded protein, then one can promote expression of more properly folded TraG* species in vivo by optimizing expression temperature and induction levels. One can also test other expression systems such as pBAD24 (arabinose inducible), pMAL or pGEX to prompt proper folding. It is also possible to test the legitimacy of TraG* dimerization as being a native

property by expressing full TraG protein in-vivo and observing its behavior using SEC-MALS. One can then use Western blot to tag TraG and use an antibody against the tag to determine if the dimer/oligomer bands observed on the gels consist of TraG proteins in vivo. Tagging TraG with a proximity monitoring fluorescent tag that use fluorescence resonance energy transfer (FRET) to determine close protein-protein interaction can also allow for visualization of TraG oligomerization in vivo. TraG* on its own is non-functional and the full length TraG is required for the protein to carry out its functions (Audette et al., 2007). Hence, all in vivo experiments should be performed using full length TraG.

It would also be ideal to map exposed regions of TraG* in the separated monomer and oligomer species using hydrogen-deuterium exchange (HDX). The same can be performed using the full length TraG which would allow us to determine whether the C-terminal domain on its own (TraG*) inherits a structure similar to that on the full length protein. This could help clarify whether the TraG* monomeric species is a flexed/stretched native structure or an experimental artifact by comparing with full length TraG map.

Efforts to express Ulp1 SUMO-protease in lab and use it to cleave the SUMO-His tag from cbu0156 are still underway. The idea to construct C-terminal truncate of Ulp1 (Ulp1*; 25.3 kDa) covering the active catalytic site from amino acids 404 to 621 and testing for the expression of the truncate version should be investigated. This idea has been inspired by that fact that the commercially available Ulp1 protease is also a truncate version of its full length (Mossessova & Lima, 2000). Once successful results are obtained, crystallization trials of purified cbu0156 can easily be performed.

In order to obtain more lasting and better quality crystal in condition 0.2 M Calcium Acetate, 0.1 M Sodium Acetate: Acetic Acid pH 4.5, 30% (w/v) PEG 400, one needs to slow the speed of crystallization by altering parameters such as decreasing incubation temperature, adding glycerol into protein sample, decreasing protein concentration and using paraffin oil to narrow the path of the drop-reservoir equilibration. On the other hand, the slow crystallization of TraG* in condition C8 (0.1 M Bis-Tris Propane pH 7.0, 1.8 M Magnesium Sulfate) of MCSGII can be improved by screening against more precipitating agents such as MPD and PEGs. In conclusion, TraG* seems to be a soluble T4SS protein that has a great potential for crystallization. Once monodisperse samples of TraG* are successfully obtained, one can further enhance crystallization of this protein using the famous surface lysine methylation method. This method is known to improve the rate of success of protein crystallization by chemical methylation of lysine residues on the protein (Sledz et al., 2010) and should be tested on TraG* for better result.

Since TraG* seems to be crystallizing fairly easily, crystallization trials using random matrix microseeding should also be attempted to increase the chance of better quality crystals to form. Random matrix microseeding can be further applied to perform crystallization trials on the full length TraG using the already obtained TraG* crystals. The expression of full length TraG may produce single TraG species in solution assuming the truncated version (TraG*) does not fold properly on its own and causes non-native behavior. This possibility is further supported by the in-vivo mating assay studies which showed that full length TraG was required for mating pair stabilization function of the protein during F-plasmid transfer (Audette et al., 2007). When expressing full length TraG, the insolubility of the TraG N-terminal half should be kept in mind since it may lead to formation of protein aggregates in solution. The expression of the full protein

will also need to be optimized due to its large size and its function-related physical flexibility as discussed in Section 3.1.3.2.

5. REFERENCES

Achtman, M., Willetts, N., & Clark, a. J. (1972). Conjugational complementation analysis of transfer-deficient mutants of Flac in Escherichia coli. *Journal of Bacteriology*, *110*(3), 831–842.

Arutyunov, D., & Frost, L. S. (2013). F conjugation: Back to the beginning. *Plasmid*, *70*(1), 18–32. http://doi.org/10.1016/j.plasmid.2013.03.010

Audette, G. F., & Hazes, B. (2007). Development of Protein Nanotubes from a Multi-Purpose Biological Structure. *Journal of Nanoscience and Nanotechnology*, *7*, 2222–2229. http://doi.org/10.1166/jnn.2007.650

Audette, G. F., Irvin, R. T., & Hazes, B. (2004). Crystallographic analysis of the Pseudomonas aeruginosa strain K122-4 monomeric pilin reveals a conserved receptor-binding architecture. *Biochemistry*, *43*(36), 11427–11435. http://doi.org/10.1021/bi048957s

Audette, G. F., Manchak, J., Beatty, P., Klimke, W. a., & Frost, L. S. (2007). Entry exclusion in F-like plasmids requires intact TraG in the door that recognizes its cognate TraS in the recipient. *Microbiology*, *153*, 442–451. http://doi.org/10.1099/mic.0.2006/001917-0

Audette, G. F., Van Schaik, E. J., Hazes, B., & Irvin, R. T. (2004). DNA-binding protein nanotubes: Learning from nature's nanotech examples. *Nano Letters*, *4*, 1897–1902. http://doi.org/10.1021/nl048942f

Baca, A. M., & Hol, W. G. J. (2000). Overcoming codon bias: A method for high-level overexpression of Plasmodium and other AT-rich parasite genes in Escherichia coli.

