

CHARACTERIZATION OF A HIGHLY DYNAMIC REGION IN

TRAG FROM THE F PLASMID

NICHOLAS BRAGAGNOLO

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 BIOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

SEPTEMBER 2019

© NICHOLAS BRAGAGNOLO, 2019

Abstract

The *Escherichia coli* F plasmid is representative of conjugative type IV secretion system for the transmission of mobile DNA elements in bacteria, a contributor to widespread antibiotic resistance. The TraG protein of this system consists of a membrane-bound N-terminal domain and a periplasmic C-terminal domain, denoted TraG*. TraG* is essential in preventing redundant DNA transfer. In the donor cell it interacts with TraN within the outer membrane to facilitate mating pair stabilisation. However, TraG* also interacts with a cognate TraS in the inner membrane of the recipient cell to prevent conjugation when the recipient cell carries the same plasmid. This thesis presents structural studies of TraG*; Thermofluor, circular dichroism and HDX-MS experiments showed N-terminal truncation mutants displayed higher stability and less disordered content relative to full-length TraG*. The 45 N-terminal residues of TraG* were predicted to be highly dynamic, possibly serving as a flexible linker between two independently functioning domains.

Acknowledgments

Firstly, I would like to thank my supervisor Dr. Gerald Audette for his guidance and dedication to his students. His support in the decisions I made for this project taught me valuable lessons about the proper way to pursue a scientific question and I have the utmost gratitude for his contributions. I would also like to thank my advisor Dr. Vivian Saridakis for inspiring me to never settle for mediocrity and aspire for high-quality data every time I perform an experiment. I would also like to thank Dr. Yi Sheng and Dr. Derek Wilson for serving on my defense committee and providing me valuable feedback on my project. I am appreciative of the administrators in the department of biology for their assistance.

I thank my fellow Audette lab members, past and present. Ayat Yaseen for training me in many vital techniques and for welcoming me in the lab, Fettah Erdogan for his sage advice on handling TraG*, and Christina Rodriguez for her friendship and her continued assistance in the lab. I'd like to thank many members of the Wilson lab; Irina, Lucienne, Xiaojing, John and Lisa, for their thoughtful discussions and aid in experimentation, especially Cristina Lento and Peter Liuni for helping me learn how to perform mass spectrometry on my own. I am grateful for my undergraduate thesis supervisor Dr. Katalin Hudak and her lab members that trained me and developed my foundation as a scientist. To the many students I've taught as a teaching assistant at the university, thank you for developing my ability in conveying a lesson in a concise and understandable way. Acknowledgements are deserved by many faculty of science members, and colleagues at York University and other institutions that I'm happy to call friends, especially those in the Life Science Building; there are too many to name but know that your contributions to my academic career and to my life are innumerable.

Lastly, I would like to thank my friends and family. My fellow musicians, including my brother John, for teaching me that life without artistry and self expression is like a protein crystal that does not diffract. My brother Mark for showing me that chefs can work just as many hours as grad students and for occasionally making me amazing food. My dog Tanner for being the best possible distraction when I need him to be. And prominently my parents, Margie and Phil, for their unconditional love and support, their undying belief in me despite my self-doubts, and for providing me with the opportunities to walk whatever I path I choose.

Table of Contents

Abstract	ii
Acknowledgments	iii
Table of Contents	iv
List of Tables	vii
List of Figures	viii
List of Abbreviations	x

Chapters

Chapter 1 – Introduction

1.1 The Mounting Threat of Bacterial Antibiotic Resistance as Aided by Horizontal Gene Transfer	1
1.2 Evidence for Urgency in Studying Gram-Negative Bacteria	3
1.3 Gram-Negative Bacterial Conjugation by T4SS	
1.3.1 <i>The Utility of Conjugative T4SS</i>	4
1.3.2 <i>Initiation of Conjugation, Pilus Assembly and Extension</i>	5
1.3.3 <i>Pilus Retraction and Exclusion Events</i>	6
1.3.4 <i>Conjugative DNA Transfer</i>	8
1.4 The F plasmid and F pilus	9
1.5 Entry Exclusion and Incompatibility Groupings	
1.5.1 <i>The Discovery of Superinfection Immunity</i>	11
1.5.2 <i>Plasmid Incompatibility</i>	13
1.5.3 <i>Entry Exclusion in Different Incompatibility Groups</i>	13
1.6 Characteristics of TraG and TraG*from the F plasmid	
1.6.1 <i>Insights on TraG Structure and Function</i>	18
1.6.2 <i>Solved Structures of Proteins Analogous to TraG</i>	19
1.7 Intrinsic Disorder in Proteins	

1.7.1 Common Functions of Dynamic Proteins.....	20
1.7.2 Pi-Pi Interactions in Disordered Regions of Proteins.....	23
1.8 Topic of Research: Structural Study of TraG*	25
Chapter 2 – Materials and Methods	
2.1 Reagents and Equipment	28
2.2 Analysis of TraG _F Using Software for the Detection of Disorder	29
2.3 Cloning of TraG* Constructs	
2.3.1 Plasmid Extraction of pET28a.....	29
2.3.2 PCR Amplification of traG* and traG* _Δ	30
2.3.3 RE Digestion, Ligation and Transformation of traG* Mutants into pET28a.....	30
2.4 Expression and Purification of TraG* Mutants	
2.4.1 Optimization of TraG* Expression.....	32
2.4.2 Purification Schemes of TraG* Variants.....	32
2.4.3 Visualization of Proteins using SDS PAGE.....	33
2.4.4 Desalting and Quantification of TraG*.....	33
2.5 Screening and Optimization of Crystallisation conditions of TraG*	
2.5.1 Screening of Crystallisation conditions.....	34
2.5.2 Thermofluor Assay.....	35
2.6 Analysis of TraG* Structural Properties	
2.6.1 Circular Dichroism.....	37
2.6.2 Global HDX MS.....	38
Chapter 3 – Results and Discussion	
3.1 Software Mediated TraG Structural Predictions PSP.....	40
3.1.1 PSP Software Analysis of TraG _F Reveals a Putative IDR.....	42
3.2 Successful Cloning of TraG* Constructs	45
3.3 Expression and Purification of TraG* Constructs	46
3.4 Thermofluor Analyses of TraG*His and HisTraG* _Δ Reveal Differences in	50
Protein Stability	
3.5 Crystallisation of TraG*His	55
3.6 CD of TraG* Variants Provides Evidence of an N-terminal IDR	59

3.7 Analysis of TraG* by Global HDX MS Confirms Dynamicity in its N-terminal Residues	64
Chapter 4 – Conclusions and Future Work	
4.1 The N-terminal IDR of TraG* and its Potential Role for Protein Function	69
4.2 Future Work	72
Chapter 5 – References	
Appendices	
Appendix A	89
Appendix B	93

List of Tables

Table 1.1 A Comprehensive List of the TraG* Protein Variants used in this Study.....	27
Table 2.1 Primers Used in the Creation of Plasmid Constructs.....	29
Table 3.1 Secondary Structure of HisTraG* and HisTraG* _Δ	61

List of Figures

Figure 1.1 Common mechanisms of Horizontal Gene Transfer.....	2
Figure 1.2 The Structure of Canonical Gram-Negative T4SS and a Representation of the Eex Mechanism.....	7
Figure 1.3 The Anatomy of the F plasmid.....	10
Figure 1.4 Genetic Organization of the Exclusion Gene-Containing Regions.....	15
Figure 1.5 Sequence Alignment of Entry Exclusion Proteins.....	17
Figure 1.6 The Paradigmatic Continuum of Protein Structure.....	22
Figure 1.7 Schematic of TraG _F , TraG* and TraG* _Δ	27
Figure 3.1 A Predictive Model of TraG by Phyre2.....	41
Figure 3.2 Software Mediated Identification of IDRs in TraG from the F plasmid.....	44
Figure 3.3 Cloning of TraG Plasmid Constructs.....	46
Figure 3.4 Optimized Ni ²⁺ IMAC Purification Schemes of TraG* Variants.....	47
Figure 3.5 Optimized Ni ²⁺ IMAC Purification Schemes of TraG* _Δ Variants.....	49
Figure 3.6 Thermofluor Analysis of TraG*His and HisTraG* _Δ	53
Figure 3.7 Optimization of Buffer Conditions for TraG* Stability using Thermofluor.....	54
Figure 3.8 Purified TraG* Constructs and TraG*His Crystals.....	57
Figure 3.9 X-Ray Diffraction of TraG*His Spherulites.....	58
Figure 3.10 The CD spectra of HisTraG* and HisTraG* _Δ	60
Figure 3.11 Urea denaturation of HisTraG* and HisTraG* _Δ as Measured by CD Spectroscopy.....	62
Figure 3.12 Denaturation Experiments Sampling Refined Concentrations Between 3-4M Urea.....	63

Figure 3.13 ESI-MS analysis of native folds for HisTraG* and HisTraG* Δ	66
Figure 3.14 Global HDX-MS analysis of HisTraG* and HisTraG* Δ	67

List of Abbreviations

aa	Amino acid	HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
ATP	Adenosine Triphosphate	HGT	Horizontal Gene Transfer
BCA	Bicinchoninic Acid	Hfr	High frequency of recombination
Bp	Base pairs	JCSG	Joint Center for Structural Genomics
BSA	Bovine Serum Albumin	IDR	Intrinsically Disordered Region
CD	Circular Dichroism	IDT	Integrated DNA Technologies
CHES	N-Cyclohexyl-2-aminoethanesulfonic acid	IM	Inner Membrane
CIP	Calf Intestinal Phosphatase	Imid	Imidazole
Da	Daltons	Inc	Inclusion
DD	Double Digest	IPTG	Isopropyl- β -D-thiogalactoside
DNA	Deoxyribonucleic acid	Km	Kanamycin
		LB	Luria Bertani
Eex	Entry Exclusion	LPS	Lipopolysaccharide
EI	Exclusion Index	MCS	Multiple Cloning Site
fin	Fertility Inhibition	MCSG	Midwest Center for Structural Genomics
FPLC	Fast Protein Liquid Chromatography	MDR	Multiple Drug Resistant
GE	General Electric	MES	2-ethanesulfonic acid
GST	Glutathione-S-Transferase	MIR	Multiple Isomorphous Replacement
HDX	Hydrogen-Deuterium Exchange		

MOPS	3-(N-morpholino)propanesulfonic acid	PPI	Protein-Protein Interaction
Mpf	Mating pair formation	PSP	Phase Separator Predictor
Mps	Mating pair stabilization	RNA	Ribonucleic acid
MS	Mass Spectrometry	rpm	Rotations per minute
MWCO	Molecular Weight Cut-Off	SAD	Single-Wavelength Anomalous Diffraction
m/z	Mass-to-Charge Ratio	SDS	Sodium Dodecyl Sulfate
NaAc	Sodium Acetate	PAGE	Polyacrylamide gel electrophoresis
NEB	New England Biolabs	Sfx	Surface Exclusion
NTA	Nitrilotriacetic acid	Spp.	Several Species
OD	Optical Density	tra	Transfer
OM	Outer Membrane	T_m	Melting Temperature
ompA	Outer Membrane Protein A	TRESI	Time Resolved Electrospray Ionization
oriT	Origin of Transfer	Tris	Tris(hydroxymethyl)aminomethane
PBS	Phosphate Buffered Saline	T4SS	Type 4 Secretion System
PCR	Polymerase Chain Reaction	UV	Ultraviolet
PDB	Protein Databank	Vir	Virulence
PEG	Polyethylene Glycol		
PMSF	Phenylmethylsulfonyl fluoride		

CHAPTER 1.0 – INTRODUCTION

1.1 The Mounting Threat of Bacterial Antibiotic Resistance as Aided by Horizontal Gene Transfer

Multidrug-resistant (MDR) bacterial pathogens; classified as strains that exhibit a phenotype with resistance to three or more antibiotic classes, are a substantial threat to public health as the number of accessible antibiotics that can efficiently eliminate infections of MDR species are perilously limited (Thomson and Bonomo, 2005). Increasing numbers of these species have been observed globally, largely aided by poor nosocomial hygiene, excessive use of antibiotics in animal agriculture, and over-prescription of broad spectrum antibiotics (Chang et al., 2015; Pelat et al., 2015). As well, it is estimated that 23,000 deaths in the U.S.A. alone are attributed to infections by MDR bacteria annually (Chang et al., 2015).

Genes conferring antibiotic resistance are evolved and persist as a result of selective pressures in bacterial populations; rapid cellular reproduction and the plasticity of the bacterial genome allows for their propagation through colonies of these pathogens by methods of DNA replication such as binary fission and methods of DNA transmission through horizontal gene transfer (HGT) (Allen et al., 2010). HGT allows for the accelerated dissemination of advantageous genes independent of the speed of replication. Beneficial genes would include those that code for drug efflux pumps, antibiotic inactivating enzymes, altered drug targets and many other virulence factors (Figure 1.1) (Sultan et al., 2018). HGT methods include transformation through the import of extracellular DNA, viral transduction through the successful entry and recombination of bacteriophage-packaged genes, and conjugation (Gogarten and Townsend, 2005; Gyles and Boerlin, 2014). Genes are transferred by bacteria in a donor-controlled fashion using conjugation, in which mobile genetic elements such as plasmids or chromosomally integrated conjugative

elements (ICEs) are replicated in a donor cell and transported into a recipient cell using a Type IV secretion system (T4SS) (Bennett, 2008). As conjugation systems are ubiquitous in the bacterial kingdom, inhibition of plasmid conjugation has been suggested as a tool to prevent the spread of antibiotic resistance genes, thus preventing the proliferation of MDR pathogens (Baron, 2013, 2010; Williams and Hergenrother, 2008).

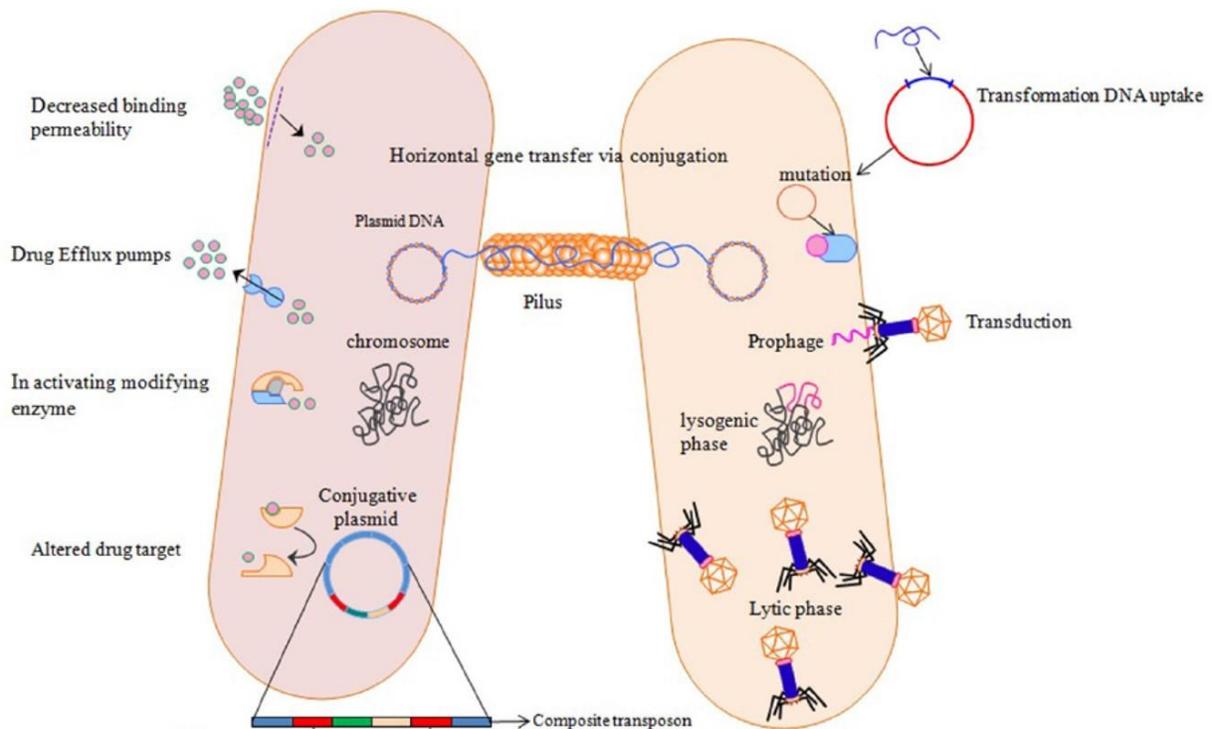


Figure 1.1: Common mechanisms of acquiring and transmitting genetic material in bacterial species and examples of the encoded functions of these genes. The resistance mechanisms shown above include modifications of the bacterial envelope components to alter permeability of drugs, substrate efflux pumps, enzymes that modify and inactivate antimicrobial molecules, and the production of an altered homologue of a drug target (Sultan et al., 2018). The acquisition of these genes occurs through either mutation or HGT methods such as the transformation of extracellular DNA, transduction of viral DNA and integration into the bacterial chromosome, and conjugative transfer of a plasmid through a pilus (Gogarten and Townsend, 2005; Gyles and Boerlin, 2014). This transfer of DNA through the pilus is a common model for conjugation by the T4SS (Costa et al., 2016). Image adapted from (Sultan et al., 2018).

1.2 Evidence for Urgency in Studying Gram-Negative Bacteria

Gram-negative bacteria differ from gram positive bacteria in the composition of their cell wall; the classification of species into these groups is classically based on Gram Staining techniques through the binding of peptidoglycan by hexamethyl pararosaniline chloride (crystal violet) (Gregersen, 1978). The bacterial envelope of gram-negative species consists of an outer lipid bilayer with bound lipopolysaccharides (LPS), a periplasm consisting of peptidoglycan, and an inner cytoplasmic membrane (IM). Gram positive species lack an outer membrane (OM); instead they feature a cell wall composed of peptidoglycan with lipoteichoic acid and a cytoplasmic lipid membrane. The presence of an OM provides gram-negative species with an added permeability barrier relative to gram positive bacteria, resulting in differences in the penetration and retention of molecules (Grohmann et al., 2003). This provides gram-negative bacteria a degree of intrinsic antibiotic resistance, as some drugs and antibiotics effective against gram positive bacteria cannot permeate the OM of gram-negative species.

There are a number of medically relevant gram-negative bacterial species proving to be hazards to public health, including *Acinetobacter* spp., *Campylobacter* spp., *Helicobacter pylori*, *Haemophilus influenzae*, *Pseudomonas aeruginosa* and a number of *Enterobacteriaceae* spp. (Exner et al., 2017; Lermينياux and Cameron, 2018). All of these species and more have been found as MDR strains resistant against ureidopenicillins, third or fourth generation cephalosporins, carbapenems, fluoroquinolones, and other antibiotics normally effective to these species (Exner et al., 2017). As these antibiotics, in particular carbapenems, are typically used as the last line of effective treatment for infections with multidrug resistant gram-negative species, the emergence of human pathogens resistant to them has been deemed a critical clinical issue by the World Health Organization (WHO) (Exner et al., 2017; Lermينياux and Cameron, 2018). Exasperating the

situation, the mobile gene *mcr-1* has been reported in microbiota of farm animals from China, the U.K., Denmark, the U.S. and Germany (Exner et al., 2017). This gene provides resistance against the polypeptide colistin, a reserve antibiotic for the treatment of critically ill patients in instances of infection by carbapenem MDR gram-negative bacteria; novel antibiotics are desperately needed.

Many efforts have been made in combatting MDR gram positive bacteria; novel antimicrobial agents and antibiotics from new classes have been developed to target these species (Giske et al., 2008). However less progress has been made in the development of novel drugs effective in targeting MDR gram-negative pathogens. Therefore, there is a vital need for further research into the modes of antibiotic resistance in gram-negative bacteria or humanity will regress to a state mirroring that of pre-antibiotic eras.

1.3 Gram-Negative Bacterial Conjugation by T4SS

1.3.1 The Utility of Conjugative T4SS

Conjugation in gram-negative bacteria is enabled by T4SS; large multi-protein complexes that span inner and outer cellular membranes and are the most ubiquitous secretion system in nature (Costa et al., 2015). Their prevalence throughout the bacterial kingdom is due to the evolution of the conjugal T4SS in a common ancestor; natural selection has favoured conjugation, as donor-mediated HGT appears to be a necessity for survival based on the requirement for rapid evolution in the competitive environment of unicellular organisms (Vogan and Higgs, 2011). As well, conjugation has been observed to occur between bacterial species to transmit plasmids with beneficial genes across broad host ranges (Grohmann et al., 2003; Lerminiaux and Cameron,

2018). As conjugation is arguably the most common method by which antibiotic resistance genes are transmitted, a thorough understanding of the conjugation process is paramount in the development of novel drugs to target the T4SS and prevent dissemination of antibiotic resistance (Cabezón et al., 2017; Lujan et al., 2007; Waksman, 2019).

1.3.2 Initiation of Conjugation, Pilus Assembly and Extension

As the conjugative transfer of DNA is a metabolically costly process for bacterial cells, the activation of T4SSs are strictly regulated (Koraimann and Wagner, 2014). In the production of conjugative systems, gene expression is limited by one or both of the following mechanisms: signalling molecules (autoinducers or sex pheromones), and low constitutive gene expression (Kohler et al., 2019). The regulation of conjugation in plasmids from the IncF inclusion family occurs solely by the latter process and was termed fertility inhibition (*fin*) (Koraimann et al., 1996). The gene coding for the activator of the transfer (*tra*) operon expression, *traJ*, is controlled by *finOP* regulatory genes. The downregulation of TraJ expression is controlled by the antisense RNA FinP, and the protein FinO upregulates the expression of FinP to reduce F transfer 10 - 2000 fold (Koraimann et al., 1996). The representative member of the IncF family, the F plasmid, is classified as 'derepressed' as it has an insertion sequence 3 (IS3) element in the *finO* gene, and the *finP* gene has extremely low expression (Frost et al., 1994). This results in high constitutive expression and perpetual transfer of the F plasmid. The T4SS of the F plasmid and its conjugative process will be detailed in the following sections.

In bacterial mating, conjugative pili are assembled and extended from T4SS to bind a neighbouring cell, then retracted to bring the potential recipient cell in close proximity to the donor

cell such that a mating bridge can be formed (Schröder and Lanka, 2005). Pilus assembly requires the processing of propilin into pilin, TraA, through the cleavage of a ~50 amino acid (aa) leader peptide by Leader peptidase B and the subsequent acetylation of the N-terminus by TraX (Arutyunov and Frost, 2013). TraQ acts as a chaperone and allows for the correct insertion and accumulation of TraA in the inner membrane to prime the protein subunits for assembly (Lawley et al., 2003). TraQ and TraX are subunits specific to the F pilus. A depiction of the F pilus displaying the T4SS subunits is shown in Figure 1.2A.

A multitude of T4SS auxiliary proteins perform the assembly of the pilus tip; the ATPase TraC is responsible for extracting the TraA monomers from the inner membrane and provides energy for their assembly (Kerr and Christie, 2010). Other proteins known to be responsible for pilus tip assembly include TraL, TraE, TraK and the N-terminal half of TraG; where one or more of these subunits are predicted to act as a chaperone to attach a molecule of IM phospholipid to each assembling TraA subunit (Costa et al., 2016; Hu et al., 2019). Polymerization of the pilus occurs in a stoichiometric complex with phospholipid molecules and forms a five-start helical assembly, where each strand is composed of 12.8 subunits per helical turn (Costa et al., 2016). This creates a protein nanotube with an outer diameter of 87Å and an inner diameter of 28Å. To promote assembly of TraA subunits from the base of the pilus and extend the nanotubule, the structural stalk and channel proteins TraB, TraK and TraV are required (Lawley et al., 2003).

1.3.3 Pilus Retraction and Exclusion Events

When the pilus comes in contact and attaches to a neighbouring cell, a signalling event occurs and the retraction of this pilus is performed by subunits TraF, TraH, TrbI, and interacting

partners TraW and TrbC (Arutyunov and Frost, 2013; Lawley et al., 2003; Shala-Lawrence et al., 2018). Many of these proteins have multiple proposed functions; TrbB and TraF both contain thioredoxin-like domains and are thought to be responsible for chaperone activity, and aid in the correct conformational folding of TraH, TraU, and TraN into the transferosome complex in the final mating pair formation step (Arutyunov et al., 2010; Lento et al., 2016).

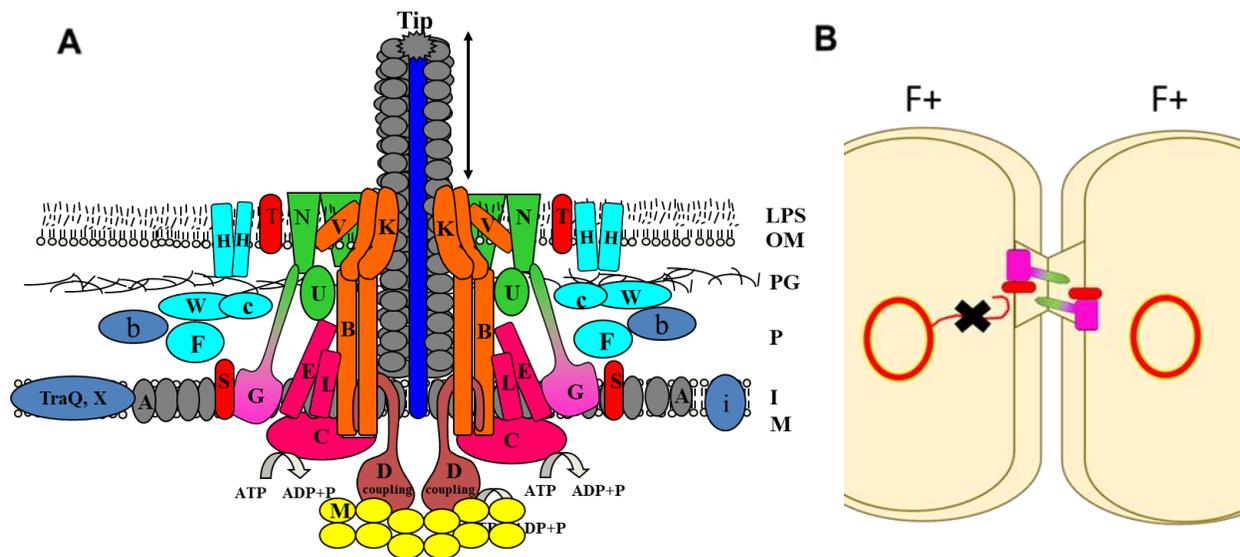


Figure 1.2. **A)** The structure of the canonical gram-negative T4SS, the F pilus. Components are colour coded based on their function; proteins in fuchsia are responsible for pilus assembly, orange-labeled proteins are structural, those in dark blue are for pilin processing, in light blue are components responsible for pilus retraction and mating pair formation, mating pair stabilization proteins are in green, and exclusion proteins are in red (Arutyunov and Frost, 2013). The N-terminal transmembrane region of TraG is responsible for pilus assembly, however the protein as a whole is considered to be important for mating pair stabilisation; the periplasmic C-terminus is predicted to interact with TraN to aid in the stabilization of the conjugative pore (Klimke and Frost, 1998). Adapted from (Lawley et al., 2003). **B)** A basic representation of the proposed mechanism by which entry exclusion functions in preventing donor-donor transfer of plasmids with identical exclusion genes. If neighbouring bacteria hold the same plasmid they will have cognate TraS and TraG in their inner membranes (Audette et al., 2007). As conjugation proceeds, if surface exclusion malfunctions then entry exclusion is essential in preventing wasteful donor-donor gene transfer; the C-terminal region of TraG interacts with its cognate TraS on the neighbouring cell's inner membrane, preventing conjugative transfer and separating the mating junction.

After a potential recipient cell is brought in proximity to the donor, mating pair stabilization (Mps) occurs. Mps is a multistep process where the OM of neighbouring cells come in close contact and exclusion events occur to prevent donor-donor transfer (Arutyunov and Frost, 2013). The first step in Mps involves interaction between TraN on the OM of the donor cell with OM protein A (OmpA) and LPS moieties on the recipient cell to maintain close contact during conjugation (Klimke and Frost, 1998). Next, surface exclusion (Sfx) occurs as mediated by the OM lipoprotein TraT; if cognate TraT is expressed in the recipient cell's membrane indicating the presence of the same plasmid in the recipient cell, then an interacting partner (currently not established) is theorized to prevent mating pore formation and conjugation will be prevented (Garcillán-Barcia and de la Cruz, 2008). If this quality control step fails, then entry exclusion (Eex) is relied upon to prevent donor-donor DNA transfer. The IM protein TraS and the periplasmic C-terminal half of TraG are interacting partners in this event; if the potential recipient has the same plasmid TraG in the donor cell is thought to scan the IM of the recipient cell and interact with the cognate TraS to arrest conjugation (Figure 1.2B)(Audette et al., 2007). If the recipient cell does not have a cognate TraS in its IM, and thus does not have the same plasmid, the C-terminus of TraG is predicted to extend to contact TraN in the OM and further the formation of the mating junction (Klimke and Frost, 1998).

1.3.4 Conjugative DNA Transfer

As part of the conjugation process, the DNA to be transferred is processed through the formation of a nucleoprotein complex using TraM and the relaxase TraI, which nicks circularized plasmid DNA at the origin of transfer (oriT) to form a single stranded DNA (ssDNA) topology. TraM binds as a tetramer to aid in DNA unwinding and chaperones the complex to the IM mating

pore formed by TraD, a hexameric ring ATPase that pumps the nucleoprotein complex through the pore into the recipient cell when all other T4SS have performed their function for mating pair formation (Mpf) (Lu et al., 2008; Peng et al., 2014). Rolling circle replication allows for asymmetric replication of the ssDNA in the donor cell, thus reforming the plasmid while the mobile ssDNA is being transferred to the recipient cell (Wawrzyniak et al., 2017). In the recipient cytoplasm the nucleoprotein complex is removed and the ssDNA is recircularized and replicated, thus converting the recipient cell into a donor for the newly integrated genes (Arutyunov and Frost, 2013; Guglielmini et al., 2014).

1.4 The F plasmid and F pilus

The Fertility (F) plasmid was the first factor confirmed by Joshua Lederberg and Edward Tatum to be responsible for bacterial conjugation in the model organism *Escherichia coli* K-12, indicating the prevalence of the F plasmid in molecular biology (Tatum and Lederberg, 1947). As in all conjugative plasmids, the F plasmid contains all genes an F⁺ cell requires to properly perform a donor-recipient conjugative DNA transfer. In related plasmids these genes are titled *vir* for virulence, however in the F plasmid they are the *tra* (for transfer) genes (Fernandez-Lopez et al., 2016). The nomenclature difference results from the distinction of F plasmid from other sex-related plasmids, such as those that code for P, N, W, or X pili. The genetic region involved in F conjugation is significantly longer (~34kb) and contains more genes than those of VirB/D4-like pili forming plasmids (~15 kb) (Lawley et al., 2003). Additionally, all *tra* genes are transcribed from a single promoter, making the 34kb *tra* operon the longest transcript found in *E. coli*. This transcript consists of genes coding for the core complex proteins required for conjugation; including DNA transfer proteins such as the relaxase and chaperone proteins that form the

nucleoprotein complex with the ssDNA to cross the mating bridge, as well as the components of the conjugative pilus and the pumping platform for the nucleoprotein complex (Arutyunov and Frost, 2013). Other proteins include those responsible for Mps, Sfx to prevent contact of two F⁺ cells, and Eex to prevent donor-donor transfer. An image of the genetic order of the F plasmid is shown in Figure 1.3 with a more detailed view of the *tra* region.

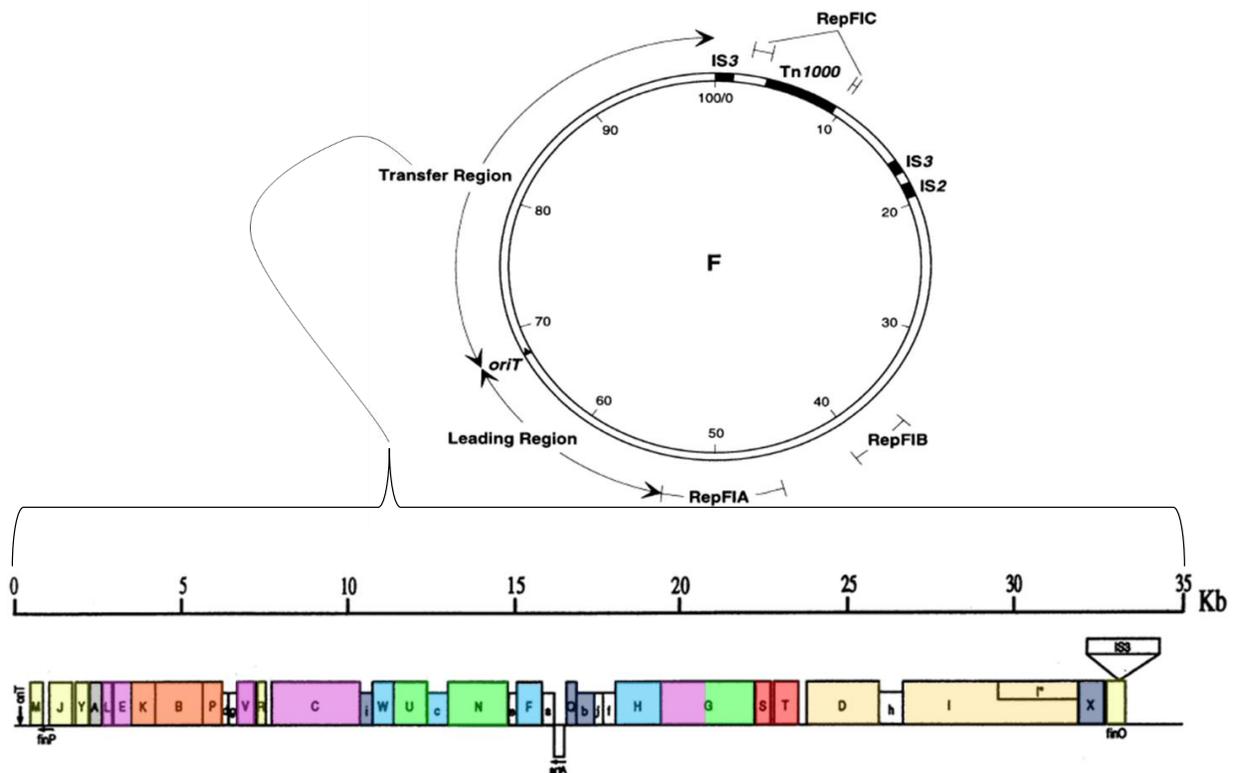


Figure 1.3: The anatomy of the F plasmid, approximating positions of important genetic markers in a 100kb map. The origin of transfer (*oriT*) represents the position where TraI nicks the plasmid, the small black triangle represents the direction of ssDNA transfer. Insertion sites 2 and 3 (IS2 & 3) and Tn1000 represented by black boxes are simple and composite transposable elements, respectively (Roberts et al., 2008). RepFIA and B are replicons which function as origins of replication, while the FIC replicon is non-functional (Villa et al., 2010). The leading region contains many genes involved in the regulation of the *tra* genes and their gene products; for more information see (Manwaring et al., 1999). The transfer region is the 34kb *tra* operon; all *tra* and *trb* genes are shown, as well as antisense genes *artA* and *finP* (Frost et al., 1994). I* represents a gene product predicted to be expressed within the *traI* gene, but later this gene region was determined to be a separate domain of the TraI protein responsible for DNA transfer activity (Matson and Ragonese, 2005). The colours are coded in the same fashion as in Figure 1.2A; however yellow genes are those involved in conjugation signalling, origin nicking, relaxase activity, and T-DNA transport. White genes are those with no known function to date. Figure adapted from (Firth et al., 1996; Laura S Frost et al., 1994).

The pili assembled by T4SSs can be categorized by their size and rigidity, which typically dictates their function (Arutyunov and Frost, 2013; Bradley, 1980). Thin flexible pili confer Mps, while rigid pili are encoded by T4SS that have an additional ATPase, VirB11 (orthologous to TrbB), and perform Mpf but do not perform Mps or pilus retraction (Arutyunov and Frost, 2013). Thick flexible pili are characteristic of the more complex T4SS that are involved in pilus retraction, Mps, and Mpf. The F pilus is the representative member of the thick, flexible T4SS pili, and the plasmid encoding it is the best characterized conjugative plasmid in gram-negative bacteria (Arutyunov and Frost, 2013; Bradley, 1980). Therefore, the F pilus is an ideal T4SS to study for obtaining a representative view of conjugation in gram-negative bacteria and to understand their mechanisms of pathogenesis.

1.5 Entry Exclusion and Incompatibility Groupings

1.5.1 The Discovery of Superinfection Immunity

Eex is an important function in conjugative systems to prevent the inefficient transfer of a plasmid from a donor cell to another cell that has the same plasmid (Frost et al., 1994). After Sfx is permitted and cells come in close proximity, Eex is performed before Mpf and completion of the mating bridge. TraG in the donor cell is thought to scan the IM of the recipient cell and interact with the cognate TraS to arrest donor-donor conjugation (Audette et al., 2007). This is performed in a plasmid specific manner; the TraG proteins of the closely related F and R100 transfer apparatuses present in donor cells will recognize their cognate TraS in a recipient cell but will not recognize TraS from a homologous plasmid. The result of this phenomenon was seen early in the history of conjugation research; Lederberg *et al.* discovered that cells harbouring the F plasmid

were poor F conjugation recipients, and later the inefficient transfer was termed 'superinfection immunity' (Lederberg et al., 1952; Watanabe, 1963). Superinfection refers to the process of lethal zygosis, a phenomenon that was seen when an excess of High frequency of recombination (Hfr) F donors were placed in contact with F⁻ recipients; many of the recipient cells died (Skurray and Reeves, 1973). The primary cause of cell death was determined to be extensive damage in the recipient cell membrane resulting in cell permeability due to extensive cell-cell contacts mediated by pili during conjugation (Ou, 1980; Viljanen, 1987). Metabolic costs due to excessive DNA transfer contribute to lethal zygosis in a lesser extent. F-mediated superinfection immunity is attributed to exclusion genes *traS* and *traT*, and the exclusion process is essential in preventing lethal zygosis through excessive mating (Ou, 1980; Skurray and Reeves, 1974).

There exists an association between exclusion and sex pilus type; sex factors carrying the same sex pilus type would be excluded, however plasmids with different sex factors permit conjugation even if the pilus type is identical (Edwards and Meynell, 1968; Meynell and Ewins, 1973). When the plasmid was transferred to a donor by transduction this phenomena did not occur in cells despite an active entry exclusion system, however even when entry exclusion was abolished it was observed that any second F factor in an F-containing strain would be removed (Dubnau and Maas, 1968; Watanabe et al., 1968). Therefore it was concluded that two phenomenon contribute to superinfection immunity: plasmid incompatibility and entry exclusion (Garcillán-Barcia and de la Cruz, 2008; Novick, 1969; Watanabe et al., 1968) .

1.5.2 Plasmid Incompatibility

Incompatibility in mobilizable elements can be defined as the inability of two plasmids to be propagated stably in the same cell line (Couturier et al., 1988). It is a manifestation of relatedness; elements involved in plasmid replication control and partitioning are identical, causing malfunction in the inheritance of both plasmids into the daughter cells. Incompatibility is also linked to copy number, as the number of origins of replication of the same grouping affects whether they are replicated (Novick, 1987). Speed of replication and toxicity of the plasmid also affects the survivability of these plasmids, as outgrowth during replication is theorized as the mechanism for the loss of one plasmid in a similar pair (Velappan et al., 2007). As incompatibility is a universal property of plasmids, it allows for the formation of a suitable classification scheme (Carattoli et al., 2005; Novick, 1987). The main families are considered to be IncF, IncP, IncA/C, IncH, IncI, and IncN based on the plasmid multilocus sequence typing database (Jolley and Maiden, 2010). An indeterminate multitude of subfamilies exist due to the continuous discovery and synthesis of novel plasmids.

1.5.3 Entry Exclusion in Different Incompatibility Groups

Plasmid inclusion families are seen to have differences in their exclusion systems in terms of gene organization and the structure of the encoded proteins (Garcillán-Barcia and de la Cruz, 2008). The previously described exclusion system is that of the IncF-like group, and the genetic organisation for Eex and Sfx genes on these plasmids follow the order seen in Figure 1.4. In comparing the genetic organisation of plasmid groups from gram-negative bacteria, synteny (the physical localisation of genetic loci) is maintained for Eex genes in relation to other transfer-related

genes (Garcillán-Barcia and de la Cruz, 2008). Plasmid families have some differences in their genetic organisation; many have their *traS*-like gene before their *traG*-like gene in the locus, and most plasmids do not have a *traT*-like Sfx gene at all. Many groupings lack a *traG*-like gene for exclusion as well; IncI-like and ICEs R391 and SXT only have a *traS*-like exclusion gene, and the pheromone responding conjugative elements have their own type of exclusion system. Plasmids most commonly found in gram-negative bacteria are from the inclusion families shown to have both TraG- and TraS- like systems; IncF-like, IncP-like, and IncN-like (includes IncW)(Fernandez-Lopez et al., 2016). There exists the potential for evolutionary selection for TraG/TraS-like Eex systems against TraT-like Sfx systems, as supported by the observed increase in exclusion index (EI) in Eex in comparison to Sfx (Garcillán-Barcia and de la Cruz, 2008). EI is calculated as the frequency of transfer of a plasmid given to a plasmid-free recipient divided by the frequency of transfer to a recipient carrying the same plasmid; for the F-plasmid in *E. coli* the EI was 100-300 because mating was observed at 100-300 times higher frequency when the plasmid was not present in the recipient in comparison to donor-donor exchange (Achtman et al., 1977). In comparing EI of donor-recipient exchanges for the F plasmid, *traT* point mutants showed a smaller reduction in mating than the *traS* point mutants (Achtman et al., 1980). An increased exclusion activity was seen in the Eex system, which was attributed an EI of 200, while an EI of 20 was associated to the Sfx system. indicating a higher reliance on the Eex system for preventing erroneous donor-donor transfers. The exclusion process performed by these systems is also seen to be gene-dose dependent; when *traS* and *traT* were cloned into a multicopy plasmid the EI increased to 10,000 (Achtman et al., 1977).

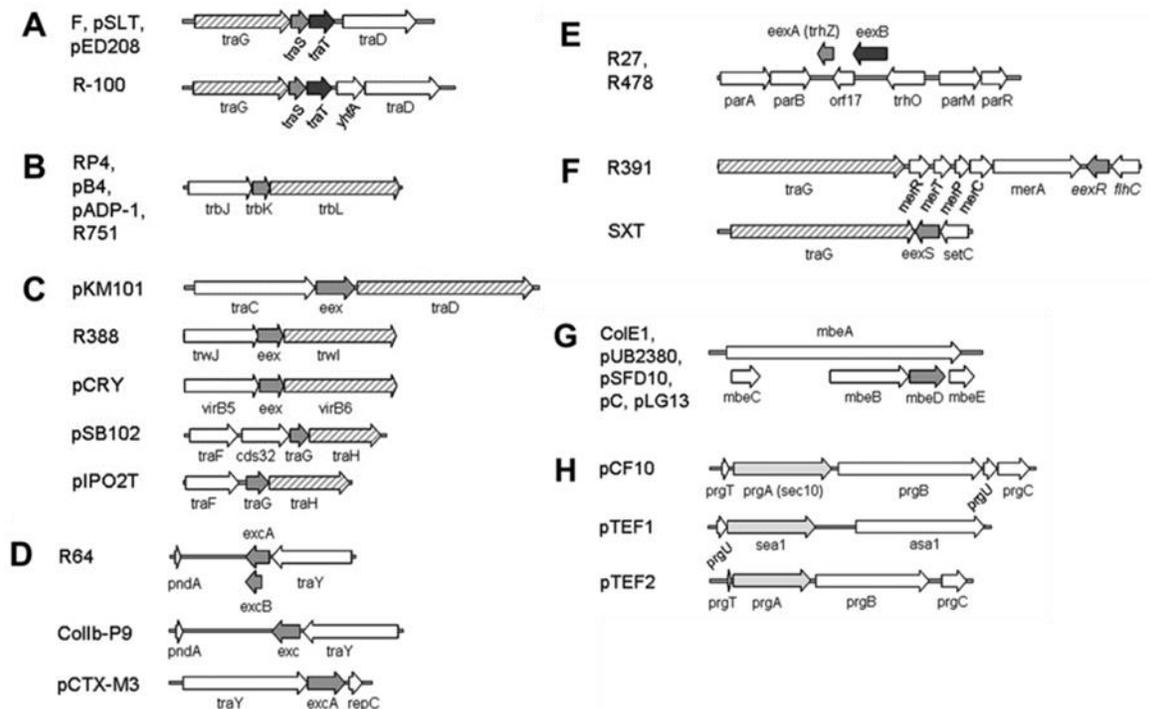


Figure 1.4: Genetic organization of the exclusion gene-containing regions in the plasmid incompatibility groups: **A)** IncF-like; **B)** IncP-like; **C)** IncN-like; **D)** IncI-like; **E)** IncH-like; **F)** ICEs R391 and SXT; **G)** ColE1-like; **H)** Pheromone-responding conjugative plasmids, more commonly found in gram positive bacteria. Arrows corresponding to the direction of transcription in exclusion genes are shadowed depending on the system; light gray is for exclusion genes of pheromone-responding conjugative plasmids, intermediate gray for genes encoding inner membrane exclusion proteins similar to *traS* and dark gray representing genes encoding outer membrane exclusion proteins similar to *traT*. *traG* homologs are dashed, while bordering genes with different functions are filled in white. Figure from (Garcillán-Barcia and de la Cruz, 2008).