International Journal for Parasitology, *30*(2), 113–118. http://doi.org/10.1016/S0020-7519(00)00019-9

Beare, P. A., Gilk, S. D., Larson, C. L., Hill, J., Stead, C. M., Omsland, A., … Heinzen, R. A. (2011). Dot/Icm type IVB secretion system requirements for Coxiella burnetii growth in human macrophages. *mBio*, *2*(4). http://doi.org/10.1128/mBio.00175-11

Beare, P. A., Unsworth, N., Andoh, M., Voth, D. E., Omsland, A., Gilk, S. D., … Heinzen, R. A. (2009). Comparative genomics reveal extensive transposon-mediated genomic plasticity and diversity among potential effector proteins within the genus coxiella. *Infection and Immunity*, *77*(2), 642–656. http://doi.org/10.1128/IAI.01141-08

Beeckman, D. S. A., & Vanrompay, D. C. G. (2010). Bacterial secretion systems with an emphasis on the chlamydial Type III secretion system. *Current Issues in Molecular Biology*, *12*(1), 17–41. http://doi.org/v12/17 [pii]

Bian, Z., & Normark, S. (1997). Nucleator function of CsgB for the assembly of adhesive surface organelles in Escherichia coli. *EMBO Journal*, *16*(19), 5827–5836. http://doi.org/10.1093/emboj/16.19.5827

Bishop, J. B., Martin, J. C., & Rosenblum, W. M. (1991). A light scattering method for qualitatively monitoring aggregation rates in macromolecular systems. *Journal of Crystal Growth*, *110*(1–2), 164–170. http://doi.org/10.1016/0022-0248(91)90880-E

Blankschien, M. D., Potrykus, K., Grace, E., Choudhary, A., Vinella, D., Cashel, M., & Herman, C. (2009). TraR, a homolog of a RNAP secondary channel interactor, modulates transcription. *PLoS Genetics*, *5*(1). http://doi.org/10.1371/journal.pgen.1000345

Borgstahl, G. E. O. (2007). How to use dynamic light scattering to improve the likelihood of growing macromolecular crystals. *Methods in Molecular Biology (Clifton, N.J.)*, *363*, 109–

129. http://doi.org/10.1007/978-1-59745-209-0_6

Burrus, V., & Waldor, M. K. (2004). Shaping bacterial genomes with integrative and conjugative elements. *Research in Microbiology*, *155*(5), 376–386. http://doi.org/10.1016/j.resmic.2004.01.012

Cascales, E., & Christie, P. J. (2003). The versatile bacterial type IV secretion systems. *Nature Reviews. Microbiology*, *1*(2), 137–149. http://doi.org/10.1038/nrmicro753

Chandran Darbari, V., & Waksman, G. (2015). Structural Biology of Bacterial Type IV Secretion Systems. *Annual Review of Biochemistry*, *84*, 603–29. http://doi.org/10.1146/annurev-biochem-062911-102821

Connell, I., Agace, W., Klemm, P., Schembri, M., Mărild, S., & Svanborg, C. (1996). Type 1 fimbrial expression enhances Escherichia coli virulence for the urinary tract. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(18), 9827–9832. http://doi.org/10.1073/pnas.93.18.9827

Cooper, T. F. (2007). Recombination speeds adaptation by reducing competition between beneficial mutations in populations of Escherichia coli. *PLoS Biology*, *5*(9), 1899–1905. http://doi.org/10.1371/journal.pbio.0050225

Craig, L., Pique, M. E., & Tainer, J. A. (2004). Type IV pilus structure and bacterial pathogenicity. *Nature Reviews Microbiology*, *2*(5), 363–378. http://doi.org/10.1038/nrmicro885

Craig, L., Taylor, R. K., Pique, M. E., Adair, B. D., Arvai, A. S., Singh, M., … Tainer, J. A. (2003). Type IV pilin structure and assembly: X-ray and EM analyses of Vibrio cholerae toxin-coregulated pilus and Pseudomonas aeruginosa PAK pilin. *Molecular Cell*, *11*(5), 1139–1150. http://doi.org/10.1016/S1097-2765(03)00170-9