In comparing *traS* genes for different incompatibility groupings, low sequence homology is observed even in members of the same inclusion family. TraT from the plasmid pED208 (TraT_{pED208}) of the IncF family has the lowest inter-family protein homology to TraT_F at 80% aa identity, however TraS_{pED208} has no identity to TraS_F or TraS_{R100} (Garcillán-Barcia and de la Cruz, 2008). Even the closely related F and R100 plasmids show a TraS homology of only 17% with no region displaying greater identity (Figure 1.5A)(Audette et al., 2007).

Exclusion genes have proven to be important for the stability of a conjugative plasmid; F mutants with both *traS* and *traT* mutations are unstable and have not been isolated (Achtman et al., 1980, 1977). This is supported by F plasmid-mediated superinfection immunity, which was determined to be mediated by TraT and TraS (Ou, 1980). The presence of TraS however was deemed to be more important for colony survival than TraT (Garcillán-Barcia and de la Cruz, 2008; Ou, 1980). As Eex by TraS and TraG also provide a higher EI than Sfx by TraT, it is theorized that the disruption of Eex systems would be severely detrimental to bacterial colony survival (Garcillán-Barcia and de la Cruz, 2008).

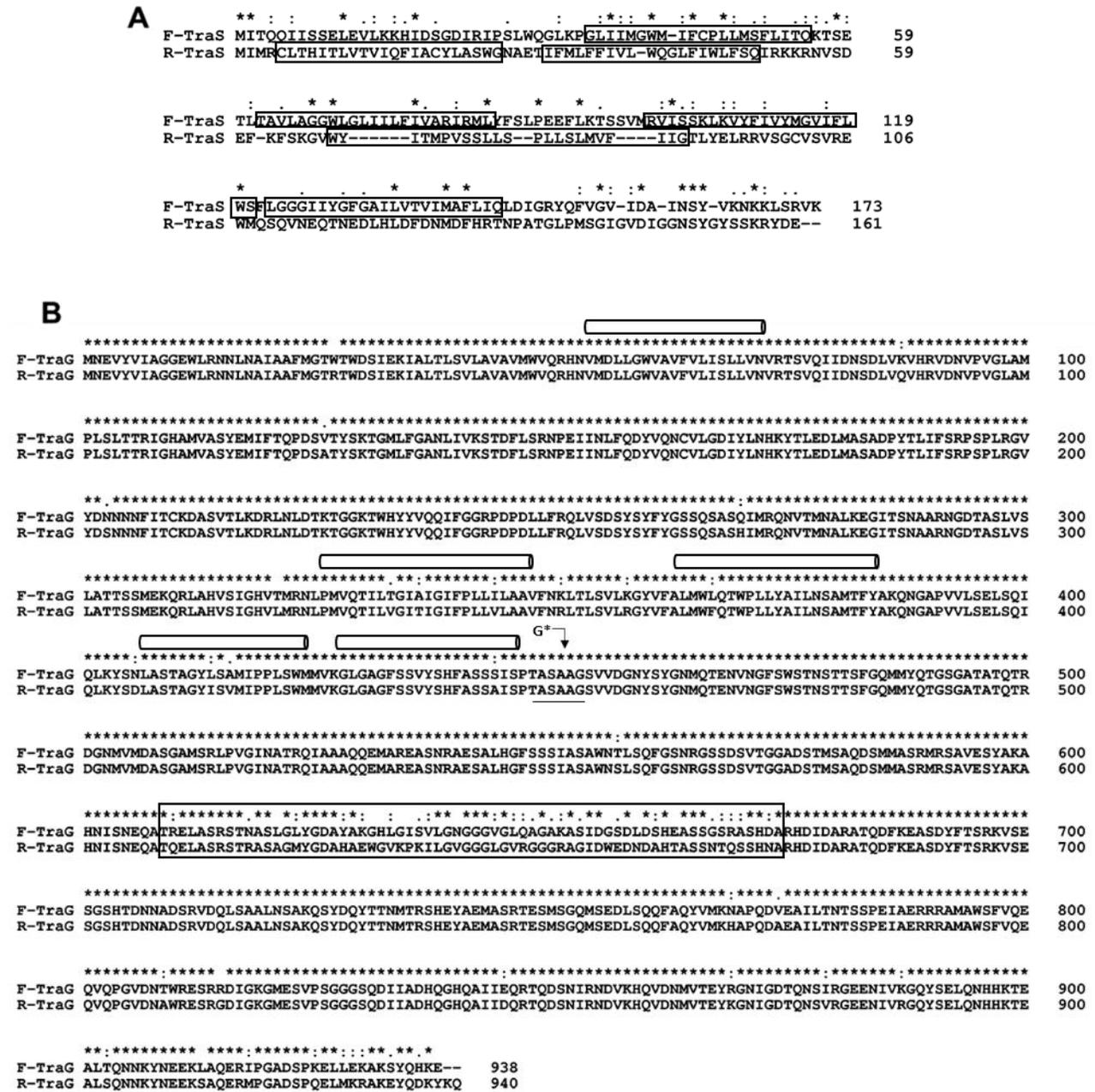


Figure 1.5: Sequence alignment of entry exclusion proteins **A)** TraS and **B)** TraG in related F-like plasmids F and R100 as aligned by Clustal W (Thompson et al., 1994). Strictly, highly and moderately conserved residues are indicated by the symbols ‘*’, ‘:’ and ‘.’, respectively. TraS proteins have an overall sequence identity of 17 %, and the predicted transmembrane helices are boxed (Audette et al., 2007). For TraG, the region between residues 610 and 673 (boxed) displays the greatest dissimilarity between the two proteins, at 55.7% identity versus 93% overall identity between the two proteins. Five transmembrane helices (aa 54–73, 327–348, 364–385, 407–426 and 428–447) in the N-terminus are indicated by cylinders above the sequences. The predicted signal peptidase cleavage site is at residue 451, and the residues forming the –3 to +1 peptidase region (aa 449–452) are underlined. Images adapted from (Audette et al., 2007).

1.6 Characteristics of TraG and TraG* from the F plasmid

1.6.1 Insights on TraG Structure and Function

F TraG (TraG_F) is a 102.5 kDa protein with an N-terminal transmembrane domain bound in the IM and a periplasmic C-terminal domain from residues A452-E939 denoted TraG* (Audette et al., 2007). It is one of the largest proteins encoded by the *tra* region of the F plasmid. Frameshift mutations in the C-terminal domain affect Mps while frameshift mutations in the N-terminus affect the polymerisation of pilus subunits (Willetts and Achtman, 1972). It was thought that the presence of a signal peptidase I cleavage site after residue A451 releases TraG* into the periplasm to fulfill its role in Mps (Firth and Skurray, 1992). However TraG_F must have both domains intact in order for Mps to occur (Audette et al., 2007). This indicates that TraG* does not become cleaved to function as a separate protein, and therefore there is a strong dependence on intact TraG for performing Mps and Eex. In performing Eex, TraG is predicted to scan the neighbouring recipient IM to interact with TraS, which would require the protein to interact over two layers of OM. This distance can be approximated to 35-40 nanometers (nm), assuming donor and recipient cell membranes are proximal after Sfx (Silhavy et al., 2010). As this distance is likely unachievable by a single globular protein, the C-terminal region of TraG is theorized to undergo a reversible conformational change and thrust into the mating pore formed after TraN interacts with OmpA of the recipient cell (Audette et al., 2007). In this predicted model TraG maintains an intact structure and extends TraG* through the mating pore into the periplasmic space of the recipient cell to perform its role in entry exclusion. A TraS-TraG complex has not been detected in mating cells through crosslinking or immunoprecipitation experiments, nor interaction detected using a bacterial two-hybrid system (Audette et al., 2007). As well, interaction between TraN and TraG has not been detected using similar methods (Klimke et al., 2005). The likely cause for the lack of

detectable assembly is due to the small number of interacting protein partners in a mating pair, as well as the transient nature of the protein-protein interactions (PPIs).

Audette and colleagues determined that there is a region in the aa sequence of TraG_F and TraG_{R100} that shows 55.7% similarity while the proteins hold an overall sequence identity of 93% (Figure 1.5B)(Audette et al., 2007). Residues T610-A673 were predicted to be responsible for the trans-exclusive interaction between TraG and a cognate TraS, as TraG will not interact with TraS in the same IM. The F and R100 TraS sequences have an overall sequence identity of 17% despite the overall homology of these plasmids. Lack of protein homology in the TraS family for the full-length protein, and in the TraG family for residues T610-A673, indicates that these proteins are responsible for plasmid specificity in the Eex process. As residues T610-A673 are within the domain of TraG termed TraG*, studying the structure of the periplasmic domain has the potential to elucidate the mechanism by which TraG interacts with the cognate TraS of a recipient cell to perform Eex.

1.6.2 Solved Structures of Proteins Analogous to TraG

Currently there is a lack of structural data for TraG*. In the IncN-type plasmids the model T4SS is the Ti pilus from *Agrobacterium tumefaciens*, largely due to the utility of this system in the genetic modification of plants mediated by the Ti plasmid (Christie, 2004; Valvekens et al., 1988). Figure 1.4C shows the plasmid pCRY (similar to the Ti plasmid) contains the gene product VirB6, a protein orthologous to TraG*. While there are no structures of VirB6 solved, the structure of VirB8 orthologous to the N-terminus of TraG has been solved in numerous species; *A. tumefaciens*, *Brucella suis*, *Bartonella quintana*, *Bartonella tribocorum*, and *Rickettsia typhi*

(Bailey et al., 2006; Gillespie et al., 2015; Smith et al., 2012). As well, TraH, a VirB8 homolog from the plasmid pIP501 (from IncN family, similar to pIPO2T) in *Enterococcus faecalis* was solved (Fercher et al., 2016). All of these structures feature a Nuclear Transport Factor 2-like fold and show high structural homology despite low sequence homology, with backbone root mean square deviation (rmsd) values ranging between 2.6–2.8 Å in comparing all structures (Fercher et al., 2016). These VirB8 structures indicate the protein functions as a dimer, however TraH is monomeric. However, a VirB8 homolog from pKM101, TraE, was shown to form hexamers when isolated and when interacting with the VirB6 homolog from this plasmid, TraD, thus implying the diversity in oligomerization possible in these systems (Casu et al., 2018). These studies indicate the potential for TraG to have an N-terminus similar in structure to VirB8 despite low sequence homology, and therefore may function as a monomer, a dimer or a hexamer. As TraG* is separated from this N-terminus there is the potential that the protein will not oligomerize without the membrane-bound half. There is evidence indicating the evolution of the *traG* gene in the F plasmid involved the fusing of an ancestral homolog of genes *virB6* and *virB8* to produce a single gene (Christie, 2016).

1.7 Intrinsic Disorder in Proteins

1.7.1 Common Functions of Dynamic Proteins

The importance of intrinsically disordered regions (IDRs) in proteins was understated for much of the history of structural biology (Uversky, 2013). The prominent view of protein function was that the amino acid sequence dictates a highly specific three dimensional structure inherent in its activity (Dunker et al., 2002). However disorder in a protein sequence plays an important role

in its function; disordered regions of proteins are commonly involved in PPIs, protein-nucleic acid binding, phosphorylation, or as flexible linkers (Dunker et al., 2002; Dyson and Wright, 2005; Reddy Chichili et al., 2013). A continuum to the extent of disorder intrinsic to a native protein structure is observed, from well-defined single domain folds, to multidomain proteins with disordered flexible regions, to disordered molten globules, and to extended random coils (Dyson and Wright, 2005; Mittag et al., 2010). Figure 1.6 displays examples of proteins in which the degree of intrinsic structural disorder is relevant to its function.

The occurrence of unstructured regions >50 residues is common in functional proteins (Dyson and Wright, 2005; Uversky, 2013, 2002). Intrinsically disordered proteins (IDPs) are more common in eukaryotes and viruses due to the requirement of complexity in their morphologies; it is suggested that there exists a link between intrinsic disorder and evolution (Xue et al., 2012). There are many examples of prokaryotic IDPs however, such as the proteins that regulate the assembly of large multiprotein complexes such as FlgE in the bacterial flagellum (Barker et al., 2017), and Ffh and FtsY of the ribosome (Dyson and Wright, 2005; Estrozi et al., 2011; Focia et al., 2004). In both cases there exists an intrinsically disordered region (IDR) that serves as a linker region for two independent functional domains required in achieving a variety of conformations for important PPIs. As well, several bacterial regulatory proteins have conserved short intrinsically disordered linker regions called Q-linkers (Dyson and Wright, 2005; Reddy Chichili et al., 2013; Wootton and Drummond, 1989). Therefore, it is not uncommon in bacterial proteins for disordered regions to be important flexible linkers of folded domains.

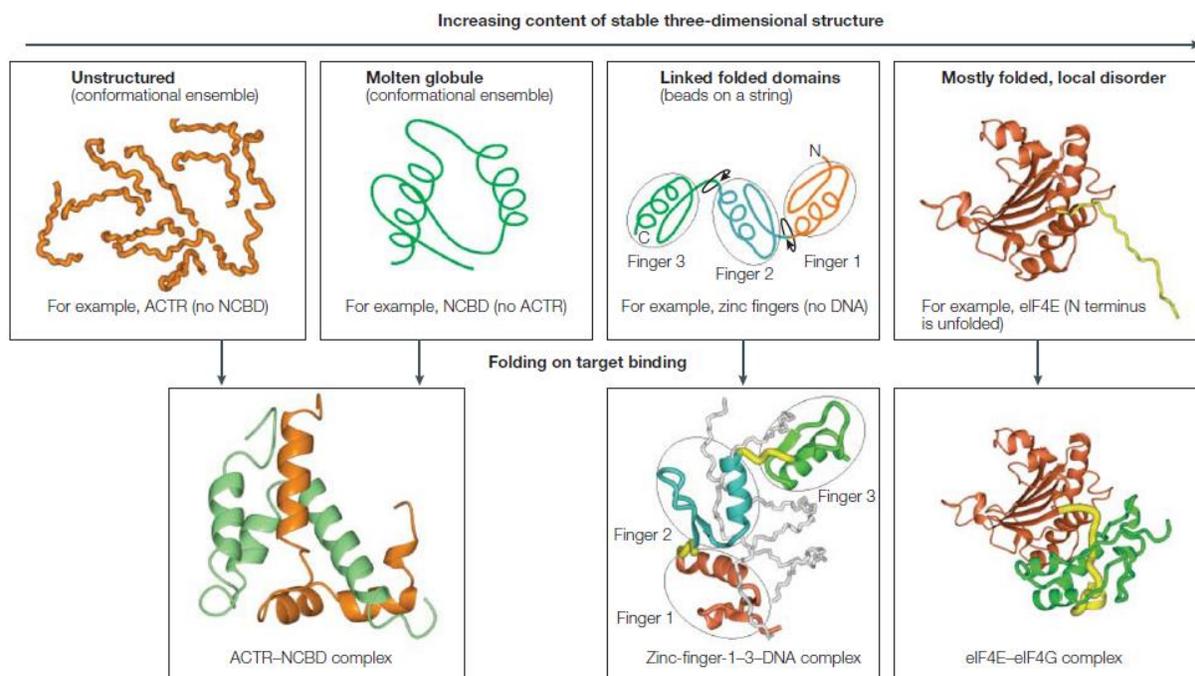


Figure 1.6: The top panels show the paradigmatic continuum of protein structure, with the degree of conformational order increasing from left to right, and examples therein. The activator for thyroid hormone and retinoid receptors (ACTR) is an example of an unstructured conformational ensemble and exists as random coils in solution (Martinez-Yamout et al., 2002). An example of the next milestone in protein dynamics is the molten globule nuclear receptor co-activator binding domain (NCBD) of the cyclic-AMP-responsive-element-binding protein (CREB)-binding protein (CBP). As seen in the bottom left panel, ACTR functions as a cryptic binding protein to form a mutually folded structural complex with the NCBD of CBP. The zinc finger of transcription factor IIA displays conformational freedom in linkages between folded domains, and finally the free eukaryotic translation initiation factor (eIF)4E displays high order with limited local disorder (Dyson and Wright, 2005; Gross et al., 2003). The remaining bottom two panels show the way these proteins interact with targets based on the stability of their three-dimensional structure. Figure from (Dyson and Wright, 2005).

1.7.2 Pi-Pi Interactions in Disordered Regions of Proteins

Protein phase separation is driven by numerous physical principles and chemical interactions; multivalent, electrostatic, cation-pi, pi-pi, and the hydrophobic effect have all been proposed to contribute to this phenomenon (Kim et al., 2016; Li et al., 2012; Lin et al., 2016; Vernon et al., 2018; Yeo et al., 2011). Phase separation occurs when a supersaturated solution of a molecule spontaneously separates into two phases, a dense phase and a dilute phase, that then stably coexist (Boeynaems et al., 2018; Li et al., 2012). It is a demixing of a homogeneous solution such that distinct regions are occupied by distinct species, as promoted by the reversible self-concentration of a protein yielding phase separations; from liquid-liquid, to liquid-gel, to liquid-solid. Phase separations by proteins are important as they are implicated in the formation of membraneless organelles for cellular organization, signaling, RNA processing, biopolymerisation and pathological aggregation (Mitrea and Kriwacki, 2016; Sfakianos et al., 2016; Su et al., 2016; Taylor et al., 2016). Many phase separating proteins have large IDRs that do not form a stable folded structure (Boeynaems et al., 2018). Interactions with solvent are typically in competition with non-covalent bonds at the surface of a well-folded protein, such as protein-protein hydrogen bonds (Efimov and Brazhnikov, 2003). However interactions common in IDRs, such as planar pi-pi interactions, are predominant in these regions, and as there is a lack of competition with water molecules for pi-pi contacts solvation of hydrogen bond donors and acceptors is promoted in these regions (Efimov and Brazhnikov, 2003; Vernon et al., 2018). High levels of solvent interaction can be detected through Hydrogen Deuterium Exchange Mass Spectrometry (HDX MS) as the availability of backbone and sidechain hydrogen atoms in IDRs promotes a rapid exchange profile (Kaltashov et al., 2013; Oganessian et al., 2018; Resetica and Wilson, 2013).

Pi-pi interactions are non-covalent bonds and involve p orbitals of atoms consisting of one electron interacting with another single occupancy p orbital, where one orbital is electron rich and the other is electron poor as caused by the atom's proximal covalent bonds (Haugh and Hirschfelder, 1955; Kertesz, 2019). These interactions are commonly associated with aromatic rings where induced quadrupolar interactions provide the non-covalent bonds formed by π (pi) orbitals; these orbitals are present in any bound sp^2 hybridized system. As the 20 amino acids all have peptide bonds, which feature sp^2 hybridized orbitals throughout the atoms of the planar amide, there is a potential for every residue in a protein to perform this interaction (McGaughey et al., 1998; Vernon et al., 2018). However, those amino acids with small side chains that leave the peptide backbone exposed; Gly, Ser, Thr and Pro, are residues in which amide pi orbitals commonly perform this interaction in proteins. As mentioned, aromatic rings stack to perform pi-pi interactions and therefore amino acids Tyr, Phe and Trp also participate using their side chains (McGaughey et al., 1998; Riley and Tran, 2017; Vernon et al., 2018). Other side chains with conjugated pi systems in their side chains include His, Gln, Asn, Glu, Asp and Arg (McGaughey et al., 1998; Vernon et al., 2018).

While any combination of pi-pi interaction is possible between residues in an IDR, it was found that the vast majority of contacts involve one of the six amino acids containing side chains with non-aromatic sp^2 hybridized atoms, or the peptide bond (Vernon et al., 2018). 36% of all observed pi-pi interactions occurring throughout solved protein structures in the PDB do not involve an aromatic residue, and that face-to-face contacts of peptide bonds occur as often as aromatic face-to-face contact. As mentioned earlier, solvation is high in IDRs and it was recognized that there is a high correlation between the number of water contacts and the probability that a residue is involved in a pi-pi contact. It was also found that planar pi-pi interactions occur

more often at positions associated with disorder, as they have an overall secondary structure with less rigidity. This is especially true of regions responsible for phase separation. Of the phase separating IDRs solved in the PDB the single property of long-range (≥ 5 residues apart or different chains) pi contact propensity is sufficient for marking them as outliers in comparison to the rest of the proteome. Through these principle properties, Vernon and others were able to make a novel software predicting IDRs in a protein based on propensity for pi-pi contacts (Vernon et al., 2018). Their work showed that most phase separating proteins can make significant non-local planar pi-interactions and based on their primary sequence analysis method, IDRs recognized with this engine have a high propensity for phase separation.

1.8 Topic of Research: Structural Study of TraG*

Determining the high-resolution structure of proteins involved in entry exclusion would be beneficial in further understanding the mechanisms of the conjugation process. However, IDRs in a protein would interfere with the generation of an ordered protein crystal, a necessity in the generation of a high-resolution diffraction pattern. The main hypothesis of the research described herein was that TraG* contained an N-terminal IDR; removal of this predicted IDR should result in increased protein stability and crystallization ability. TraG*_F, rather than intact TraG_F, was chosen for study due to the hydrophobic, membrane-bound N-terminal region of the full-length protein. Intermembrane proteins tend to aggregate at concentrations required for crystallisation, requiring specialized crystallization methods, thus providing difficulty in the purification and structural solution of proteins involved in entry exclusion such as TraG and TraS (Bayburt and Sligar, 2010). As well the large size of TraG_F would make the full-length protein challenging to overexpress in *E. coli*. Therefore TraG* was chosen to be expressed due to its reduced size and

ease of purification, as its solubility in the periplasm will permit cytoplasmic expression. Furthering the importance of this study is the encompassing of the predicted region involved in performing TraS interactions within the TraG* sequence (Audette et al., 2007). As the F pilus is the representative member of gram-negative conjugative T4SS, an understanding of the entry exclusion process of this system has the potential to develop a new paradigm for the conjugative process. Disrupting Eex to promote ceaseless conjugation thus causing lethal zygosis, or to prohibit conjugation completely, could allow for the development of a novel class of antibiotics that reduces the proliferation of antibiotic resistance.

This thesis will discuss the issues arising in TraG*_F protein crystallisation that led to the hypothesis that the N-terminal residues of TraG* are intrinsically disordered. This hypothesis was supported by novel structural prediction software based on long range planar pi-pi interactions (Vernon et al., 2018). This allowed for the design of a truncation mutant of the N-terminal 45 residues of the protein; termed TraG*_Δ, the protein begins at Thr497 of the TraG_F (Figure 1.7)(See Appendix A for Sequence Information). Constructs of C- and N- terminally hexahistidine (His₆) tagged TraG* and TraG*_Δ were cloned, expressed and purified (Table 1.1). Performing subsequent global HDX MS, circular dichroism (CD) and thermofluor assays confirmed that removal of this region enhances protein stability, and comparisons to the full-length TraG* confirm the truncated region behaves in a manner characteristic of high dynamicity. It is theorized that this region serves as a flexible linker between the membrane-bound N-terminal domain of TraG responsible for pilus assembly and the periplasmic C-terminal domain responsible for interacting with both TraN and TraS, thus providing the inherent flexibility required for mediating TraG* contacts.

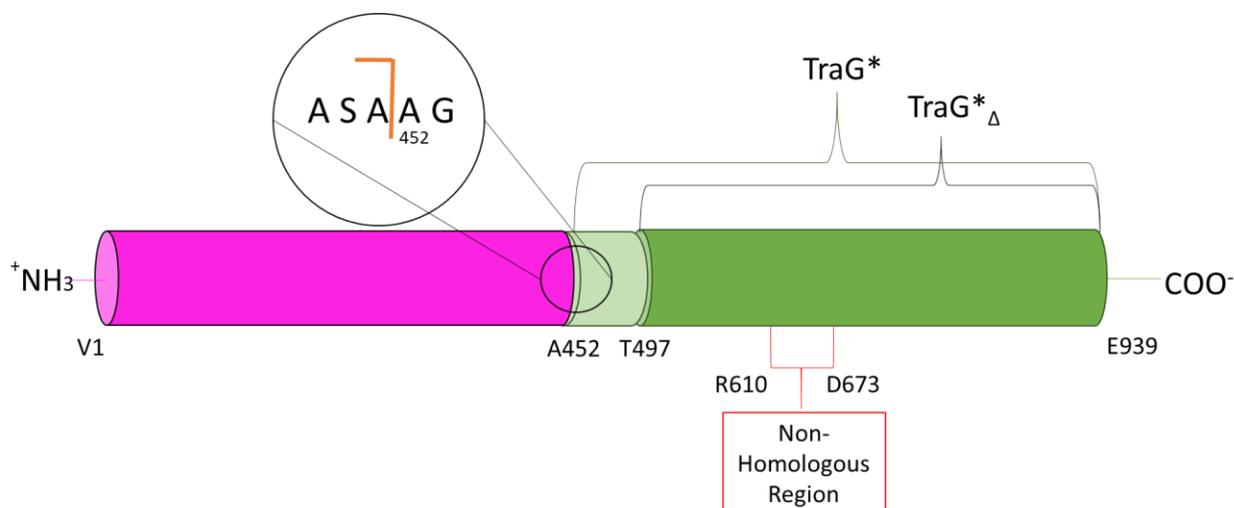


Figure 1.7: Schematic of TraG_F, TraG* and TraG*_Δ. The N-terminal domain in fuchsia from V1-A451 is inner membrane bound and responsible for pilus assembly, and the C-terminal domain in green from A452-E939 termed TraG* is responsible for Mps an Eex. The region from A452-A496 is hypothesized to be highly dynamic, and the protein from T497-E939 is termed TraG*_Δ. This construct contains the non-homologous region from R610-D673 predicted to be responsible for interacting with TraS to perform Eex.

Table 1.1: A comprehensive list of the TraG* protein variants used in this study, as well as the plasmid they were expressed from and their respective abbreviations. Full protein sequence information is listed in Appendix A.

Protein	Plasmid	Abbreviations	TraG Residues
C-terminally His ₆ -tagged TraG*	pET28a:G*His	TraG*His, G*His	A452-E939
N-terminally His ₆ -tagged TraG*	pET28a:HisG*	HisTraG*, HisG*	
C-terminally His ₆ -tagged TraG* _Δ	pET28a:G* _Δ His	TraG* _Δ His, G* _Δ His	T497-E939
N-terminally His ₆ -tagged TraG* _Δ	pET28a:HisG* _Δ	HisTraG* _Δ , HisG* _Δ	

CHAPTER 2.0 – MATERIALS AND METHODS

2.1 Reagents and Equipment

All chemicals and antibiotics used in this study were purchased from Sigma-Aldrich, VWR, Thermo-Fisher, and BioBasic unless otherwise indicated. Restriction Enzymes and cloning reagents were purchased from New England Biolabs (NEB); DNA ladders and protein markers were purchased from Thermo-Fisher. Qiagen and Anatrace crystallisation screens were used for initial vapour diffusion experiments. Plasmid vectors pET28a, pET28a:G* and pET28a:G*His as well as *E.coli* BL21 (DE3) and DH5 α were obtained from cryogenic lab stocks. Primers purchased from Integrated DNA Technologies (IDT) used in this study are listed in Table 2-1.

Polymerase Chain Reaction (PCR) was performed using a TECHNE TC3000-G thermocycler. All sonication was carried out using a Sonic Dismembrator 5000 (Fisher Scientific). Liquid chromatography was performed utilizing an ÄKTA purifier 10 from General Electric (GE) Healthcare using columns from GE and Wyatt. Agarose and Sepharose resins for protein purification were from Fisher Scientific. Concentration and dialysis of proteins was performed on 30,000 Dalton (Da) Molecular weight cut-off (MWCO) centrifugal concentrator (Millipore, Corning). DNA quantification was estimated using a NanoDrop 2000 spectrophotometer (Thermo Scientific). Protein quantification assays were conducted using a Synergy H4 Hybrid microplate reader from BioTek. Electrospray ionization mass spectrometry was performed on a Waters Synapt G1 operating in positive ion mode. Circular Dichroism experiments were conducted on a J-815 CD Spectrometer from Jasco. Dynamic Scanning Fluorimetry (ThermoFluor) experiments were performed on a RotorGene Q qPCR from Qiagen and SYPRO orange was used to detect protein unfolding (Life, ThermoFisher). Observations of crystallisation experiments were carried out using a Nikon SMZ1500-Fibre Lite MI 150 stereoscopic microscope.

Table 2.1: Primers used in the creation of plasmid constructs pET28a:HisG*, pET28a:HisG* Δ and pET28a:G* Δ His. The underlined portion of the primer sequences show their respective restriction enzyme cut sites.

Construct	Restriction Enzyme	T _m (°C)	Primer Sequence (5'–3')
pET28a:HisG*_Fwd	BamHI	55	TTAGGATCCGCAGGCAGTGTGTTG
pET28a:HisG*_Rev	HindIII	56	CGCAAGCTTTTATTCTTTATGCTGGTAACTCTTTGC
pET28a:HisG* Δ _Rev			
pET28a:HisG* Δ _Fwd	BamHI	62	CGAGGATCCACACAGACCCGTGACGGTAATATG
pET28a:G* Δ His_Fwd	NcoI	61	CGACCATGGACACAGACCCGTGACGGTAATATG
pET28a:G* Δ His_Rev	HindIII	56	CGACAAGCTTTTCTTTATGCTGGTAACTCTTTGC

2.2 Analysis of TraG_F Using Software for the Detection of Disorder

The primary sequence of TraG_F was input into the analysis servers PrDOS, DisEMBL 1.5, SCRATCH, MFDp2, IUPred2A, Phyre2, I-TASSER, and the IDR Phase Separator Predictor (PSP) (Cheng et al., 2005; Dosztányi et al., 2009; Linding et al., 2003; Magnan and Baldi, 2014; Mészáros et al., 2018; Mizianty et al., 2013; Roy et al., 2010; Vernon et al., 2018). Decisions in designing primers for cloning of a TraG* mutant were based on the results of these analyses.

2.3 Cloning of TraG* Constructs

2.3.1 Plasmid Extraction of pET28a

Overnight cultures of DH5 α *E. coli* with plasmids pET28a and pET28a:G*His were prepared in 10mL Luria Bertani broth (LB) with 50 μ g/mL Kanamycin (Kan) at 37°C with shaking at 200 rotations per minute (rpm). Plasmid extraction was performed using the GeneJET plasmid miniprep kit from Thermo-Fisher following the manufacturer's protocol. The extracted plasmids

were quantified using a NanoDrop 2000 Spectrometer from Thermo Scientific and qualified using a 1.2% agarose gel.

2.3.2 PCR Amplification of traG and traG* Δ*

Primers were designed based on guidelines from IDT and analyzed using the IDT OligoAnalyzer[®]. Melting temperatures were maintained within 5°C between forward and reverse primers, GC content was between 35-65%, and the free energy to generate self-dimers, hairpins or heterodimers was kept above -9.0kcal/mol. Q5 DNA polymerase was used to amplify the genes from 200ng of pET28a:G* plasmid. The PCR protocol was performed as follows: initial denaturation at 98°C for 30s, then 30 cycles of denaturation, annealing and extension at 98°C for 10s, 55°C for 20s, and 72°C for 28s respectively. Then a final extension was performed at 72°C for 10 minutes. PCR products were purified using the GeneJET PCR Purification kit from Thermo-Fisher following the manufacturer's protocol. Amplicons were then quantified by NanoDrop 2000 and qualified using a 1.2% agarose gel.

2.3.3 RE Digestion, Ligation and Transformation of traG Mutants into pET28a*

The pET28a vector and purified amplicons were double digested using the appropriate restriction enzymes (RE) in 1X NEB CutSmart Buffer. 1µg of DNA was digested with 20,000 units of each RE at 37°C for 30 minutes. After this incubation, 1 unit of Calf Intestinal Alkaline Phosphatase (CIP) was added to the double digested (DD) pET28a samples to de-phosphorylate the 5' and 3' ends of the linearized plasmid to prevent self-ligation. These samples were then incubated for an additional 30 minutes at 37°C. DD plasmids and amplicons were then isolated

through separation on a 1.2% agarose gel by performing electrophoresis at 100V for 40 min, followed by gel extraction using the GeneJET Gel Extraction kit from Thermo-Fisher using the manufacturer's protocol. The DNA samples were then quantified by NanoDrop 2000.

NEBs ligation calculator was then used to design ligation experiments in 1:5 and 1:7 proportions of plasmid to amplicon, and a plasmid self-ligation control. Ligation was performed overnight at 4°C using T4 DNA ligase in proportions described by the manufacturer's protocol. Heat inactivation was performed at 65°C for 10 minutes, and after cooling at 4°C 5µL of the ligation reaction was added to 90µL of DH5α competent cells (protocol for the generation of competent cells found in Appendix B). The cells were incubated at 4°C for 20 min, heat-shock transformed at 42°C for 90 seconds, cooled at 4°C for 5 minutes and then incubated at 37°C for 1 hr following the addition of 900µL of LB to equilibrate the transformed cells. The cells were spun down at 5000 x g for 10 minutes at room temperature, then resuspended with 100µL of LB. 30µL of each cell resuspension was spread plated onto a LB Kan⁺ (50µg/mL) plate, as well as control plates LB Ampicillin⁺ (100µg/mL) and LB with no antibiotic to check for cell viability. Plates were kept at 37°C for 16-24 hrs.

Colonies grown from LB Kan⁺ plates were aseptically picked and used to inoculate 10mL of LB Kan⁺ which was then incubated at 37°C overnight. Plasmid extraction was performed as described in Section 2.2.1 and restriction enzyme digestion was performed as described above to check for insertion of the proper gene through performing agarose gel electrophoresis. The plasmids with the correct insert size were sent to BioBasic (Markham, ON) for dideoxy sequencing of the multiple cloning site using the T7 Promoter and Terminator universal sequencing primers. Plasmids determined to have the correct *traG** genes inserted in the proper frame were used to transform BL21 (DE3) competent cells as in the above heat shock protocol.

2.4 Expression and Purification of TraG* Mutants

2.4.1 Optimization of TraG* Expression

Initial expression trials of untagged TraG* and TraG*His were performed previously at 37°C (Ergodan, 2016). A purification after 16°C overnight incubation was performed on TraG*His and results were compared to purifications performed at 37°C for 4 hours. The optimized protocol follows: a 10mL feeder culture of BL21 *E. coli* with the desired TraG* plasmid construct was grown in LB with 50µg/mL Kan overnight at 37°C. This feeder culture was used to inoculate 1L of LB with 50µg/mL Kan and 1mM Glucose. After the 1L culture had reached the mid-logarithmic growth phase ($OD_{600nm} = 0.5 - 0.8$), ethanol was added to a final concentration of 1% to induce a stress response and promote chaperone protein assembly (Chhetri et al., 2015). Expression of the TraG* variant was then induced through addition of IPTG to a final concentration of 1mM. An incubation was performed for 16 hours at 18°C with shaking at 200rpm, and cells were then pelleted at 5,000 x g for 20 min at 4°C. HisTraG*, TraG* Δ His and HisTraG* Δ expressions were performed in the same way.

2.4.2 Purification Schemes of TraG* Variants

Cell pellets from 500mL LB were resuspended with IMAC Loading Buffer (20mM Tris pH 7.5, 200mM NaCl, 10% glycerol). Protease inhibitors phenylmethane sulfonyl fluoride (PMSF), Benzamidine HCl and Aminocaproic acid were added to final concentrations of 1.25mM, 5mM and 5mM respectively. Lysis was performed through sonication at 25% amplitude with a 20s on, 40s off cycle for 10 min; the lysate was then separated from cell debris through centrifugation at 25,000 x g for 30 min at 4°C.

Ni²⁺ IMAC was performed through a XK26 Ni²⁺-NTA agarose column using an Äkta Purifier 10S FPLC (GE Healthcare). The lysate was pumped into the column at 1mL/min and the beads were washed with 100 mL of IMAC Loading Buffer with 30mM Imidazole (Imid), and then 75 mL of IMAC Loading Buffer with 50 mM Imid. 3mL fractions were taken as the protein was eluted with 75mL IMAC Loading Buffer with 200mM Imid for TraG*His and 300mM Imid for HisTraG*_Δ and TraG*_ΔHis. A final elution was performed with 50mL of IMAC with 500mM or 1M Imid.

2.4.3 Visualization of Proteins using SDS PAGE

Samples of interest were aliquoted and appropriately diluted to a 15μL volume, then mixed with 5 μL of 4 X SDS buffer (160mM Tris pH 6.8, 4.0% SDS, 20% glycerol, 0.0012% Bromophenol blue, 20% β-mercaptoethanol). After boiling at 100 °C for 5 min, 10 μL of each sample was loaded onto a 12.5% SDS PAGE gel and electrophoresis was performed. The gel was stained with Coomassie solution (50% methanol, 10% glacial acetic acid, 0.1% Coomassie Brilliant Blue R250) to visualize protein bands.

*2.4.4 Desalting and Quantification of TraG**

Fractions from the affinity purification with high quantities of HisTraG* relative to impurities were pooled and dialyzed using a 30kDa MWCO centrifugal concentrator column (Corning, Millipore) at 5,000 x g at 4°C into 10mM HEPES pH 7.0, 10% glycerol. The protein was concentrated to 5mL and then added to a 26 10 HiPrep desalting column (GE) using the Akta FPLC with 10mM HEPES pH 7.0, 10% glycerol.

The properly exchanged TraG* sample was quantified by Bicinchoninic Acid Assay (BCA) using the protocol described on the Pierce™ BCA Protein Assay Kit (Thermo Scientific). 12.5 µL volumes of various dilutions of the protein sample were mixed with 237.5 µL of the working reagent and loaded on a 96 well plate (Sarstedt). After incubation at 37°C for 30 minutes the OD_{562 nm} of each sample was determined using the Synergy H4 Hybrid plate reader. After protein concentration was determined the HisTraG* sample was loaded onto a 30kDa concentrator column and centrifuged to a final volume determined to bring the protein to a concentration of 30 mg/mL. This concentration would then be confirmed by comparing samples with predicted masses of TraG* to samples of BSA with known protein concentration using SDS PAGE (Section 2.4.3). The protein was aliquoted, flash frozen with liquid N₂, and stored at -80°C.

2.5 Screening and Optimization of Crystallisation conditions of TraG*

2.5.1 Screening of Crystallisation conditions

TraG* protein samples were loaded in a variety of concentrations as 0.75 µL amounts to the sitting drop wells of a 96 well plate from Axygen. 90 µL of a screening buffer was added to the reservoir well, and 0.75 µL of that reservoir solution was mixed with the protein solution drop in the corresponding sitting well. This was performed using the buffers from the MCSG -1, -2, -3, and -4 kits from Anatrace, as well as PUREPEG and TOP 96 from Anatrace and the JCSG+ screen from Qiagen. These plates were sealed to allow vapour diffusion to occur and were kept in a place that experiences little vibration at both 4°C and 19°C. Results were checked periodically under a stage microscope with a birefringent lens.

Conditions that were determined to produce TraG* crystals were optimized through altering pH of the buffer, salt concentration, precipitant content, protein content, temperature and drop sizes in 24 well hanging drop plates. Typically, the plate was set up with 1 mL of mother liquor in the reservoir, with 1 μ L of this solution placed on a plastic coverslip. Then 1 μ L of the protein solution was added to the reservoir solution droplet and the coverslip was placed onto the corresponding reservoir such that the drop was hanging above and diffused with the reservoir due to effective sealing of the coverslip using high vacuum grease.

Diffraction screening of crystals was performed on a Rigaku MicroMax-007 HF rotating anode X-ray generator with a Saturn944+ charge coupled device detector. Potential TraG* crystals were looped with appropriately sized microloops (Mitegen), added to a cryogenic solution which consisted of conditions mimicking that of the reservoir but with a discrete addition of either glycerol, ethylene glycol, or an increased amount of the same precipitant used in the reservoir. The crystal was flash frozen in liquid N₂ and mounted on the system where diffraction screening was performed for 10 seconds of X-ray exposure with 4 images collected at 60.0mm crystal-to-detector distance with angles from 0°-2° and 70°-72°.

2.5.2 Thermofluor Assay

To quantify the beneficial properties of buffers and salts that seemed to aid the stability of TraG* mutants in vapour diffusion experiments, thermofluor experiments were performed. Thermofluor, or differential scanning fluorimetry, involves the use of a fluorescent hydrophobic probe to determine the optimal solvent-protein interactions using the melting temperature (T_m) of the protein as an indicator of stability (Ericsson et al., 2006; Lavinder et al., 2009; Phillips and de

la Peña, 2011; Seabrook and Newman, 2013). A protein in a specific buffer and salt condition is heated from 25°C to 99°C in a gradient fashion while fluorescence absorption is measured. As the temperature is increased the protein begins to unfold and hydrophobic sidechains of its residues allow for the hydrophobic probe to bind, causing an increase in the fluorescent signal. Fluorescence then decreases again as proteins begin to denature and aggregate. The T_m as dictated by the point at which the fluorescent signal is increasing at its highest rate, is indicative of a set of buffer conditions that aid in maintaining protein homogeneity, solubility and folding stability; the parameters most suitable for crystallisation of a protein would be those which provide the highest melting temperature with a well-defined melting curve shape.

SYPRO Orange dye was the probe used for thermofluor experiments in this thesis, and it was prepared through serial dilution from a 5000X commercial stock (Life, ThermoFisher) to a final concentration of 10X in the buffers of interest, with 1mg/mL of TraG* (purified with 0% glycerol) in a total of 50µL. Fluorescence output was set to 470nm and detected at 610nm, with a gain of 5, and the samples were heated from 25°C to 99°C at a ramp rate of 1°C/min. The scheme used was similar to that shown in (Seabrook and Newman, 2013); a total of 14 buffers were tested at 50mM concentration with either 50mM NaCl or 200mM NaCl, and each condition was performed in triplicate, with a lysozyme positive control in 1xPBS, dye-only negative control and a protein-only negative control in every experiment. Further optimization was performed with the buffer conditions that promoted the highest melting temperature, wherein 8 common salts used in crystallography replaced NaCl.

2.6 Analysis of TraG* Structural Properties

2.6.1 Circular Dichroism

Circular dichroism (CD) is a widely used and well-developed technique for the determination of protein secondary structure (Micsonai et al., 2015; Provencher and Glöckner, 1981). The technique relies upon the differential absorption of left- and right-hand circular polarized light. In this case, the electric field of a photon has a circularly rotational direction relative to the propagation direction of the photon, where the photon vector remains constant in magnitude (Greenfield, 2007; Kelly et al., 2005). When these photons pass through an asymmetric (chiral) molecule, the speed, absorbance and the wavelengths differ depending on the direction of their polarization. As the sum of the vectors for the right- and left- handed polarized light forms an ellipse, the change in ellipticity (measured as $\Delta\epsilon$) of a substance as a function of the wavelength of incident light is reported in performing CD experiments, even though the change in absorbance of the differentially polarized light is what is measured in CD spectrometer.

The $\Delta\epsilon$ of TraG* mutants were measured by CD spectroscopy at 2 μ M over wavelengths 190-260nm. To normalize the CD spectra of the proteins the spectrum of the solvent, 10mM HEPES pH 7.0, 10% glycerol, was subtracted from the experimental spectra. All spectra were obtained through a continuous scan performed at 100 μ m/min with $\Delta\epsilon$ measurements every 0.1nm, with an accumulation factor of 8. Urea denaturation experiments were performed to provide an indication of protein stability when comparing TraG* with and without the putative IDR, as denaturing studies are common for determining the relative stability of protein mutants (Glover et al., 2016; Griko et al., 2001; Matsuo et al., 2007). TraG* variants were added to solutions with different urea concentrations in 10mM HEPES pH 7.0, 10% glycerol such that the final concentration of protein was 2 μ M. Samples were incubated at 25°C for 1 hour and measured on

the CD spectrometer. Initially urea concentrations of 0.5, 1, 2, 3, 4, 5, and 6M to determine the approximate range for protein denaturation, which was narrowed to 3–4M. Then concentrations of 3.2, 3.4, 3.6 and 3.8 M were tested. Spectra measurements were performed with the same scanning protocol as stated above, and all results were normalized to their respective solvent conditions. Deconvolution of all CD data was performed using BeStSel, an algorithm for protein fold recognition and secondary structural determination using input CD spectra, unique in its ability to distinguish parallel from antiparallel β -sheets (Micsonai et al., 2015).