Dobrindt, U., Hochhut, B., Hentschel, U., & Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nature Reviews. Microbiology*, *2*(5), 414–424. http://doi.org/10.1038/nrmicro884

Duguid, J. P., Anderson, E. S., & Campbell, I. (1966). Fimbriae and Adhesive Properties in Salmonellae. *Journal of Pathology and Bacteriology*, *92*(1), 107-. article. http://doi.org/10.1002/path.1700920113

Filutowicz, M., Burgess, R., Gamelli, R. L., Heinemann, J. A., Kurenbach, B., Rakowski, S. A., & Shankar, R. (2008). Bacterial conjugation-based antimicrobial agents. *Plasmid*. http://doi.org/10.1016/j.plasmid.2008.03.004

Firth, N., Ippen-ihler, K., & Skurray, R. a. (1996). Structure and Function of the F Factor and Mechanism of Conjugation. *Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology*, 2377–2401.

Firth, N., & Skurray, R. (1992). Characterization of the F plasmid bifunctional conjugation gene, traG. *Molecular & General Genetics : MGG*, *232*(1), 145–153. http://doi.org/10.1007/bf00299147

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. *Microbiological Reviews*, *58*(2), 162–210.

Frost, L. S., Leplae, R., Summers, A. O., & Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nature Reviews. Microbiology*, *3*(9), 722–732. http://doi.org/10.1038/nrmicro1235

Furuya, E. Y., & Lowy, F. D. (2006). Antimicrobial-resistant bacteria in the community setting. *Nature Reviews Microbiology*, *4*(January), 36–45. http://doi.org/10.1038/nrmicro1325

Gil, H., Benach, J. L., & Thanassi, D. G. (2004). Presence of Pili on the Surface of Francisella

tularensis. *Infection and Immunity*, *72*(5), 3042–3047. http://doi.org/10.1128/IAI.72.5.3042-3047.2004

Girardeau, J. P., Bertin, Y., & Callebaut, I. (2000). Conserved structural features in class I major fimbrial subunits (Pilin) in gram-negative bacteria. Molecular basis of classification in seven subfamilies and identification of intrasubfamily sequence signature motifs which might Be implicated in quaterna. *J Mol Evol*, *50*(5), 424–442. http://doi.org/10.1007/s002390010045

Girón, J. a, Ho, a S., & Schoolnik, G. K. (1991). An inducible bundle-forming pilus of enteropathogenic Escherichia coli. *Science (New York, N.Y.)*, *254*(5032), 710–713. http://doi.org/10.1126/science.1683004

Griffiths, A., Miller, J., Suzuki, D., & Al., E. (2000). *An Introduction to Genetic Analysis. 7th edition.* San Fransisco: W. H. Freeman.

Guglielmini, J., Néron, B., Abby, S. S., Garcillán-Barcia, M. P., La Cruz, D. F., & Rocha, E. P. C. (2014). Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Research*, *42*(9), 5715–5727. http://doi.org/10.1093/nar/gku194

Hames, B. D. (1998). *Gel electrophoresis of proteins : a practical approach*. Oxford University Press.

Hazes, B., & Frost, L. (2008). Towards a systems biology approach to study type II/IV secretion systems. *Biochimica et Biophysica Acta*, *1778*, 1839–1850. http://doi.org/10.1016/j.bbamem.2008.03.011

Hazes, B., Sastry, P. a, Hayakawa, K., Read, R. J., & Irvin, R. T. (2000). Crystal structure of Pseudomonas aeruginosa PAK pilin suggests a main-chain-dominated mode of receptor

binding. *Journal of Molecular Biology*, *299*(4), 1005–1017. http://doi.org/10.1006/jmbi.2000.3801

Helaine, S., Dyer, D. H., Nassif, X., Pelicic, V., & Forest, K. T. (2007). 3D structure/function analysis of PilX reveals how minor pilins can modulate the virulence properties of type IV pili. *Proceedings of the National Academy of Sciences*, *104*(40), 15888–15893. http://doi.org/10.1073/pnas.0707581104

Jakubowski, S. J., Krishnamoorthy, V., Cascales, E., & Christie, P. J. (2004). Agrobacterium tumefaciens VirB6 domains direct the ordered export of a DNA substrate through a type IV secretion system. *Journal of Molecular Biology*, *341*(4), 961–977. http://doi.org/10.1016/j.jmb.2004.06.052

Judd, P. K., Mahli, D., & Das, A. (2005). Molecular characterization of the Agrobacterium tumefaciens DNA transfer protein VirB6. *Microbiology (Reading, England)*, *151*(Pt 11), 3483–92. http://doi.org/10.1099/mic.0.28337-0

Keeling, P. J., & Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nature Reviews. Genetics*, *9*(8), 605–618. http://doi.org/10.1038/nrg2386

Klimke, W. A., Rypien, C. D., Klinger, B., Kennedy, R. A., Rodriguez-Maillard, J. M., & Frost, L. S. (2005). The mating pair stabilization protein, TraN, of the F plasmid is an outer-membrane protein with two regions that are important for its function in conjugation. *Microbiology*, *151*(11), 3527–3540. http://doi.org/10.1099/mic.0.28025-0

Kohler, P. L., Chan, Y. a., Hackett, K. T., Turner, N., Holly L., H., Cloud-Hansen, C., & Dillard, J. P. (2013). Mating pair formation homologue trag is a variable membrane protein essential for contact-independent type iv secretion of chromosomal DNA by neisseria gonorrhoeae. *Journal of Bacteriology*, *195*(8), 1666–1679. http://doi.org/10.1128/JB.02098-12

Lawley, T. D., Klimke, W. a., Gubbins, M. J., & Frost, L. S. (2003). F factor conjugation is a true type IV secretion system. *FEMS Microbiology Letters*, *224*(1), 1–15. http://doi.org/10.1016/S0378-1097(03)00430-0

Lawn, A. M., Meynell, E., Meynell, G. G., & Datta, N. (1967). Sex Pili and Classification of Sex Factors in Enterobacteriaceae. *Nature*, *216*(5113), 343-. article. http://doi.org/10.1038/216343a0

Le, A. (n.d.). Alila Medical Images. Retrieved July 4, 2014, from http://www.alilamedicalmedia.com/-/galleries/images-videos-by-medical-specialties/medical-microbiology-pathology/-/medias/103aeff4-4439-11e3-b7ef-f1a470cbde24-bacterial-conjugation-labeled-diagram

Lederberg, J., & Tatum, E. L. (1953). Sex in Bacteria: Genetic Studies, 1945-1952. *Science*, *118*(3059), 169–175. article. http://doi.org/10.1126/science.118.3059.169

Lento, C., Ferraro, M., Wilson, D., & Audette, G. F. (2016). HDX-MS and deletion analysis of the type 4 secretion system protein TraF from the Escherichia coli F plasmid. *FEBS Letters*, *590*(3), 376–386. http://doi.org/10.1002/1873-3468.12066

Lu, H. M., Motley, S. T., & Lory, S. (1997). Interactions of the components of the general secretion pathway: role of *Pseudomonas aeruginosa* type IV pilin subunits in complex formation and extracellular protein secretion. *Molecular Microbiology*, *25*(2), 247–259. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9282737

Lujan, S. A., Guogas, L. M., Ragonese, H., Matson, S. W., & Redinbo, M. R. (2007). Disrupting antibiotic resistance propagation by inhibiting the conjugative DNA relaxase. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(30), 12282–12287. http://doi.org/10.1073/pnas.0702760104

Marrero, J., & Waldor, M. K. (2007). Determinants of entry exclusion within Eex and TraG are cytoplasmic. *Journal of Bacteriology*, *189*(17), 6469–6473. http://doi.org/10.1128/JB.00522-07

Merz, A. J., & So, M. (2000). Interactions of pathogenic neisseriae with epithelial cell membranes. *Annual Review of Cell and Developmental Biology*, *16*, 423–57. http://doi.org/10.1146/annurev.cellbio.16.1.423

Miroux, B., & Walker, J. E. (1996). Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology*, *260*(3), 289–298. http://doi.org/10.1006/jmbi.1996.0399

Moodie, C. E., Thompson, H. A., Meltzer, M. I., & Swerdlow, D. L. (2008). Prophylaxis after exposure to Coxiella burnetii. *Emerging Infectious Diseases*, *14*(10), 1558–1566. http://doi.org/10.3201/eid1410.080576

Mossessova, E., & Lima, C. D. (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Molecular Cell*, *5*(5), 865–876. http://doi.org/10.1016/S1097-2765(00)80326-3

Niesen, F. H., Koch, A., Lenski, U., Harttig, U., Roske, Y., Heinemann, U., & Hofmann, K. P. (2008). An approach to quality management in structural biology: Biophysical selection of proteins for successful crystallization. *Journal of Structural Biology*, *162*(3), 451–459. http://doi.org/10.1016/j.jsb.2008.03.007

O'Cellaghan, D., Cazevieille, C., Allardet-Servent, A., Boschiroli, M. L., Bourg, G., Foulongne, V., … Ramuz, M. (1999). A homologue of the Agrobacterium tumefaciens VirB and Bordetella pertussis Ptl type IV secretion systems is essential for intracellular survival of Brucella suis. *Molecular Microbiology*, *33*(6), 1210–1220. http://doi.org/10.1046/j.1365-