2.6.2 Global HDX MS

A vital technique for the determination of protein structure and dynamics is HDX MS. $^2\text{H}_2\text{O}$ (D_2O) is used to provide an indication for the solvent-exposure of residues in a protein chain (Oganesyan et al., 2018). The exchange of the amide hydrogen is highly dependent on the higher order structure of the protein, therefore subtle perturbations in solvation affect the exchange rate (Balasubramaniam and Komives, 2013; Beveridge et al., 2013; Oganesyan et al., 2018). As well, at neutral pH the peptide backbone protons are exchanged at a rate amenable to measurement by MS. As a change in the mass-to-charge ratio (m/z) of the molecule over time can be effectively measured, the overall dynamics of a protein can be determined based on the rate of deuteration. In HDX MS techniques, an acid is used to quench the deuterium exchange reaction, followed by enzyme digestion to fragment the macromolecule into peptides allowing for a residue-by-residue analysis of solvent exposure, thus providing insight on protein structure (Oganesyan et al., 2018; Wang et al., 2002; Zhang and Smith, 1993). In global HDX MS the protein is left intact as a way to monitor the overall changes in protein structure and stability; this technique shows the sum of exchange kinetics for the individual backbone amides over the entire protein (Houde et al., 2009;

Oganesyan et al., 2018). HDX MS requires proteins to enter the spectrometer in gas phase as an ion; electrospray ionization (ESI) is a soft ionization method that allows for evaporation of the solvent such that molecules in solution are sent into the spectrometer as intact ions in individual nanodroplets (Katta and Chait, 1991; Oganesyan et al., 2018).

In the global HDX scheme performed in this thesis, D₂O mixing was performed in the ESI capillary; differential deuteration times were obtained by pulling the protein-dispensing capillary a measurable distance through an outer capillary into which D₂O is pumped, providing discrete deuteration time-points. Protein samples were prepared for MS analysis through a Zeba spin desalting column (Thermo Fisher) following the manufacturer's protocol to ensure removal of small molecule contaminants and to buffer exchange the protein into 50mM NH₄CH₃COO pH 6.8, MS grade. Native MS analysis was performed on TraG* and TraG*_Δ with 5μM protein flowing at a rate of 8μL/min. A series of deuteration time points were collected for the native folds of TraG* and TraG*_Δ to confirm the presence of a putative N-terminal IDR. 0, 1, 2, 5, 10, 20, 50 and 100mm time points were collected with 30μM protein flowing at a rate of 4μL/min, and D₂O at 16 μL/min. Fully deuterated spectra were obtained through a 1:5 mixing of the proteins with D₂O to a final protein concentration of 5μM, incubated at 25°C for 30 minutes and then electrosprayed at 8μL/min. In these experiments capillary voltage was set to 3.5 kV, the sampling cone was 150.0 and the extraction cone was 7.0, with source temperature 120°C, desolvation 250°C, cone gas flow 30.0L/hr, nanoflow gas pressure at 3 bar, desolvation gas flow at 450L/hr, the trap and transfer collision energy was 15.0 and 10.0 respectively, the trap gas flow was 3.0 mL/min and the source gas flow was 11.0mL/min. Backing pressure was raised to 3.9 millibar (mBar) to accommodate ESI MS of a protein larger than 50kDa (Wilm, 2011).

CHAPTER 3.0 – RESULTS AND DISCUSSION

3.1 Software Mediated TraG Structural Predictions

The full length TraG sequence was entered into the protein homology/analogy recognition engine (Phyre2), resulting in sequences with poor homology to TraG (Kelley et al., 2015). As with most structural predictors, Phyre2 uses sequence homology to compare an entered primary sequence to those of proteins with solved structures in the protein databank (PDB). A subunit of the vacuolar proton ATPase from *Thermus thermophilus* solved by cryo-electron microscopy displayed the highest template coverage at 27%, with a confidence value of 91% and identity of 7% (Figure 3.1A) (Schep et al., 2016). The modeled residues corresponded to residues S657-K912 of TraG. This protein was only recognized when the full length TraG sequence was used for the query, when TraG* is entered the highest template covered was 3% corresponding to residues S751-V766 of the protein with a confidence of 7.3% and identity of 25%. As well, the secondary structure analysis of TraG* by Phyre2 revealed 56% of the protein sequence is predicted to be α -helical, 5% showed a potential for forming β -strands (with low confidence), but overall 56% is unstructured (Figure 3.1B). Failure to detect matches for significant regions of a protein sequence entered into structural predictors is typically caused by a lack of a sufficient number and diversity of homologous sequences to the query, indicating the presence of a novel fold not present in the current structural database (Kelley et al., 2015; Zhang, 2009). As mentioned in section 1.6.2., there are few homologous proteins to TraG, the regions that are predicted to be disordered may be unrecognized as they are part of a novel fold.

3.1.1 PSP Software Analysis of TraG_F Reveals a Putative IDR

Sequence analyses of full length TraG_F using a variety of popular structural prediction software provided variable results in their structural predictions of TraG*. Many of the algorithms described in Section 2.2 provided predictions similar to those seen in Figure 3.2A from the software IUPred2A and ANCHOR 2A. IUPred2A bases its structural disorder predictions on theoretical redox state and potential for binding to global protein partners, while ANCHOR 2A uses homology to known protein structures to recognize disordered regions with the potential for becoming ordered upon binding (Dosztányi et al., 2009; Mészáros et al., 2018). The two programs similarly recognized ordered secondary structures in the membrane-bound residues M1-A440, however predictions identifying disordered regions of TraG* showed contradiction. In many instances, the primary sequences that result in peak values which cross the predictive threshold for IDRs in one software show a trough in the other software, and vice versa. As mentioned in Section 3.1 failure of multiple predictive software highly dependent on analogy to achieve consensus regarding the secondary structure of a region is likely indicative of a lack of structural data for homologous proteins (Kelley et al., 2015; Zhang, 2009). As TraG*_F is a synthetic construct that is not normally expressed without the N-terminal domain, the successful expression and purification of TraG*_{His} and untagged TraG* in previous studies provides evidence of a stable fold inherent in the protein (Ergodan, 2016). A synthetic protein with many closely interspersed highly disordered regions would likely aggregate, as most native IDPs evolved to prevent self-aggregation and perform their *in vivo* function (Uversky et al., 2008).

The novel phase separation predictor (PSP) software for predicting phase separating IDRs with long range pi-pi contacts proved more insightful in determining which regions of TraG are putative IDRs and can be targeted for mutation or deletion (Vernon et al., 2018). The region from

residues 447-498 was predicted to cause phase separation in the protein, with 27 of these residues surpassing the score threshold of 4.0 thus providing TraG with an overall score of 4.06. When the sequence for TraG* is entered into the PSP (Figure 3.2C), the resultant P score is 2.53 as the influence of the N-terminal region for increasing phase separation propensity is lowered as the algorithm predicts an expected level of disorder in terminal residues. However, the 48 N-terminal residues approach the threshold score for the PSP to predict the described region as phase separating, thus the N-terminus may provide a degree of disorder sufficient to prevent good crystallization. When the sequence of TraG* Δ is input in the PSP (Figure 3.2C) the P score is lowered to 1.37, a substantial reduction for the deletion of 45 amino acids. This provided ample evidence to proceed with the cloning of TraG* Δ mutants.

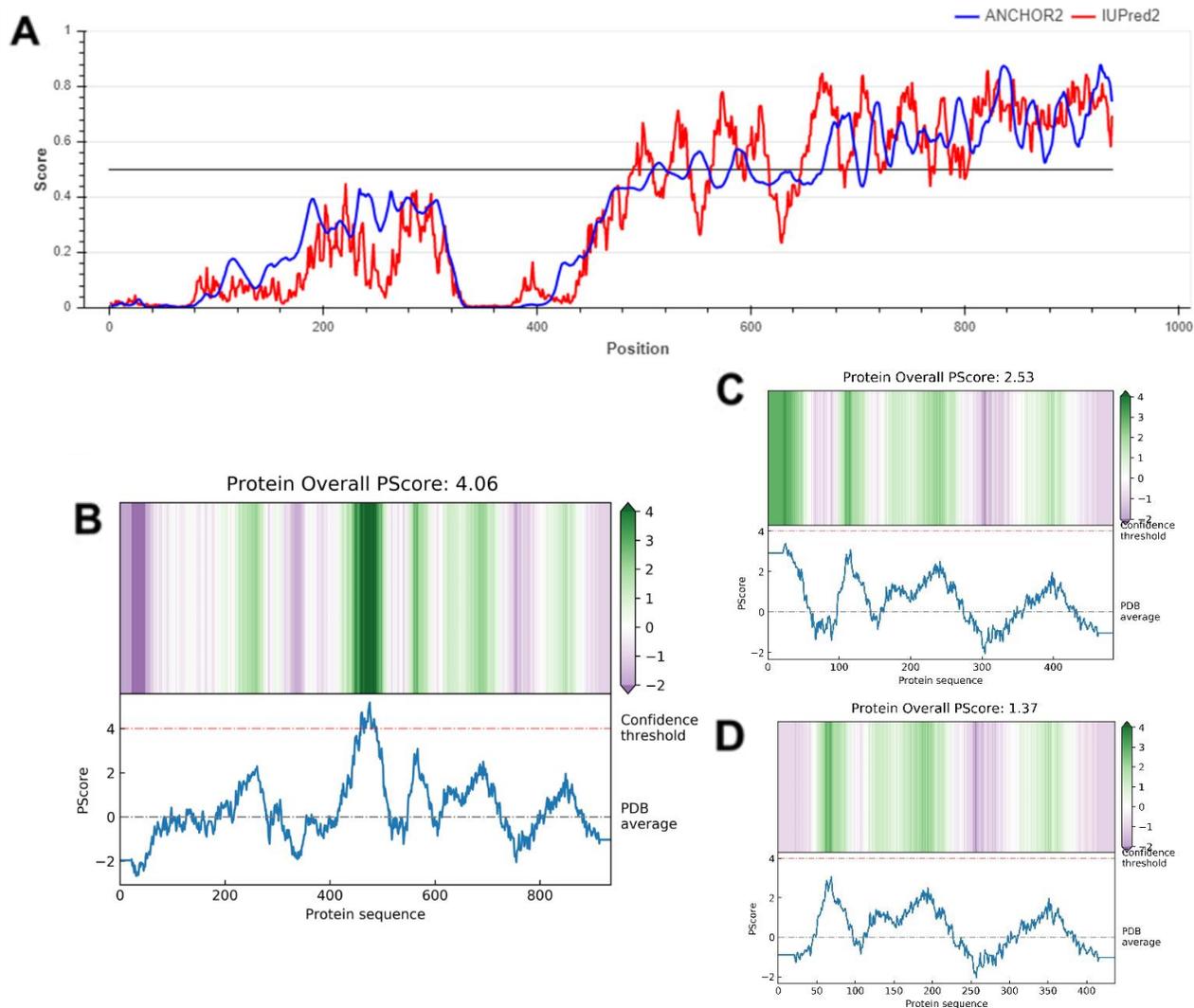


Figure 3.2: Software mediated identification of IDRs in TraG from the F plasmid. **A)** IUPred2A and ANCHOR 2A algorithms were used in tandem to analyze the TraG primary sequence and produce the graph, shown as red and blue lines respectively (Dosztányi et al., 2009; Mészáros et al., 2018). **B)** the resultant 2D heat map and graph produced by entering the TraG sequence into the PSP (Vernon et al., 2018). The only region predicted to be an IDR based on phase separation is therefore the putative linker from residues P447-Q498, as it is the only region that surpasses the designated confidence threshold. TraG* shows a lower predicted capacity for phase separation as seen in **C)** due to the algorithm’s treatment of N-terminal residues. As a large region with high predicted disorder in its N-terminus results in difficulty in crystallisation, the truncation mutant TraG* $_{\Delta}$ was conceived. **D)** the PSP result for the TraG* $_{\Delta}$ sequence, predicting low overall phase separating propensity.

3.2 Successful Cloning of TraG* Constructs

DNA extraction of the necessary starting products for cloning was successful, with pET28a empty vector seen as pure and at the expected size of 5369bp (Figure 3.3A). pET28aG* was observably larger than the expected mass of 6656 bp due to differences in DNA topology, where the lowest MW band is supercoiled DNA, and increasing weights are further relaxed conformations.

The amplicons replicated from pET28aG* for the cloning of plasmids encoding for N-terminally His₆-tagged TraG* (HisTraG*), N-terminally His₆-tagged TraG*_Δ (HisTraG*_Δ) and C-terminally His₆-tagged TraG*_Δ (TraG*_ΔHis) are seen in Figure 3.3B. The primers described in Table 2.1 were successful in the amplification of the desired gene products and provided the expected sizes of 1464 bp, 1329bp and 1326bp respectively. After purification, restriction enzyme digestion, ligation and transformation steps described in Section 2.3.3, successful colony growth was achieved in all 1:5 and 1:7 ligation experiments; the DH5α colonies from the transformed 1:7 ligation product were used for further experimentation due to their optimal colony number and size. Dideoxy sequencing (BioBasic) after purification of the plasmids from O/N cultures showed proper insertion of the genes as well as the correct genetic code. Heat shock transformation into BL21 competent cells was performed successfully; the plasmid constructs mini-prepped from O/N BL21 cultures and restriction enzyme double digests of these constructs are shown in Figure 3.3C and displays the correct size of every plasmid and gene insertion. Table 1.1 lists the plasmid constructs, their abbreviations, and the abbreviation of the proteins they express.

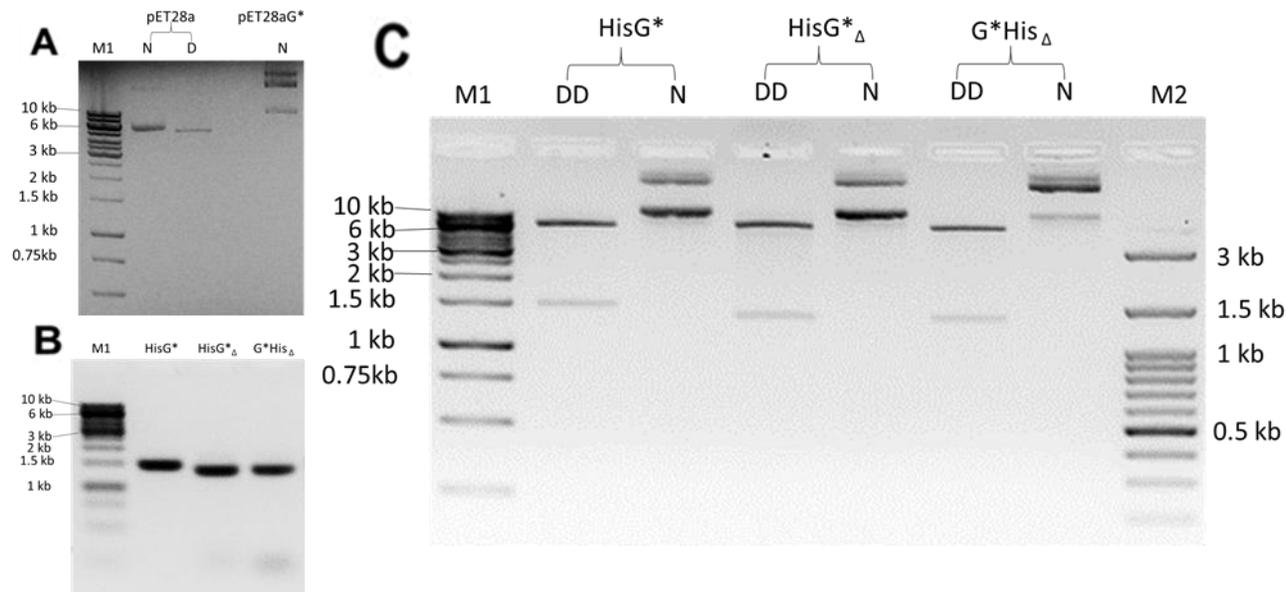


Figure 3.3: **A)** Plasmid constructs used for the cloning of *traG** displayed on a 1.2% agarose gel after electrophoresis. pET28a was the target vector for insertion of *traG** as amplified from pET28aG*, which has full length *traG** inserted in its MCS. **B)** The amplicons coding for HisG*, HisG* Δ , and G*His Δ , produced through PCR of pET28aG*. **C)** Successfully cloned plasmid constructs mini-prepped from BL21. N: undigested plasmid, D: single digest by restriction enzyme BamHI, DD: restriction enzyme double digest (BamHI and HindIII for N-terminally tagged products, NcoI and HindIII for C-terminally tagged products), M1: GeneRuler 1kb DNA Ladder (#R0491 ThermoFisher), M2: 100bp DNA Ladder H3 RTU (#DM003-R500 GeneDireX).

3.3 Expression and Purification of TraG* Constructs

The optimized 18°C O/N TraG*His expression scheme was seen to induce transcription and translation of TraG*His to a higher extent compared to expression at 37°C for 4 hours (results not shown). Chromatograms and SDS PAGE images of optimized purification schemes for the TraG* constructs are shown in Figures 3.3 & 3.4. TraG*His was seen to be isolated with slightly lower overall purity and yield in comparison to HisTraG*. A gradient purification was performed for HisTraG* as the purity and yields were ameliorated in this scheme compared to step elution schemes.

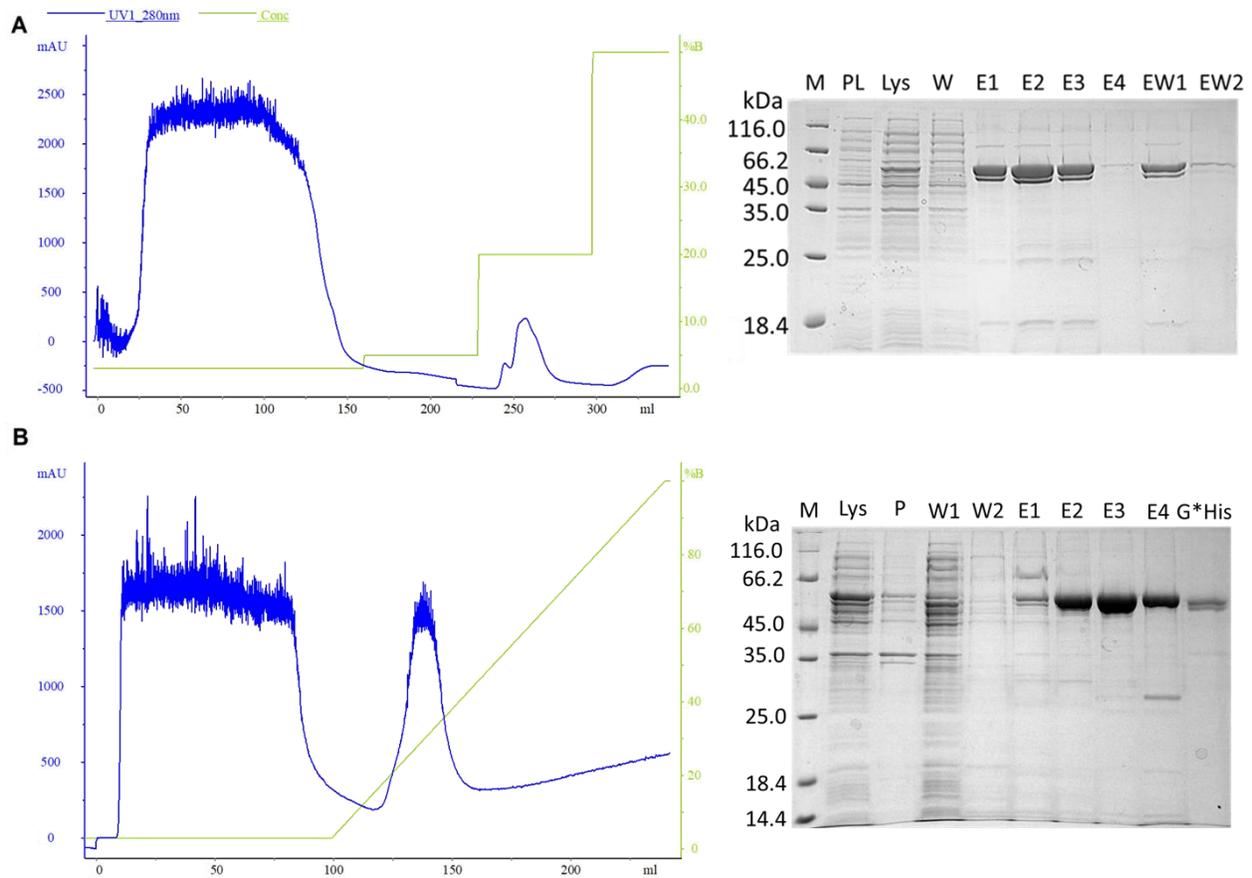


Figure 3.4: Optimized Ni^{2+} IMAC purification schemes of TraG* variants presented on a chromatogram and a 12.5% SDS PAGE gel, purified with a XK26 Ni^{2+} -NTA agarose GE column using an Äkta Purifier 10S FPLC. The purification schemes were performed using 20mM Tris pH 7.5, 200mM NaCl, 10% glycerol as the sample running buffer and eluted with 20mM Tris pH 7.5, 200mM NaCl, 10% glycerol, 1M Imid. **A)** Purification scheme of TraG*His, where PL: lysate from culture prior to induction, Lys: lysate after O/N 18°C expression, W: from 0-150mL, E1: peak at 245mL, E2: peak at 260mL, E3: end of the peak at 275mL, E4: 500mM Imid elution at 330mL, EW1: pooled 240-275mL, EW2: pooled 500mM Imid from 320-350mL. **B)** Gradient purification scheme for HisTraG*, where labelling is the same as above except P: cell debris pellet, W2: 50-100mL, E1: 120mL (127mM Imid), E2: 135mL (256mM Imid), E3: edge of peak at 145mL (323mM Imid) E4: 155mL (436mM Imid), G*His: 1 μ g of purified TraG*His. In all figures shown M: unstained protein marker (ThermoFisher #26610).