2958.1999.01569.x

Ottow, J. C. (1975). Ecology, physiology, and genetics of fimbriae and pili. *Annual Review of Microbiology*, *29*, 79–108. Journal Article. http://doi.org/10.1146/annurev.mi.29.100175.000455

Petrov, A., & Audette, G. F. (2012). Peptide and protein-based nanotubes for nanobiotechnology. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, *4*(October), 575–585. http://doi.org/10.1002/wnan.1180

Petrov, A., Lombardo, S., & Audette, G. F. (2013). Fibril-mediated oligomerization of pilin-derived protein nanotubes. *Journal of Nanobiotechnology*, *11*(1), 24. http://doi.org/10.1186/1477-3155-11-24

Proft, T., & Baker, E. N. (2009). Pili in Gram-negative and Gram-positive bacteria - Structure, assembly and their role in disease. *Cellular and Molecular Life Sciences*, *66*(4), 613–635. http://doi.org/10.1007/s00018-008-8477-4

Qadri, F., Svennerholm, A.-M., Faruque, A. S. G., & Sack, R. B. (2005). Enterotoxigenic Escherichia coli in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clinical Microbiology Reviews*, *18*(3), 465–83. http://doi.org/10.1128/CMR.18.3.465-483.2005

Rammelkamp, C. H., & Maxon, T. (1942). Resistance of Staphylococcus aureus to the Action of Penicillin. *Experimental Biology and Medicine* , *51*(3), 386–389. http://doi.org/10.3181/00379727-51-13986

Remaut, H., & Waksman, G. (2004). Structural biology of bacterial pathogenesis. *Curr Opin Struct Biol*, *14*(2), 161–170. http://doi.org/10.1016/j.sbi.2004.03.004

Sakellaris, H., & Scott, J. R. (1998). New tools in an old trade: CS1 pilus morphogenesis.

Molecular Microbiology. http://doi.org/10.1046/j.1365-2958.1998.01088.x

Shala, A., Ferraro, M., & Audette, G. F. (2016). *Bacterial Secretion Systems: Nanomachines for Infection and Genetic Diversity, in R. Bawa, G.F. Audette & I. Rubenstein Eds. Handbook of Clinical Nanomedicine: Nanoparticles, Imaging, Therapy and Clinical Applications* (Chapter 24). Singapore: Pan Sanford Publishing.

Sledz, P., Zheng, H., Murzyn, K., Chruszcz, M., Zimmerman, M. D., Chordia, M. D., … Minor, W. (2010). New surface contacts formed upon reductive lysine methylation: Improving the probability of protein crystallization. *Protein Science*, *19*(7), 1395–1404. http://doi.org/10.1002/pro.420

Smillie, C., Garcillan-Barcia, M. P., Francia, M. V., Rocha, E. P. C., & de la Cruz, F. (2010). Mobility of Plasmids. *Microbiology and Molecular Biology Reviews*, *74*(3), 434–452. http://doi.org/10.1128/MMBR.00020-10

Taylor, R. K., Miller, V. L., Furlong, D. B., & Mekalanos, J. J. (1987). Use of phoA gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. *Proceedings of the National Academy of Sciences of the United States of America*, *84*(9), 2833–7. http://doi.org/10.1073/pnas.84.9.2833

Thanassi, D. G., Bliska, J. B., & Christie, P. J. (2012). Surface organelles assembled by secretion systems of Gram-negative bacteria: Diversity in structure and function. *FEMS Microbiology Reviews*, *36*(6), 1046–1082. http://doi.org/10.1111/j.1574-6976.2012.00342.x

Thanassi, D. G., Saulino, E. T., Lombardo, M. J., Roth, R., Heuser, J., & Hultgren, S. J. (1998). The PapC usher forms an oligomeric channel: implications for pilus biogenesis across the outer membrane. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(6), 3146–51. http://doi.org/10.1073/pnas.95.6.3146

Thanassi D, N. S. S. K. S. S. B. A. (2007). Fimbriae: Classification and Biochemistry. *EcoSal Plus*. article. Retrieved from http://www.asmscience.org/content/journal/ecosalplus/10.1128/ecosalplus.2.4.2.1

Tseng, T.-T., Tyler, B. M., & Setubal, J. C. (2009). Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiology*, *9 Suppl 1*, S2. http://doi.org/10.1186/1471-2180-9-S1-S2

van Schaik, E. J., Chen, C., Mertens, K., Weber, M. M., & Samuel, J. E. (2013). Molecular pathogenesis of the obligate intracellular bacterium Coxiella burnetii. *Nature Reviews. Microbiology*, *11*(8), 561–73. http://doi.org/10.1038/nrmicro3049

Waksman, G., & Orlova, E. V. (2014). Structural organisation of the type IV secretion systems. *Current Opinion in Microbiology*, *17*(1), 24–31. http://doi.org/10.1016/j.mib.2013.11.001

Willetts, N., & Skurray, R. (1980). The conjugation system of F-like plasmids. *Annual Review of Genetics*, *14*(1), 41–76. http://doi.org/10.1146/annurev.ge.14.120180.000353

Williams, J. J., & Hergenrother, P. J. (2008). Exposing plasmids as the Achilles' heel of drug-resistant bacteria. *Current Opinion in Chemical Biology*. http://doi.org/10.1016/j.cbpa.2008.06.015

Woodcock, D. M., Crowther, P. J., Doherty, J., Jefferson, S., Decruz, E., Noyer-Weidner, M., … Graham, M. W. (1989). Quantitative evaluation of Escherichia coli host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic Acids Research*, *17*(9), 3469–3478. http://doi.org/10.1093/nar/17.9.3469

Wurpel, D. J., Beatson, S. A., Totsika, M., Petty, N. K., & Schembri, M. A. (2013). Chaperone-Usher Fimbriae of Escherichia coli. *PLoS ONE*, *8*(1). http://doi.org/10.1371/journal.pone.0052835

Xu, X. F., Tan, Y. W., Lam, L., Hackett, J., Zhang, M., & Mok, Y. K. (2004). NMR structure of a Type IVb pilin from Salmonella typhi and its assembly into pilus. *Journal of Biological Chemistry, 279*(30), 31599–31605. http://doi.org/10.1074/jbc.M404727200

Zhang, X. L., Tsui, I. S. M., Yip, C. M. C., Fung, A. W. Y., Wong, D. K. H., Dai, X., … Morris, C. (2000). Salmonella enterica serovar Typhi uses type IVB pili to enter human intestinal epithelial cells. *Infection and Immunity, 68*(6), 3067–3073. http://doi.org/10.1128/IAI.68.6.3067-3073.2000

Zhu, H. W., Xu, C. L., Wu, D. H., Wei, B. Q., Vajtai, R., & Ajayan, P. M. (2002). Direct synthesis of long single-walled carbon nanotube strands. *Science, 296*(5569), 884–886. http://doi.org/10.1126/science.1066996

6. APPENDIX: Review of Methods and Techniques

Crystallization of proteins requires large quantities (milligrams) of purified samples. High purity and homogeneity of the protein samples are crucial for successful crystallization experiments. Dynamic light scattering (DLS) is used to characterize the polydispersity of a sample and observe the presence of different oligomeric forms or even aggregates of the protein in solution. Electrospray ionization mass spectrometry (ESI-MS) helps confirm the appropriate masses of both monomeric and oligomeric species along with form along with their stability under ionizing conditions. Other methods such as differential scanning fluorometry (DSF), CD spectroscopy and analytical ultracentrifugation are also used to help characterize the stability of the protein in different buffers, presence of different ligands and proper folding. Proper and detailed characterization of a protein for preparation in crystallography should be given more importance than the actual biochemical experiments. However, it is often the case that researchers take partial purity and the activity of a protein as sufficient criteria to perform crystallization trials.

A.1. Protein crystallization and X-ray diffraction

X-ray crystallography is a method used for identifying the atomic and molecular structure of macromolecules such as proteins or a small molecule such as a salt. The first stage towards resolving the structure of a molecule of interest is to obtain crystals of it under specific conditions. For optimal crystallization conditions, a molecule sample high in purity and homogeneity and low in polydispersity is required. Factors such as solubility in buffers used for purification as well as crystallization and the stability of the molecule under these conditions are also considered. When performing crystallization trials on a particular protein, an experimenter is screening for a condition that contains ideal parameters of pH, ionic strength, temperature, protein concentration,

salt variants, ligands or additives, precipitant variants and the actual crystallization method to use (**Figure 35**). Furthermore, it is often the case in crystallography that once crystals are obtained in a screen, the crystallization condition needs to be further optimized to obtain larger single crystals suitable for X-ray diffraction experiments.

In a crystallization set up, the concentration of the precipitant in the reservoir is usually higher than that in the drop. The concentration gradient created in this environment forces evaporation of water from the drop towards reservoir, thus lowering the volume of the drop and increasing the protein concentration. At some point during evaporation, the concentration of the protein reaches supersaturation (**Figure 35**) and assuming all other parameters including pH, ionic strength, type of buffer and additives are fine, the protein might start crystallizing.