HisTraG* and TraG*His both show two bands copurifying at the approximate size of the proteins of interest (Figure 3.4) (see Appendix A). The lower MW abundant protein in TraG*His seems to be of similar quantity to that of the higher MW species. In the HisTraG* purification the relative amount of the higher MW protein is abundant in comparison to the lower MW band. As well, an impurity of the same size appears to be highly present in the wash fractions for HisTraG*. It was predicted that the higher MW band is the full-length protein and the lower MW band is a common degradation product, possibly from the removal of the putative IDR in the N-terminus. This conclusion is supported by the increased purity of the N-terminally His-tagged protein, as N-terminal cleavage products would not have the opportunity to bind the Ni-IMAC column unlike the C-terminally His-tagged variant which would allow for the purification of products with N-terminal degradation.

TraG* Δ His was not seen to be expressed, while HisTraG* Δ purified optimally in comparison to the other constructs (Figure 3.5A and B). Furthering the above argument is the evident purity of the HisTraG* Δ ; no degradation products are seen. The lack of expression in TraG* Δ His might be due to the requirement for some residues to serve as a N-terminus; without them the region may be excessively solvent exposed such that the formation of its proper secondary structure is prevented, causing instability and degradation during expression. The comparison between 1 μ g of HisTraG* Δ , 0.5 μ g of TraG*His and 1 μ g of BSA show accuracy in BCA quantitation (Figure 3.5B).

The purification schemes were proven to be optimal for all proteins except TraG* Δ His through observation of the lanes displaying lysate, washes and elution fractions (Figures 3.4 and 3.5). In lysate samples an abundance of the expected 54.2 kDa protein in comparison to the protein contaminants is seen. As well the washes show low levels of the desired proteins with high

amounts of impurities, such that the chromatograms show a nearly baselined UV_{280nm} before the peak shown to be abundant TraG* is eluted. In many cases, a UV_{280nm} peak is seen in the chromatograms at the 1M Imid elution despite no protein being eluted. This is likely absorbance caused by imidazole, which can absorb light at 280nm due to the conjugated pi system of the molecule.

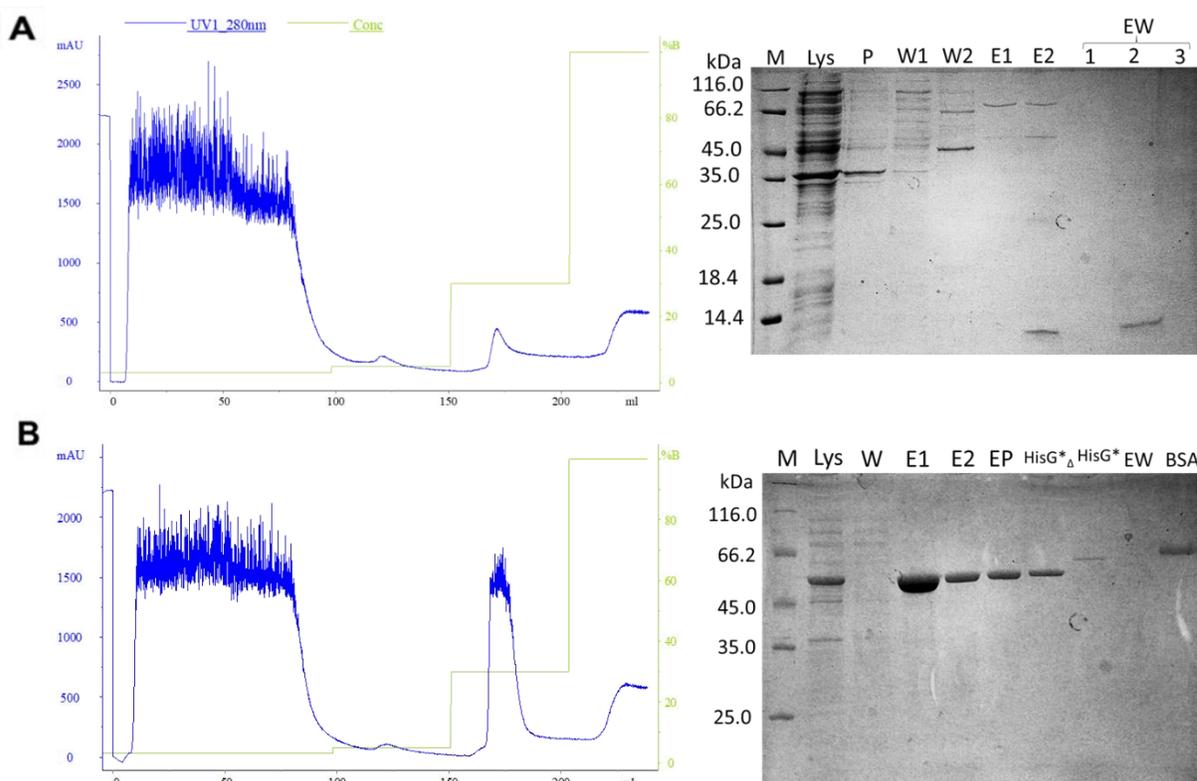


Figure 3.5: Optimized Ni²⁺ IMAC purification schemes of TraG*_Δ variants presented on a chromatogram and a 12.5% SDS PAGE gel, purified with a XK26 Ni²⁺-NTA agarose GE column using an Äkta Purifier 10S FPLC. The purification schemes were performed using 20mM Tris pH 7.5, 200mM NaCl, 10% glycerol as the sample running buffer and eluted with 20mM Tris pH 7.5, 200mM NaCl, 10% glycerol, 1M Imid. **A)** Purification scheme of TraG*_ΔHis where Lys: lysate after O/N 18°C expression, P: cell debris pellet, W1: 0-25mL, W2: 100-125mL, E1: fraction of peak at 173mL, E2: fraction of peak at 177mL, EW1:180mL, EW2: 190mL, EW3: peak at 230mL **B)** holds a similar labeling scheme for the purification of HisTraG*_Δ, however E1: 165mL, E2: 175mL, EP: pooled 160-180mL, HisG*_Δ: 1μg of HisTraG*_Δ, HisG*: 0.5μg of HisTraG*, EW: 220mL, BSA: 1μg of BSA. In all figures shown M: unstained protein marker (ThermoFisher #26610).

3.4 Thermofluor Analyses of TraG*His and HisTraG* Δ Reveal Differences in Protein Stability

The resultant melting curves from performing thermofluor analyses of TraG*His in various buffer conditions exhibited notable trends. On average, conditions with 200mM NaCl provided a lower observed T_m than when buffers had 50mM NaCl (Figure 3.6A). As well variability was high for conditions below pH 6.0 as seen by the large error bars representing standard deviation between triplicate data points. The buffers that provided TraG*His optimal T_m values were 50mM Citrate pH 6.0 and 50mM NaCH₃COO pH 5.2 at approximately 54°C (Figure 3.6A & B). These values are proximal to the protein's predicted pI of 5.61 (Appendix A). The melting curve provided when TraG*His was in 50mM citrate pH 6.0 was markedly improved in terms of the optimal shape in comparison to that of 50mM NaCH₃COO pH 5.2 (Figure 3.6B). Some conditions with pHs higher than pH 6.0 had a poor or indistinguishable increase in fluorescence throughout the melting curve, such as the data seen for 50mM MES pH 6.0 (Figure 3.6A & B). This indicates TraG*His was unstable in this buffer condition and was partially or completely misfolded and/or aggregated prior to the gradient temperature increase; the resultant melting curve starts with higher fluorescence and decreases as the protein continues to aggregate and exclude SYPRO. A notable trend seen in most successful melting curves was a higher baseline RFU in most conditions for HisTraG* Δ relative to TraG*His, which is likely an artifact of difficulties in normalizing thermofluor data.

Trends in the thermofluor analysis of HisTraG* Δ are contrary in some respects to those of TraG*His. Buffers with a pH lower than 5.2 do not confer a notable HisTraG* Δ melting curve, however many buffers that did not produce a melting curve for TraG*His due to high pH values were seen to do so with HisTraG* Δ , albeit the T_m values were lower than average (Figure 3.6A). As seen in Figure 3.6B, 50mM NaCH₃COO pH 5.2 may have provided a higher T_m to HisTraG* Δ

than TraG*His, however the melting curve of HisTraG* Δ is poor with numerous abnormalities prior to the main peak at 54.37°C. Similar to TraG*His, buffers with 50mM NaCl concentration were seen to provide a more favourable melting curve for HisTraG* Δ unfolding in comparison to buffers with 200mM NaCl. Lower NaCl concentration appeared to maintain folding of both proteins as temperature was increased. This may be caused by increased heat capacity of the solvent when less ionic strength is present, or from higher ionicity affecting intermolecular interactions by masking charges in TraG*, or both phenomena contribute to the result. As the RotorGene Q measures the internal temperature of the system rather than the temperature of the buffers themselves it is possible the buffers with higher salt content absorbed more kinetic energy at the same temperatures and therefore caused the proteins to unfold at lower temperatures. As well, increasing salt concentrations may create ‘salting out’ effects; as ionicity is increased salt ions can mask the inter- and intra- molecular ionic interactions that prevent protein unfolding at increased temperatures (Dumetz et al., 2007).

In comparing the thermofluor data from HisTraG* Δ and TraG*His it is evident that HisTraG* Δ has higher overall stability (Figure 3.6A). Without a buffer, HisTraG* Δ in ddH₂O shows a melting curve with a T_m 4.8°C higher than TraG*His in ddH₂O, and the HisTraG* Δ melting curve is superior to many of those in which TraG*His is in a buffer system (Figure 3.6A & B). The optimal buffers for both proteins were shown to have a pH near the predicted pI of the protein; the pI of HisTraG* Δ is 5.89 (Appendix A). A peculiar result was the unfavourability of TraG*His for 50mM MES pH 6.0, which was the buffer which gave the highest T_m for HisTraG* Δ . As TraG*His also had its highest T_m in a pH 6.0 Citrate buffer, it is interesting that MES destabilizes TraG*His at the same pH.

Further optimization of TraG*His buffer conditions was performed through the alteration of the pHs for the three buffers that gave the highest T_m values; 50mM NaCH₃COO pH 4.0, 50mM NaCH₃COO pH 5.2, and 50mM citrate pH 6.0. All buffers had pHs changed above and below in 0.2, 0.4, and 0.8 -unit intervals, centered around the previously used pH; the most favourable melting curve shape found was for 50mM NaCH₃COO pH 5.6 (Figure 3.7). Citrate pH 6.0 has a higher T_m values, however as vapour diffusion experiments showed NaCH₃COO pH 5.6 provided spherulite formation (as described in Section 3.5), it was deemed as the optimal buffer for attempting further optimization experiments with various salts common in crystallography.

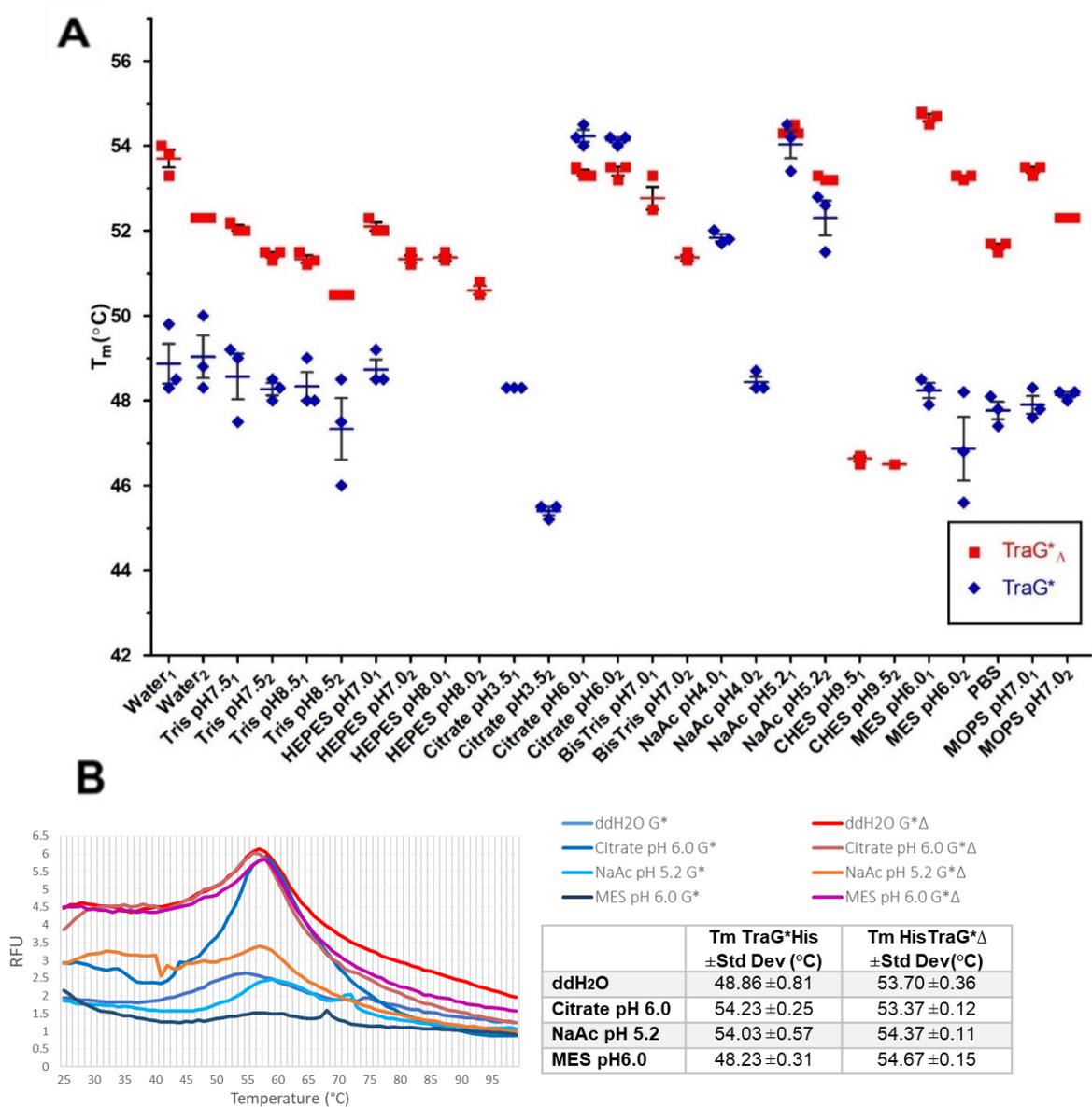


Figure 3.6: Thermofluor analysis of TraG*His shown in blue and HisTraG* Δ shown in red. **A)** the melting temperatures (T_m) in triplicate, as well as means with standard deviations, of both proteins as they were incubated in each buffer condition shown on the x-axis. Each buffer is at a concentration of 50mM (except ddH₂O and PBS), where ₁ represents the addition of 50mM NaCl in the buffer and ₂ includes 200mM NaCl. Instances where no data points appear for a buffer condition specify no clear melting temperature was observed in the melting curve, indicating the destabilizing nature of the buffer on TraG* folding. **B)** Graphs displaying fluorescence output as a function of temperature is shown for a sample of the buffer conditions, in which all buffers listed feature 50mM NaCl. NaAc represents NaCH₃COO. Mean melting temperatures with standard deviations for these conditions are shown in the table.

In analyzing the results of the thermofluor analyses for the identification of salts efficient in stabilizing TraG*His and HisTraG* Δ some similarities are seen (Figure 3.7). The salts that best stabilize both proteins based on favourable melting curve shape and T_m are those with dianions, with sulfate as the ideal anion. As well Mg^{2+} seemed to have a favourable effect on both proteins, providing optimal shapes to melting curves whenever present. $NiCl_2$ and NaF provided observably poor melting curves; despite providing high T_m values the lack of fluorescent signal increase was deemed destabilizing. The optimal salts in these buffer conditions were 50mM $(NH_4)_2SO_4$ for TraG* His and 50mM KCl for HisTraG* Δ .

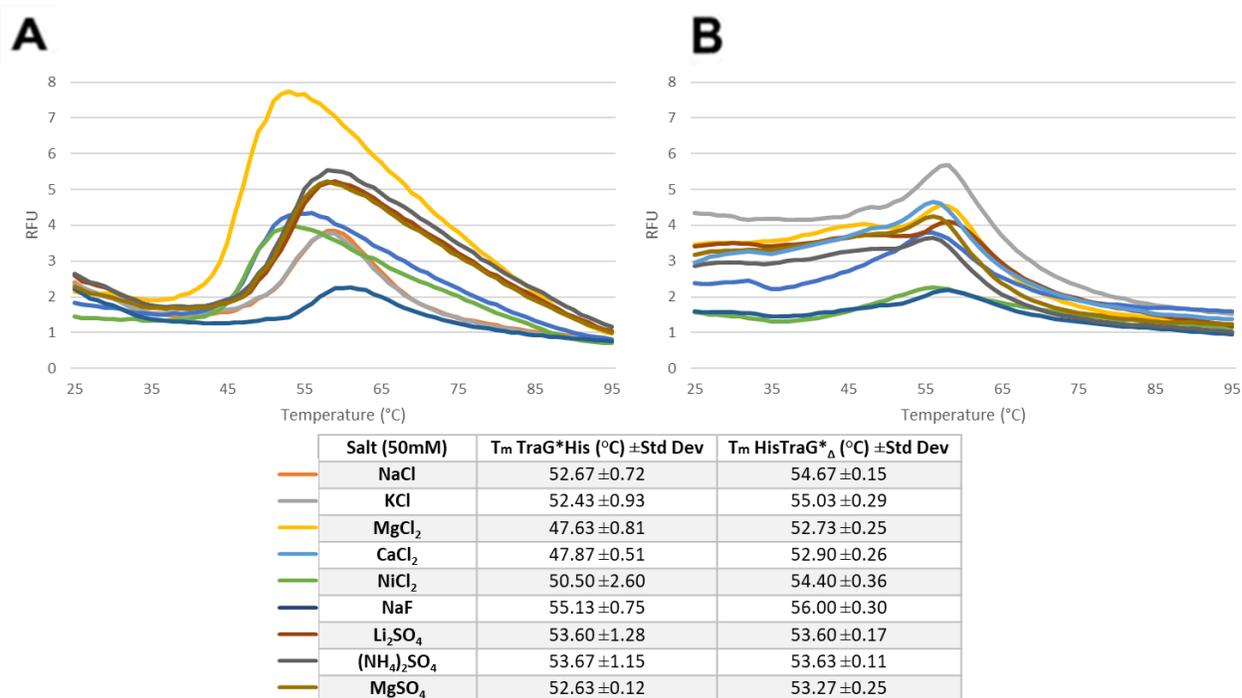


Figure 3.7: Optimization of buffer conditions for TraG* stability through screening of 9 salts common in crystallography. 50mM $NaCH_3COO$ pH 5.6 was used in **A)** for TraG*His as it was determined to be the optimal buffer, while experiments using HisTraG* Δ shown in **B)** were performed in 50mM MES pH 6.0. Averages of each fluorescence reading at every temperature point were taken for three trials to produce the above graphs, and the corresponding mean melting temperatures with standard deviations are listed in the above table.

3.5 Crystallisation of TraG*His

Previous crystallisation screening experiments of untagged TraG* showed the protein forms microcrystals in conditions from pH 6.5-8.0 with 200-250mM (NH₄)₂SO₄ and 18-25% PEG 3350 (w/v). Purified TraG*His was screened in 96 well sitting drop vapour diffusion plates with commercial screens in concentrations of 2, 5, 10, 20, 30 and 40 mg/mL. As no optimizable crystal hits were observed, 24 well hanging drop screening was performed with conditions similar to those suitable for the crystallisation of untagged TraG*. Spherulites were seen when optimizing conditions with 100mM NaCH₃COO, and the largest ones were seen in panel 1 of Figure 3.8B. As shown in the other panels, successive rounds of microseeding (μ seeding) from these spherulites and subsequent crystals grown therein changed spherulite morphology. μ seeding was performed through isolation of a crystal in 1 μ L of reservoir by use of a microloop, addition of the 1 μ L to 99 μ L of reservoir solution in a microfuge tube with a seed bead, followed by 2 minutes of vortexing prior to serial dilution and addition into droplets. A diffraction quality 3-dimensional crystal was never obtained; all spherulites shown grew mainly about two axes and had little depth, forming disk- and crater- like structures. The spherulites grown in panel 5 were imaged prior to isolation through looping into a separate droplet which was pipetted into a microfuge tube, denatured and identified as protein by SDS PAGE (Figure 3.8A). Comparing the profile of the TraG*His spherulite to a 1 μ g sample of fresh TraG*His and denatured TraG*His demonstrates that common degradation products of TraG*His rather than full length protein are responsible for producing these spherulites. As well, a 1 μ g sample of purified HisTraG* after a single freeze-thaw cycle shows new degradation bands not initially present after IMAC purification, corresponding to the same approximate size as the protein species seen in lane in which the spherulite sample was added. However, when HisTraG* Δ is purified in the same manner no degradation bands are seen.

This further supports the hypothesis for the presence of an IDR in the N-terminus of TraG*, as the removal of this region prevents any of the phenomena seen when the N-terminus is present. No diffraction quality crystals of HisTraG* or HisTraG* Δ have been observed at this moment, however crystal screening of these variants was not as thorough.

Attempts to diffract the spherulites produced from TraG*His vapour diffusion experiments were unsuccessful; the only diffraction which produced observable spots was that seen in Figure 3.9C. All other diffraction attempts produced ice rings with no observable spots, similar to the diffraction image seen in Figure 3.9F. Ice rings were more common in conditions in which low precipitant concentration was used in the reservoir solution, as less precipitant could be effectively used to cryo-protect these crystals without causing crystal degradation prior to flash-freezing. Addition of glycerol to 10% (v/v) dissolved all spherulites attempted for diffraction, however some spherulites were further cryo-protected by the addition of ethylene glycol to a final concentration of 10% (v/v). The spherulite that produced some spots from diffraction was shown to be crystalline through its birefringent properties (Figure 3.9A and B), however as the reflections are well spaced in the resulting diffraction pattern this spherulite was likely composed of a small molecule packing into a small unit cell (Figure 3.9C). As many spherulites similar to those seen in Figure 3.9D and E produced no diffraction, the composition of these spherulites cannot be confirmed.