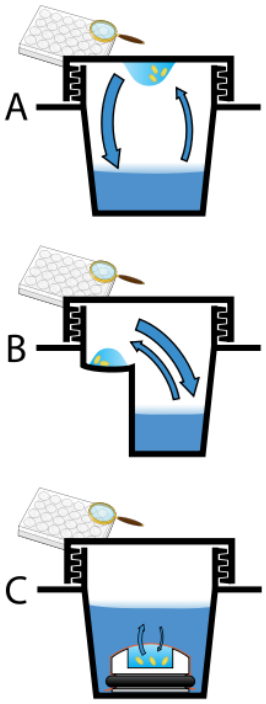


Figure 35: Three various methods used in crystal set up. A: hanging drop, B: sitting drop, C: Microdialysis (Adenosine, 2010)

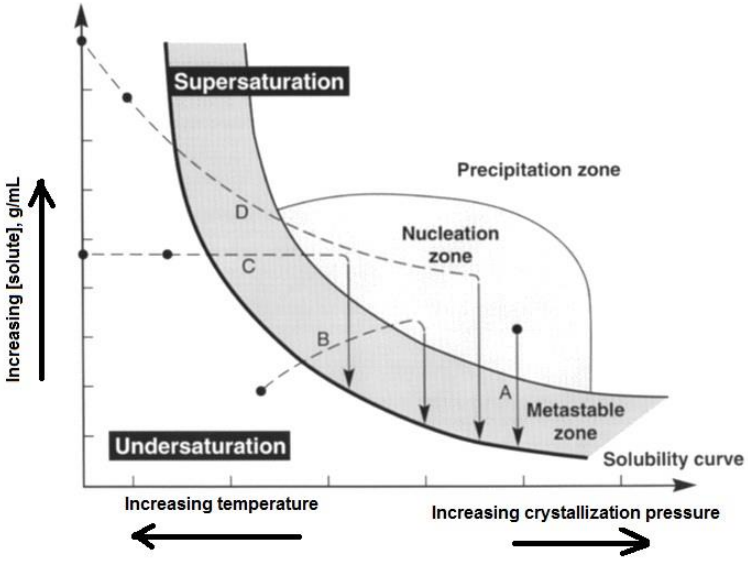


Figure 36: Protein crystallization phase diagram. A newly set up drop usually beings at the under-saturation region of the phase and increases its concentration overtime (Chayen, 1998).

Once obtained, the protein crystal is exposed to X-rays and the crystalline atoms of the molecule diffract the X-rays into specific directions. The data of the angles and intensities of these

diffracted beams is collected on a detector film and a three-dimensional image of the density map of electrons within the crystal produced. Using this density map, the mean position of each atom, chemical bonds and structural order in the crystal is then determined.

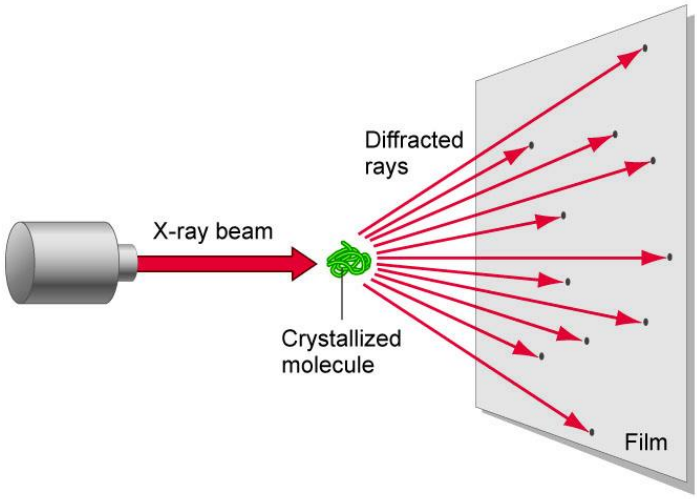


Figure 37: Schematic diagram of X-ray diffraction (260h/X ray Crystallography).

A.2. Dynamic Light Scattering

Dynamic light scattering (DLS) is a technique used mainly to analyze size distribution profile of particles in suspension or polymers in solution (Berne & Pecora, 2000). It can also be used to monitor the behavior of macromolecules in various solutions. A more advanced light scattering method that is used today is the multi-angle light scattering (MALS) or 3D-dynamic light scattering method. In the method, a monochromatic light source such as a laser, is shot through a polarizer into a sample solution. The light scattered by the particles in solution goes

through a second polarizer and is then collected by a photomultiplier detector. In MALS, there are many detectors surrounding the sampling cell to collect light scattered at various angles (Rayleigh scattering) and allow for collection of more data. The diffracted light from the molecules may interfere constructively or destructively. The collection of light diffraction data compares the intensity of light (**Figure 38**) at each angle over time and computes information on the radii, polydispersity, concentration and molecular weight of the sample being analyzed.

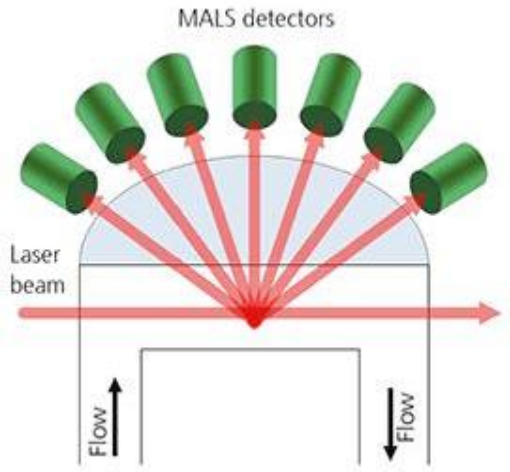


Figure 38: Schematic diagram of MALS (Malvern Instruments, 1989).

A.3. Electrospray Ionization Mass Spectrometry (ESI-MS)

ESI is a technique used to produce ions using an electrospray with high voltage to create aerosols from the liquid sample. The generation of ions from macromolecules is particularly useful in MS (**Figure 39**) because it decreases the susceptibility of these molecules to degradation during ionization (Ho *et al.,* 2003). In order to perform MS analysis, these ions are accelerated using an electrical field to bring them all to same level of kinetic energy. Accelerated ions are then deflected by a magnetic field according to their masses. The level of deflection depends on the mass and charge of the ions. The lighter and more charged the ion, the more it gets deflected. Using the electrically detected mass/charge (m/z) ratio of the ions the corresponding mass of the sample analyte is extrapolated (Clark, 2000).

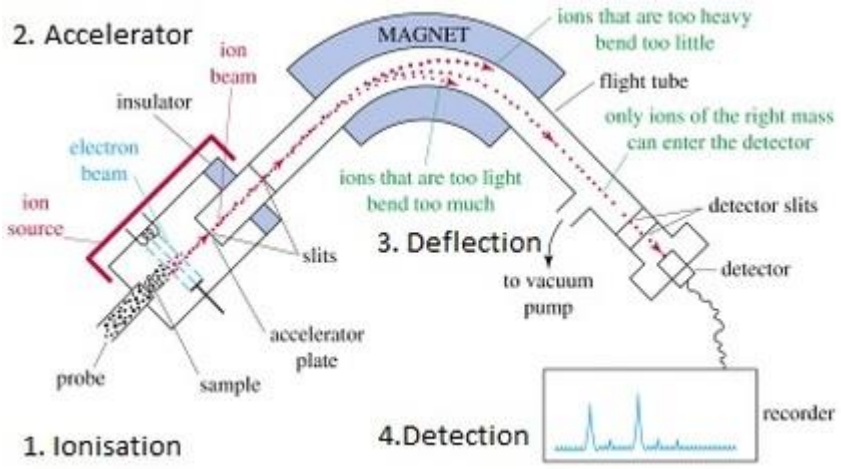


Figure 39: Schematic diagram of electrospray ionization mass spectrometry (Media Portfolio).

Appendix Reference

"260h / X Ray Crystallography." https://260h.pbworks.com/w/page/30814223/X Ray Crystallography. (Date Accessed: September 17, 2016).

Adenosine (2010). 3 methods of creating crystals for X-crystallography. Wikipedia. https://commons.wikimedia.org/wiki/File:CrystalDrops.svg. (Date Accessed: September 17, 2016).

Berne, B.J.; Pecora, R. Dynamic Light Scattering. Courier Dover Publications (2000) ISBN 0-486-41155-9

Chayen, NE (1998) Comparative Studies of Protein Crystallization by Vapour-Diffusion and Microbatch Techniq. *Acta Crystall. Sec. D Biol. Crystall.* **54,** 8–15 (September 17, 2016).

Chemical Instrumentation: Mass Spectrometry. http://chemicalinstrumentation.weebly.com/mass-spectrometry.html (Date Accessed: September 17, 2016).

Clark, J. (2000). Chemguide. The Mass Spectrometer. http://chemguide.co.uk/analysis/masspec/howitworks.html. (Date Accessed: September 17, 2016).

Ho, CS; Chan MHM; Cheung RCK; Law LK; Lit LCW; Ng KF; Suen MWM; Tai HL (February 2003). *"Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications"*. Clin. Biochem. Rev. **24** (1): 3–12.

"Malvern Instruments", Inc. *Anal. Chem.* **61,** 545A–545A (1989). http://pubs.acs.org/doi/abs/10.1021/ac00183a731 (Date Accessed: September 17, 2016).

"Media Portfolio." http://wps.prenhall.com/wps/media/objects/340/348272/wade_ch12.html. (Date Accessed: September 17, 2016)