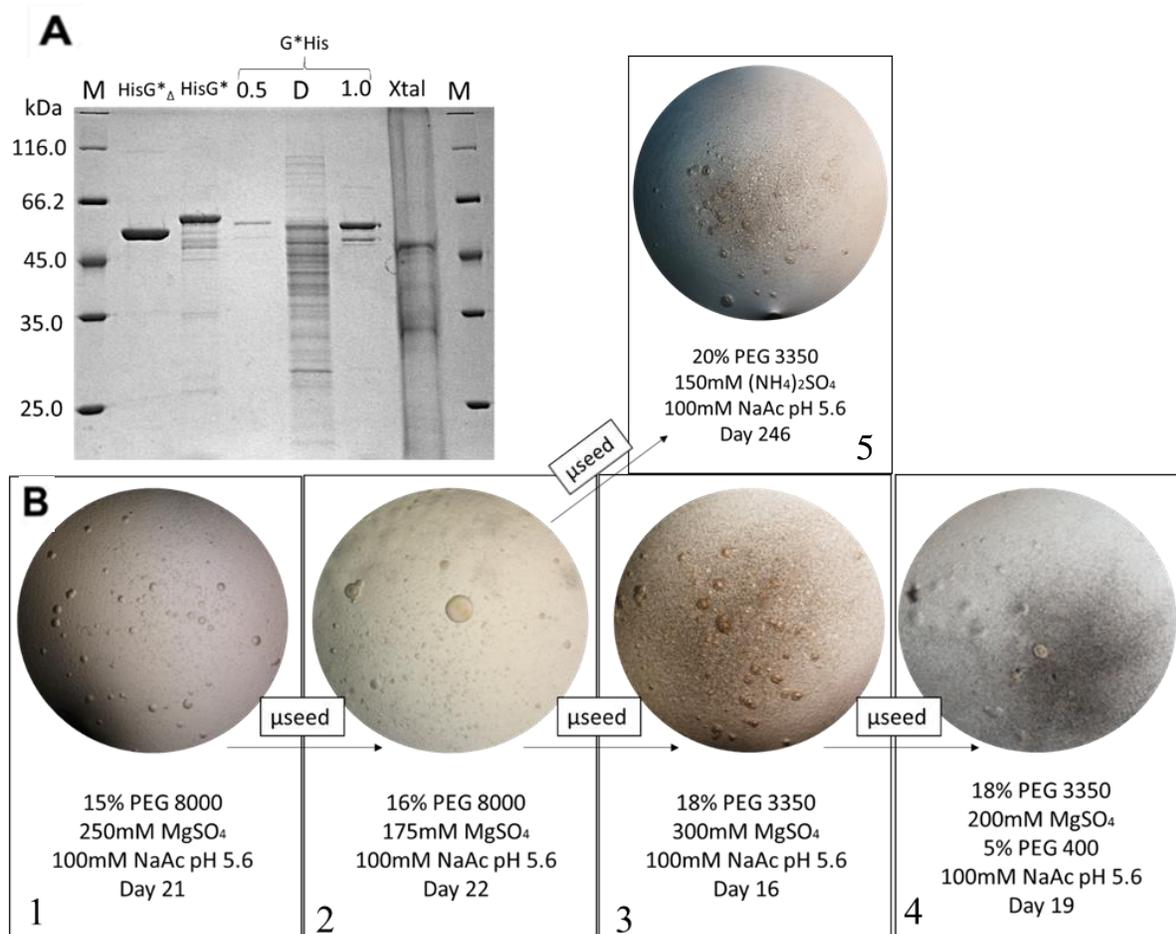


Figure 3.8: **A)** Purified TraG* constructs post desalting, concentration and quantification as shown on a 12.5% SDS PAGE. The image shows HisG* Δ : 1 μ g of HisTraG* Δ , HisG*: 1 μ g of HisTraG*, TraG*His samples include a 0.5 μ g sample, a denatured sample (left at 4 $^{\circ}$ C for 3 weeks), and a 1 μ g sample. Xtal represents the crystals seen in panel 5 which were isolated into a 1 μ L droplet containing a 1:1 mix of the reservoir solution and ddH₂O using a microloop, then pipetted into a microfuge tube and prepared for loading onto the gel as described in Section 2.3. M: unstained protein marker (ThermoFisher #26610). **B)** TraG*His crystals from hanging drop vapour diffusion experiments. Spherulites were produced from successive rounds of μ seeding and optimization, in which either 1 μ L of 40mg/mL TraG*His was mixed with 1 μ L of reservoir (Image 1), or 1.2 μ L of 40mg/mL (Image 2&5) or 30mg/mL (Image 3&4) TraG*His with 0.8 μ L reservoir and 0.4 μ L of a seed stock from the previous image's crystal (Images 2-4).

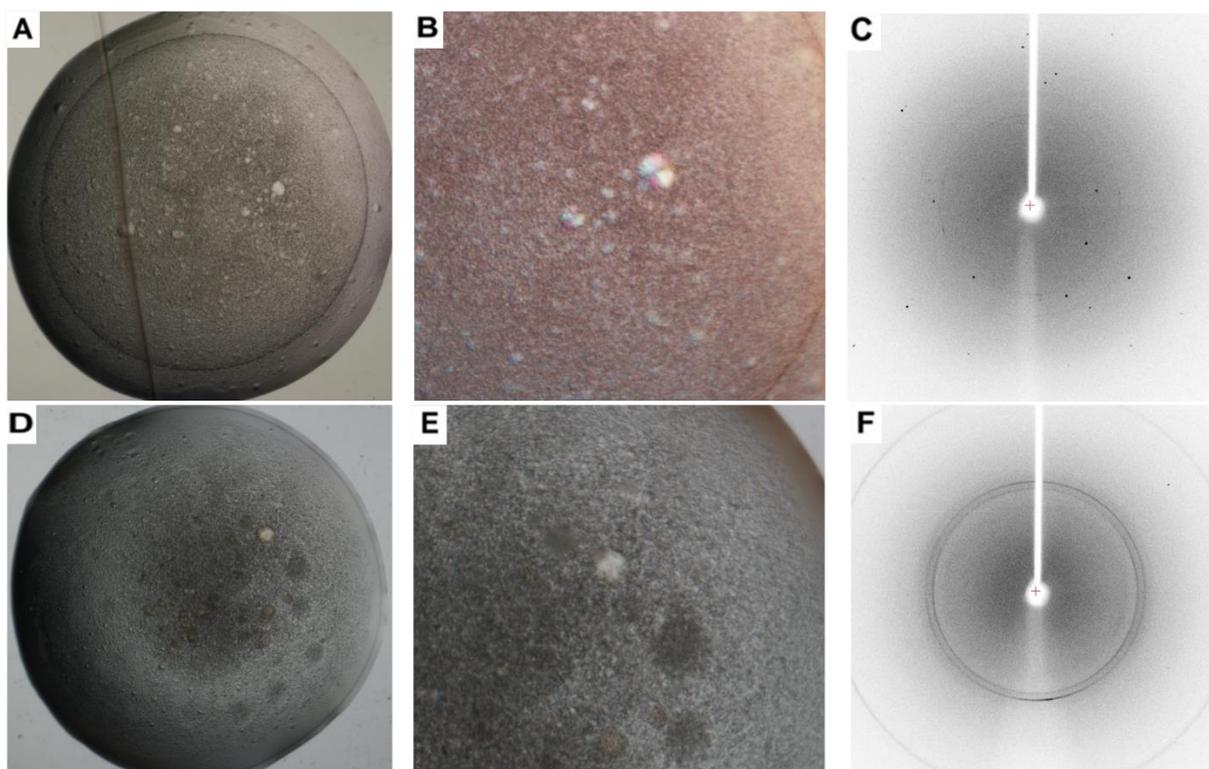


Figure 3.9: X-Ray diffraction of spherulites produced in a TraG*His vapour diffusion experiment. **A)** a hanging drop from the mixing of 40mg/mL TraG*His in a 1 μ L:1 μ L ratio with 350mM MgSO₄, 0.1M NaCH₃COO pH5.6, and 22% PEG3350 (w/v), taken after 21 days of vapour diffusion at 19°C. A zoomed in spherulite from this droplet is shown in **B)**, with a cross-polarizing filter on to display the crystalline nature of the spherulite through its bireferingent properties. This image was taken after 35 days of vapour diffusion. On the same day the spherulite focused on in **B)** was looped with a 100 μ m microloop, soaked for 2 minutes into the same buffer condition with 25% PEG3350 (w/v), flash frozen and diffracted. **C)** displays the ice rings and a diffraction pattern not consistent with a protein crystal from the 3rd image of 4 collected from diffraction screening. **D)** a hanging drop from the mixing of 50mg/mL TraG*His in a 1 μ L:1 μ L ratio with 350mM MgSO₄, 0.1M NaCH₃COO pH5.6, and 20% PEG3350 (w/v), taken after 25 days of vapour diffusion at 19°C. A zoomed in spherulite from this droplet is shown in **E)**, taken after 16 days of vapour diffusion. On day 27 the spherulite focused on in **E)** was looped with a 100 μ m microloop, soaked for 2 minutes into the same buffer condition as the resevoir but with 23% PEG3350 (w/v), flash frozen and diffracted. **F)** displays the ice rings from the 1st image of 4 collected from diffraction screening.

3.6 CD of TraG* Variants Provides Evidence of an N-terminal IDR

Findings from performing CD on HisTraG* and HisTraG* $_{\Delta}$ are aligned with the hypothesized presence of an IDR in the N-terminus of TraG*. As seen in Figure 3.10, the CD spectra of each variant appears to be mainly α -helical based on the characteristic negative $\Delta\epsilon$ peaks at 222nm and 208nm (Greenfield, 2007). This confirms that many of the predictive software were incorrectly labeling many regions of the protein as intrinsically disordered due to the lack of homologous protein sequences with solved structures. Deconvolution by BeStSel shows differences in the extent to which the two variants fold into canonical α helices (Figure 3.10, Table 3.2) (Micsonai et al., 2015). The model providing the best fit for HisTraG* showed that 76.1% of the protein was predicted to be α -helical, albeit 21.1% of helices are of a bent or imperfect topology (Figure 3.10A). Turns were of low abundance in the overall structure of the protein at 4.8%, and the remaining 19.1% of the protein was deemed as ‘other,’ which represents unstructured regions undetectable by CD. BeStSel also groups uncommon helical forms with lowered propensity to affect the resultant $\Delta\epsilon$ into the ‘other’ section, such as 3_{10} and π helices, (Geetha, 1996; Micsonai et al., 2015; Wilman et al., 2014). The spectrum of HisTraG* $_{\Delta}$, which appears to have better defined peaks at 222 and 208nm, is predicted to have 0% ‘other’ character and is an all α -helical protein, with 5.3% of the protein predicted as representative of turns between the α -helices. This provides a strong indication that the 45 residues removed from the protein are intrinsically disordered; the N-terminal IDR in TraG* could be causing distortions in the remaining protein fold based on the high relative change in the predicted percentage of assigned conformations.

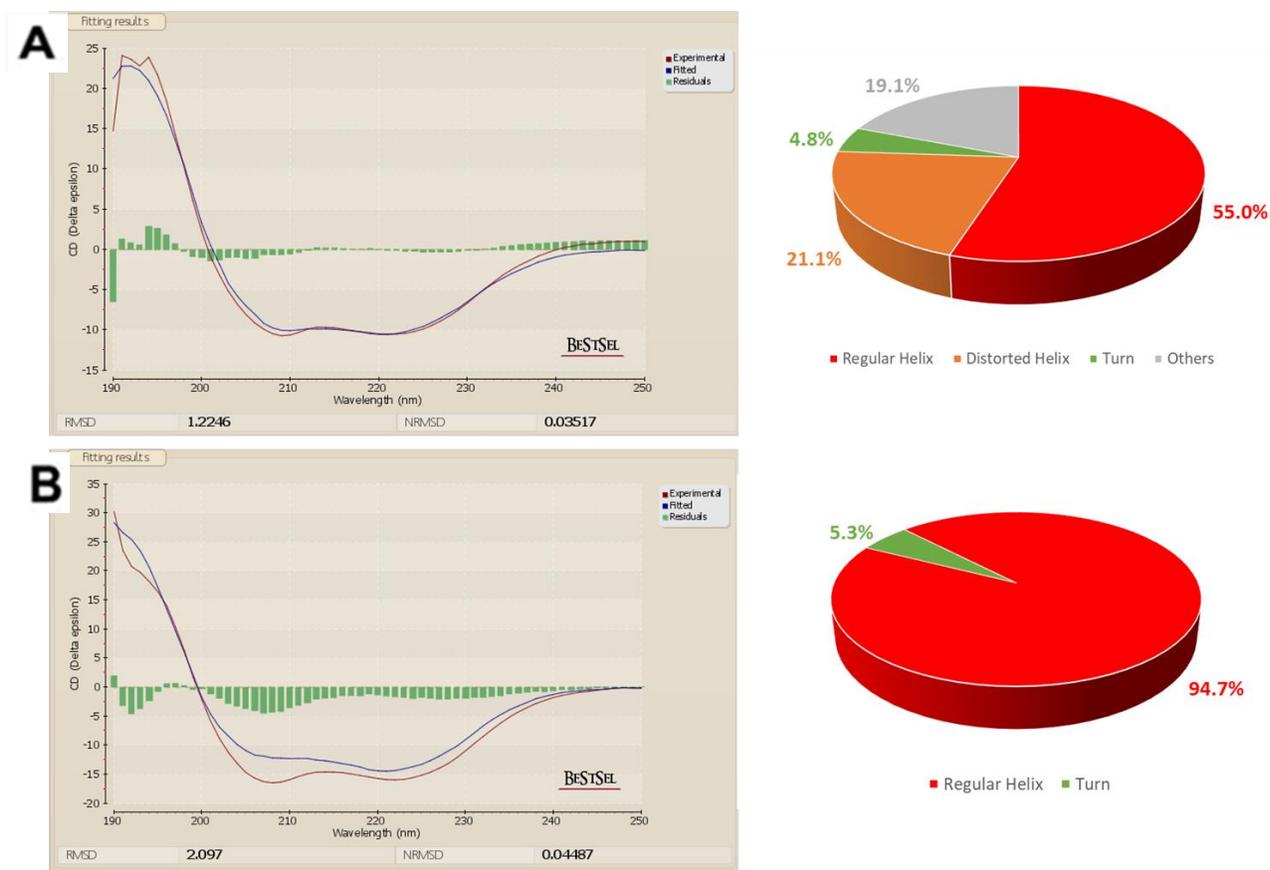


Figure 3.10: The CD spectra of **A)** HisTraG* and **B)** HisTraG* Δ as displayed by BeStSel (Micsonai et al., 2015). Experiments were performed at a protein concentration of 2 μ M, where the buffer system was 10mM HEPES pH 7.0, 10% glycerol when measured in the spectrophotometer, and the resultant data was normalized by the blank buffer. Fitting spectra of the TraG* variants (experimental data shown in red) to those of proteins with known structure (shown in blue) allowed for the prediction of the secondary structure composition of each protein, as seen in the pie charts beside each spectrum. BeStSel uses pie charts to provide a visual representation of the predicted secondary structure content of each protein based on the CD spectra. Distorted helices represent regions in which the α helix geometry is imperfect resulting in helices more closely resembling a 3_{10} helix, while regions referred to as ‘Other’ include 3_{10} and π helices, but also loops that may not change the direction of polarized UV light and are therefore undetectable by CD (Geetha, 1996; Micsonai et al., 2015; Wilman et al., 2014).

Table 3.2: Secondary structure of HisTraG* and HisTraG* Δ as predicted by analysis of CD spectra by BeStSel (Micsonai et al., 2015).

Protein	Regular α -helix	Distorted α -helix	Turns	Other
HisTraG*	55.0%	21.1%	4.8%	19.1%
HisTraG* Δ	94.7%	0%	5.3%	0%

Urea denaturation studies were performed to support the conclusion surmised from the thermofluor data that TraG* Δ has higher stability than TraG*. Spectra and tables displaying predicted secondary structure topology by BeStSel are displayed in Figure 3.11, and from comparing spectra of HisTraG* to HisTraG* Δ it is evident that the presence of residues A1-A44 destabilizes TraG*. HisTraG* displayed higher susceptibility to unfolding based on the direct decrement of $\Delta\epsilon$ at 222 and 208nm as the urea concentration was increased. The secondary structure of HisTraG* Δ remained relatively intact until incubation with 4.0M urea, while the secondary structure of HisTraG* was partially unfolded at 1.0M urea based on the topology analysis from 260-200nm by BeStSel. As data from 207-200nm is subject to variability due to the UV absorption properties of urea, the topology analyses may be inaccurate despite correction using a buffer blank.

Further investigation of the unfolding processes of TraG* variants was performed through the analysis of a finer range of urea concentrations from 3.0-4.0M (Figure 3.12). HisTraG* is shown to have a substantial reduction in secondary structure character between 3.0M - 3.2M urea (Figures 3.12 A and B respectively). The spectra of HisTraG* Δ shows $\Delta\epsilon$ values are approaching 0 but the spectrum retains the characteristic shape of an all α -helical protein until 3.6M urea (Figure 3.12 D). At 4.0M urea the CD spectra of the TraG* proteins overlap, indicating the same degree of unfolding has been reached for both protein samples. This data aids in concluding that

differences in $\Delta\epsilon$ throughout the CD experiments are due to distinct secondary structures of the respective protein samples and not due to variability in concentrations, as the molarity of a chiral molecule affects the extent to which molar ellipticity is changed during a CD experiment. Due to the presence of spectra overlap at a point in which the two proteins are not entirely unfolded (as seen by the further increase of $\Delta\epsilon$ at 5.0M urea in Figure 3.11), this assures the accuracy of CD results in dictating the secondary structure of the variants, and therefore TraG* exhibits a stabilized secondary structure by the removal of the putative N-terminal IDR.

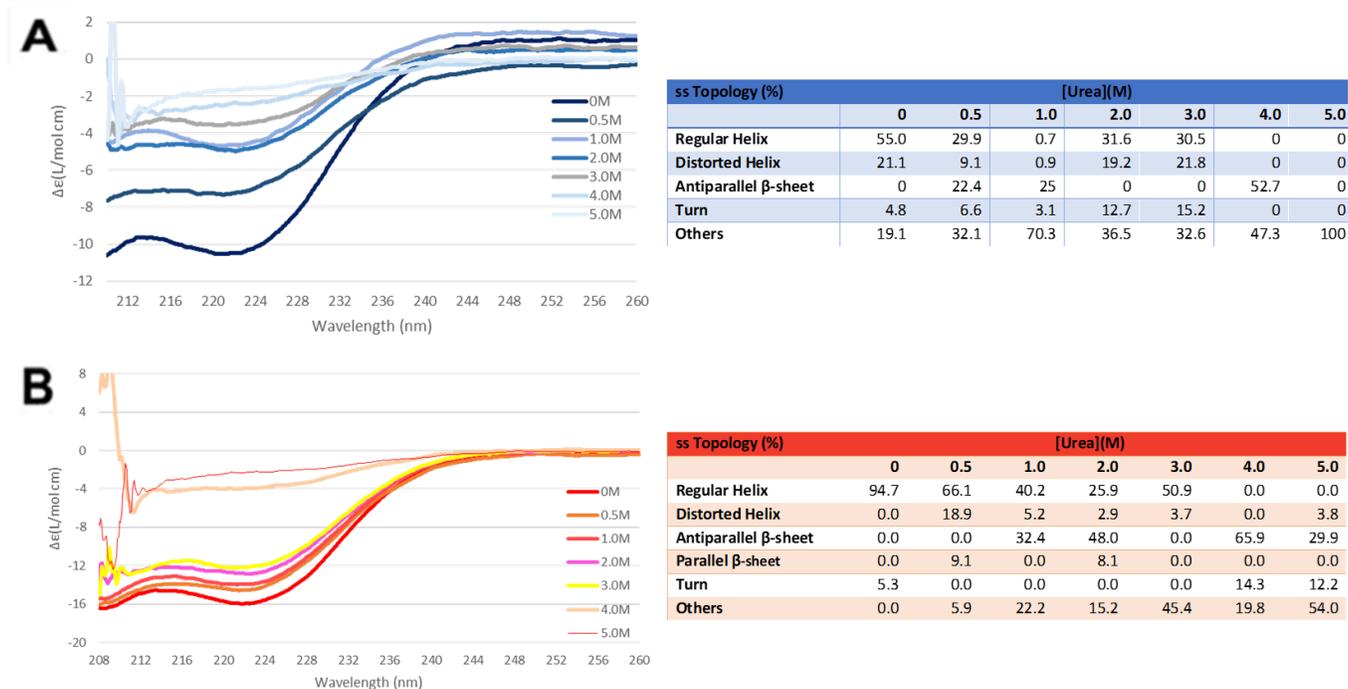


Figure 3.11: Urea denaturation of **A)** HisTraG* and **B)** HisTraG* Δ as measured by CD spectroscopy. All experiments were performed with a final protein concentration of 2 μ M and were incubated in the respective concentrations of urea (shown in the figure legends) for 1 hour at 25 $^{\circ}$ C prior to UV measurement. The resultant data was normalized by the buffer. Molar ellipticity measurements were obtained every 0.1nm from 260-190nm, however due to inconsistencies in UV absorption of urea at wavelengths between 207-190nm only data from 260-208nm is shown. Data from 260-200nm was entered into BeStSel for deconvolution, and the resultant analyses were tabulated as percentages (Micsonai et al., 2015).

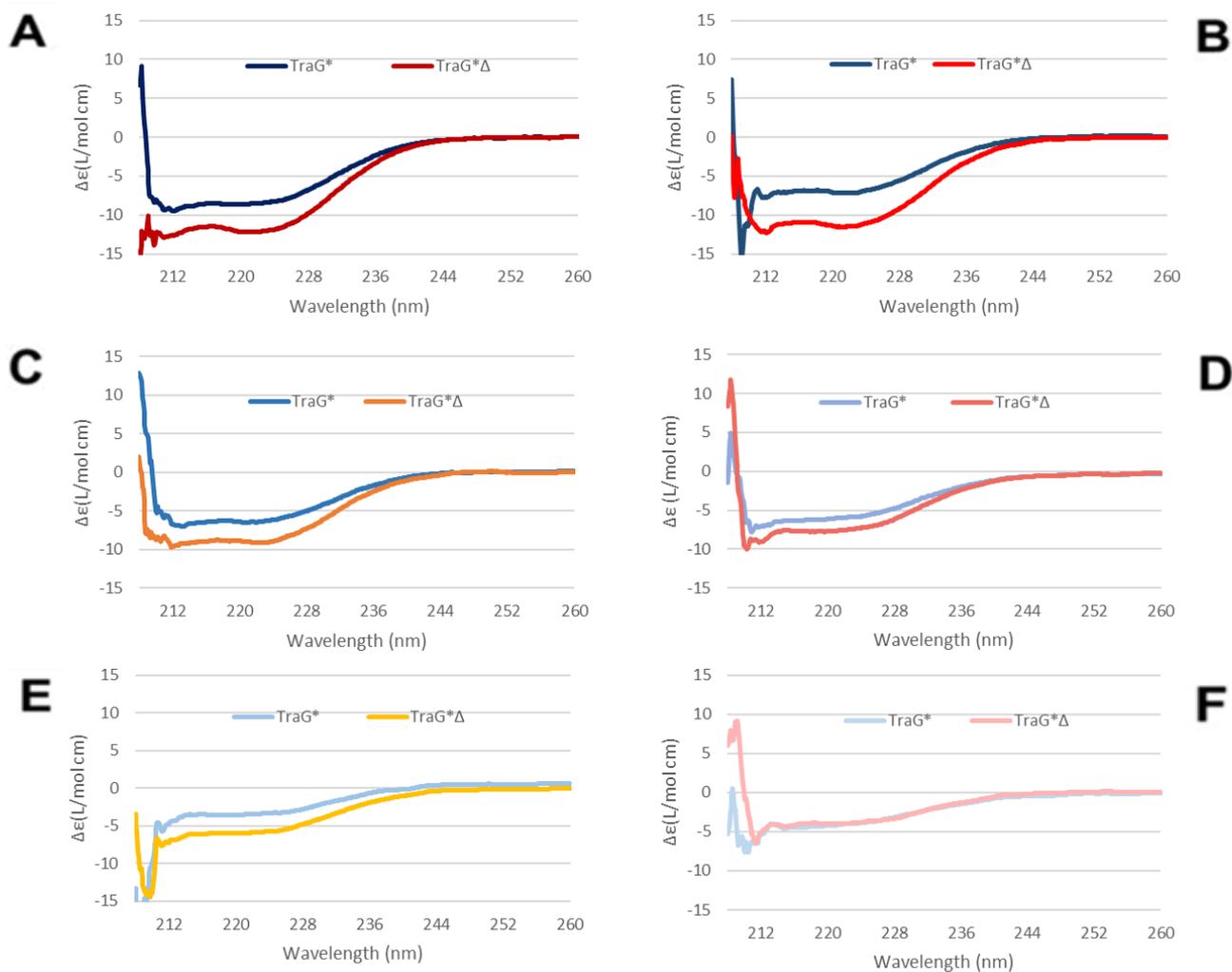


Figure 3.12: Denaturation experiments sampling refined urea concentrations between 3-4M using the proteins HisTraG* and HisTraG* Δ . Proteins at concentrations of 2 μ M were incubated at 25 $^{\circ}$ C for 1hr in **A)** 3.0M **B)** 3.2M **C)** 3.4M **D)** 3.6M **E)** 3.8M **F)** 4.0M urea prior to CD measurement. The resultant data was normalized by the blank buffer. Molar ellipticity measurements were taken every 0.1nm from 260-190nm, however due to inconsistencies in UV absorption of urea at wavelengths between 207-190nm only data from 260-208nm is shown.

3.7 Analysis of TraG* by Global HDX MS Confirms Dynamicity in its N-terminal Residues

The goal of global HDX MS experiments was to compare relative deuteration rates of the TraG* variants independent of their inherent mass such that their dynamics could be elucidated, thus providing information regarding the solvent exposure of the truncated region. ESI-MS analysis of HisTraG* and HisTraG* Δ shows that the charge states of highest abundance in both proteins were +15, +14, and +13, and were the peaks used for all analyses in comparative experiments. The protein masses observed include $54,447.91 \pm 0.32$ Da and $51,740.90 \pm 0.45$ Da respectively (Figure 3.13), while the actual masses are expected to be 56262.90 and 51573.94 Da (Appendix A). These masses were calculated from m/z values of the prominent peaks using ESIprot (Winkler, 2010). The increase of approximately 160 Da in HisTraG* Δ is likely due to H₂O and other solvent components interacting with the protein. The decrease in mass for HisTraG* is contrary to experimental evidence as no degradation products were seen when samples were analyzed by SDS PAGE. The ESIprot algorithm calculates a mass around 58.5 kDa if the m/z values of the right-most distribution of the peaks are entered, which causes all of the charge states to shift by +1 (Winkler, 2010). It is possible that masses have been incorrectly predicted by the utilized primary sequence analysis tool ProtParam, or that ionization is causing a truncation of approximately 1815 Da (~17 aa) at the higher charge states due to the higher energy input (Gasteiger et al., 2005). The first 17 residues of the protein are from the His₆ affinity tag of the pET28a vector and are likely to be cleaved due to a thrombin recognition site at this location, indicating the possibility for the resultant cleavage product to be the highly recognized species by ESI MS.

Other notable features in the spectrum of HisTraG* not appearing in the spectrum of HisTraG* Δ are the extra distributions of small shoulders immediately to the left of the +16 and

+15 charge state peaks, which correspond to a 60Da difference between the peaks, indicating the presence of a common adduct to the protein. As the most favourable buffer condition for TraG*His was determined to be in NaCH₃COO it is likely that this mass increase corresponds to the acetate anion from NH₄CH₃COO. As it is not seen in the spectrum of HisTraG* Δ it is possible that the N-terminal IDR is responsible for its binding, and thus provides the protein with enhanced stability. Small peaks intermittent in the abundant charge state distributions are likely due to a low percentage of ESI droplets that happen to contain two molecules of HisTraG*, rather than being caused by a low proportion of protein dimerization; a common occurrence when backing pressure of the Synapt G1 system is raised to 3.9 mBar to accommodate ESI MS of a protein larger than 50kDa (Wilm, 2011). Another notable aspect of the mass spectra is the increased presence in a gaussian distribution of high charge states (+25-+22) seen in HisTraG* Δ but not HisTraG*. This is indicative of a small population of unfolded or denatured proteins, which is unexpected as the deletion mutant was designed to be more stable (Donnelly et al., 2019).

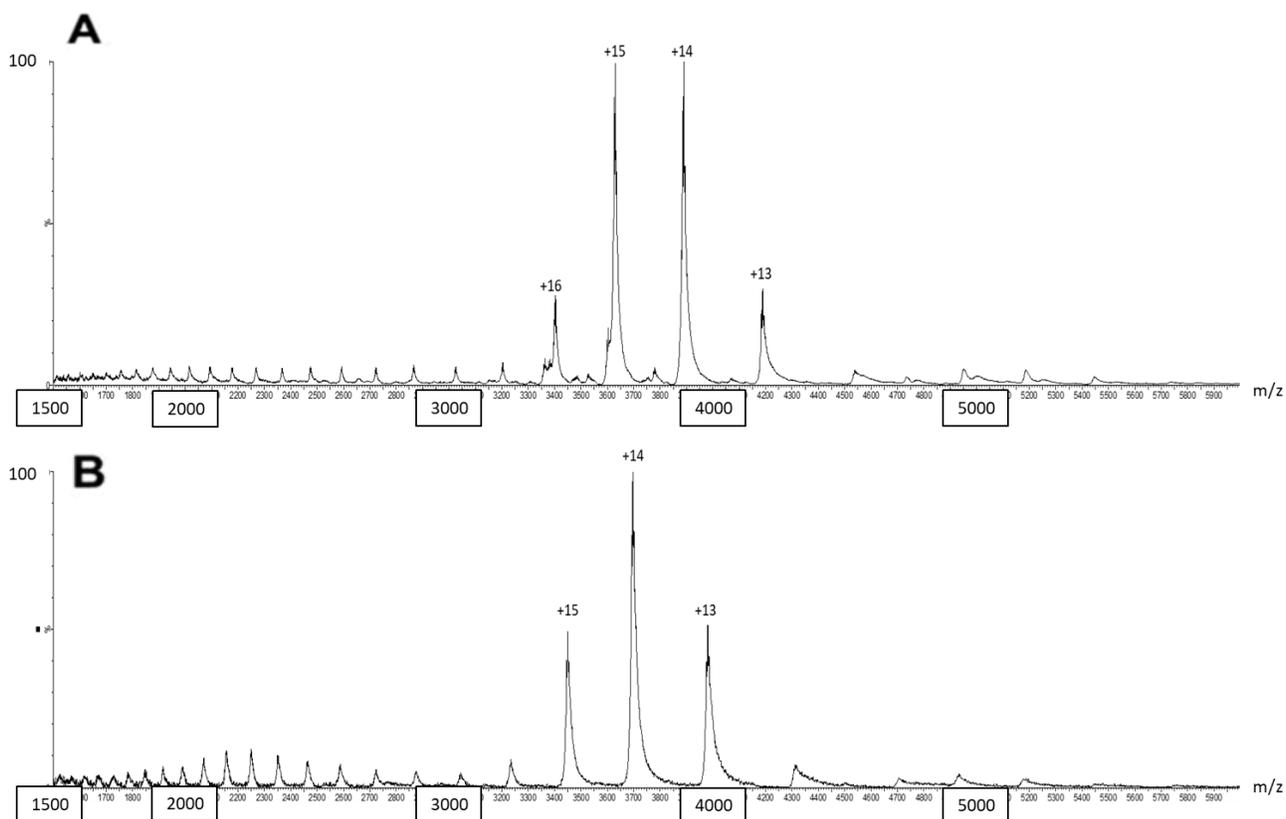


Figure 3.13: ESI-MS analysis of native folds for **A)** HisTraG* and **B)** HisTraG* Δ . Proteins in 50mM NH₄CH₃COO pH 6.8 were flowed through the ESI capillary at concentrations of 5 μ M at 8 μ L/min. Monomers for both proteins are seen in the 3200-4300 m/z range with prominent charge states of +16 to +13 for TraG* and +15 to +13 for TraG* Δ . These provided average molecular weights of 54,447.91 \pm 0.32 Da and 51,740.90 \pm 0.45 Da respectively, calculated using ESIprot (Winkler, 2010).

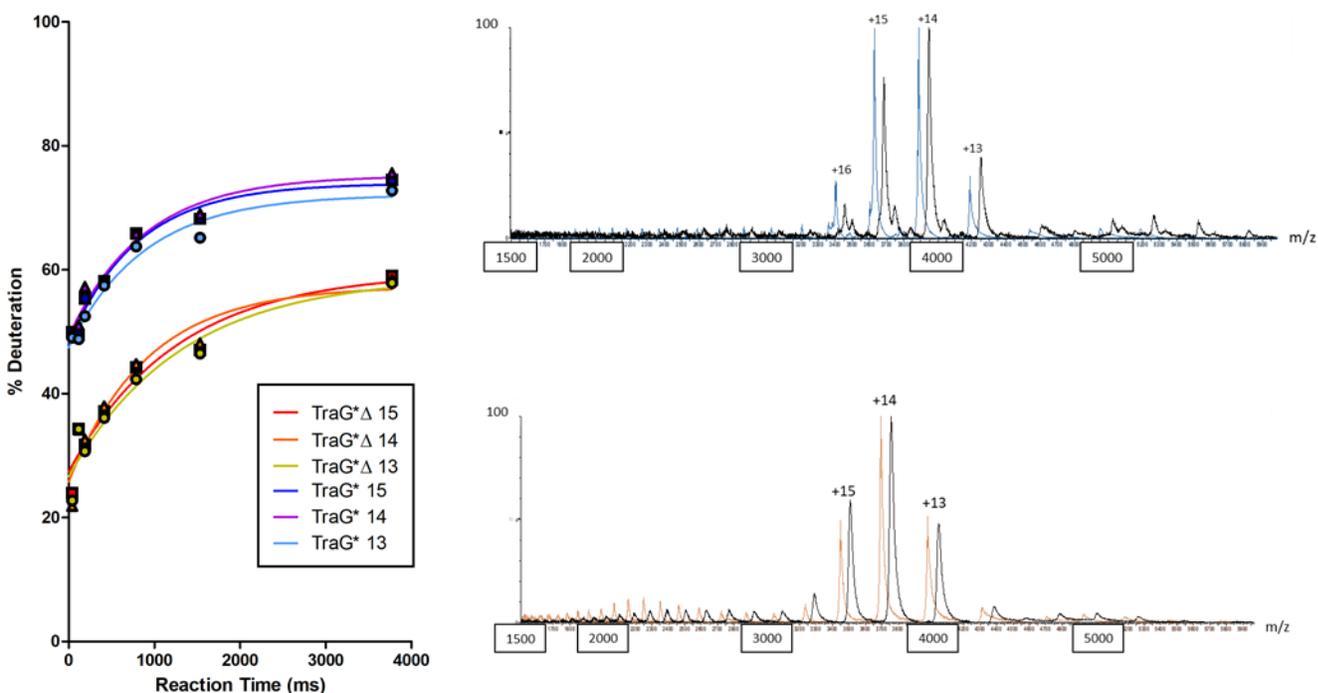


Figure 3.14: Global HDX-MS analysis of HisTraG* and HisTraG* Δ . Measurements were taken at 0, 1, 2, 5, 10, 20, 50, and 100ms time points with 30 μ M protein inserted at a rate of 4 μ L/min, and D₂O at 16 μ L/min. A fully deuterated spectrum was taken as well to provide a 100% deuteration reference point to which the other time point measurements were compared. The corresponding graphs were fit to a one phase decay distribution. The mass spectra to their right with corresponding colour schemes show undeuterated spectra in colour and the fully deuterated spectra in black.

Global HDX MS analysis of the TraG* variants at different deuteration time points further supported the hypothesis of an N-terminal IDR and provided evidence that its removal further stabilized the protein. Deuterium uptake was calculated as a percentage based on comparisons to the fully deuterated samples; fully deuterated proteins had a calculated uptake of 928.5 ± 4.35 and 969.1 ± 16.9 deuterium atoms (average of the +15–+13 charge states) for HisTraG* and HisTraG* Δ , respectively. This calculation accounts for an effect on the perceived deuteration caused by comparisons made between proteins that differ in overall protein size. The amount of

overall deuterium uptake was higher for the truncated protein however, indicating differences in TraG* dynamics upon removal of the N-terminal IDR.

Upon the comparison of deuteration rates fit to a one-phase decay using GraphPad, deuterium uptake was significantly higher in HisTraG* in comparison to HisTraG* $_{\Delta}$ (Figure 3.14). However, the overall rate of deuteration appear to be similar based on the shape of the limited growth curves ($1-e^{-kt}$), indicating TraG* and the TraG* $_{\Delta}$ perform similar dynamic movements. After 40 ms of D₂O mixing prior to ESI, deuteration was observed to be ~50% of the experimental maximum for HisTraG*, while HisTraG* $_{\Delta}$ demonstrates ~23% of maximum deuteration. As momentary exposure to D₂O caused a comparatively rapid accumulation of deuterium onto HisTraG* relative to HisTraG* $_{\Delta}$, the truncated region must be highly dynamic.

CHAPTER 4 – CONCLUSIONS AND FUTURE WORK

4.1 The N-terminal Dynamic Region of TraG* and its Potential Role for Protein Function

TraG is a multifunctional protein in the F-T4SS. It is composed of a membrane-bound N-terminal domain required for pilus generation and a C-terminal periplasmic domain (TraG*) required for Eex and Mps. TraG* therefore requires plasticity in its structure to facilitate these different functions that are thought to require numerous PPIs. Based on the findings presented in this thesis it can be stated that the N-terminal region of TraG* is highly dynamic and its presence destabilizes the protein, as confirmed through comparisons of thermofluor and CD results to the results from the truncation mutant. These characteristics are not sufficient in identifying the region as an IDR; IDRs shorter than 50 residues are not uncommon, however they typically serve a functional purpose in the native protein (Van Der Lee et al., 2014). As this region in TraG has not been shown to be intrinsically disordered, the region cannot be classified as an IDR. However the TraG* constructs explored in this report provided evidence that the region from A452-A496 is prone to degradation and is highly solvated. The region likely serves as a flexible linker in the full-length protein, however when expressed separately from the N-terminal membrane bound portion this region of TraG* displays higher dynamicity due to its expression as a synthetic construct.

As the only deletion mutant tested was from the truncation of 45 amino acids, the size of the disordered region has not been defined in full, however this finding has many implications. Many structural predictive software such as Phyre2, IUPred2A, ANCHOR 2A, and many more, improperly identified secondary structure properties of TraG*; overestimating disorder content to be present in approximately 50% of the protein (Figure 3.1B and 3.2A). The CD analysis of TraG* ensured the protein is not largely disordered; some regions are instead predicted to feature distorted helices, with 19.1% being disordered (Section 3.7). This confirms that these programs were falsely

assigning regions of TraG* as disordered due to their inability to identify homologs with known structures, enforcing the postulation regarding the novelty of this protein structure. The identification of a highly dynamic region by the PSP software was successful; therefore other popular algorithms should be modified to include disorder predictions based on long range pi-pi contact frequencies to improve their accuracy (Section 3.1).

Confirming the presence of a dynamic N-terminus in TraG* elucidates the basis for many of the phenomena observed in purification and crystallisation of TraG* and TraG* Δ . HisTraG* purified with less degradation products than TraG*His, as is logical if the N-terminus is prone to auto-cleavage (Section 3.3). It is possible that TraG* Δ His did not express properly due to a lack of protein stability when no residues are present before T497; perhaps the formation of an α -helix vital to the native fold of the structure is disrupted by increased solvent exposure when this region is N-terminal, thus lowering the cellular longevity of the protein. A similar phenomenon has been seen before in sperm whale myoglobin, however deletions were performed with knowledge of the protein structure (Ribeiro and Ramos, 2005).

The mechanism behind spherulite formation in crystallisation studies of TraG*His can be explained by the presence of a highly dynamic N-terminus of 45 amino acids. SDS PAGE of spherulites showed the main component of these crystals is from degradation products, which corroborates a basis for the duration prior to visualizable spherulite growth (Section 3.5). It is likely the time elapsed allows for sufficient degradation of TraG* such that the resultant N-terminally truncated TraG* is of sufficient abundance to be self-concentrated into a nucleation point for a crystal. The presence of full length TraG* would disrupt proper crystal formation, leading to the observed spherulites. Intermittent stacking of full-length TraG* in crystals of the protein with truncations has the potential to disrupt a proper crystal lattice from forming as it would

alter the shape of the respective unit cells. The incessant motion of the N-terminal region could prevent proper packing or disrupt a vital crystal contact and therefore prevent crystal growth in one axis of the lattice. This would explain the 2D disk-like shape of the spherulites. X-Ray diffraction of spherulites commonly yielded no discernable spots (Figure 3.9F), however some spherulites provided diffraction patterns akin to a small molecule or salt crystal based on the distantly interspersed spots indicating a well-packed, small unit cell (Figure 3.9C). These spherulites may have been an anomaly, as other spherulites have proven to be proteinaceous based on an SDS PAGE (Figure 3.6) but did not produce a diffraction pattern (Figure 3.9F).

The formation of proteinaceous spherulites in conditions that were deemed as favourable by thermofluor assays is promising for the technique. The formation of crystalline objects in the presence of sulfates and the magnesium cation, also shown to be stabilizing by thermofluor assays, may indicate their presence in forming important ionic intramolecular interactions for protein function or intermolecular interactions for crystal contacts (Sections 3.4 and 3.5). As well the increased T_m of conditions where buffer pH was approximate to the respective predicted pI of the proteins is interesting, as crystal trials are often suggested to be started with the pH of the mother liquor equal or proximal to the pI of the protein (Kantardjieff and Rupp, 2004; Kirkwood et al., 2015). Maintaining protein charge at net neutral is predicted to aid in the formation of intermolecular interactions, which is why increased stability and crystallisation propensity is seen with these conditions.

Results from global HDX MS show that the presence of the highly dynamic region in TraG* causes increased deuterium uptake, and therefore is highly solvated relative to TraG* $_{\Delta}$ (Section 3.7). However, more deuterium uptake was observed in the HisTraG* $_{\Delta}$ relative to that of HisTraG*. This is counterintuitive based on the larger size of TraG* and contrary the hypothesis

that the truncation mutant is more stable. There was also evidence that some unfolded species were present in the native mass spectra of TraG* Δ which could be the cause of the increased overall deuterium uptake of the fully deuterated TraG* Δ sample (Figure 3.13 B). The observed unfolded species could be a result of a fraction of the protein denaturing during the final desalting and buffer exchange; perhaps it does not remain stable in NH₄CH₃COO. TraG* may be stabilized by the acetate ion as indicated in thermofluor and MS experiments, while the truncation mutant does not appear to share this characteristic. The overall shape of the deuterium exchange rate curves for both proteins look similar however, indicating the loss of the N-terminal region does not drastically change the conformational states achievable by the protein, or the rate at which these folding states change.

The discussed N-terminal region of the protein is implicated in connecting the membrane-bound portion of TraG and periplasmic TraG*; it would be congruent with known moieties in structural biology that this region would serve as a highly dynamic linker region between these two domains that display different functionalities. It is logical to assume this region of the protein requires a high degree of flexibility to mediate the variety of interactions TraG* performs. In related systems, these two domains are commonly separate proteins as discussed in Section 1.6.2. Therefore, it is possible this region is merely a linker and does not contain interacting domains, making it non-essential in the protein's structural solution. *In vivo* experiments such as conjugative mating assays could not be performed to support this theory as TraG* is not functional when expressed on its own (Audette et al., 2007).

4.2 Future Work

This study suggests the presence of a highly dynamic region in residues A452–A496 of TraG_F, however the boundaries for the size of the region have not been fully identified as only one

deletion mutant was tested. To further the analysis, different mutants could be conceived with additions or deletions of the current TraG* Δ sequence to define the region that serves as a putative linker. The same experiments described in this report could be performed, however to further support the identification of a dynamic region, protein phase separation experiments should be considered as this region was predicted to be a phase separating IDR (Section 3.1.1)(Nott et al., 2015).

Alternatively, solving the structure of TraG* Δ will dictate whether the dynamic region has been completely or partially removed as determined through analysing the density at the N-terminus. Crystallisation of TraG* Δ is vital in solving the structure through X-Ray diffraction to provide an understanding of the protein's hypothesized function and indicate whether the N-terminus is an essential component of the periplasmic domain. Therefore, further crystal trials of HisTraG* Δ will be performed. As the protein is predicted to be a novel structure as there currently exists no solved structure of high homology, molecular replacement will not be possible (Dimaio et al., 2011). Multiple isomorphous replacement (MIR) and single wavelength anomalous diffraction (SAD) phasing must be performed through either heavy atom soaking of a TraG* Δ crystal or the expression of a selenomethionine derivative protein and subsequent crystallisation and diffraction (Hendrickson et al., 1990; Leahy et al., 1994).

Some unexpected results were seen throughout this report and must be confirmed as either anomalies or as results bearing a biochemical or physical basis. The lack of expression seen in TraG* Δ His was an interesting result and requires further investigation (Section 3.3). It will be confirmed that additional residues must be added to the N-terminus through the utilization of different C-terminal tags to the same protein. Constructs of N-terminally GST-tagged variant of TraG* and TraG* Δ have been made; a C-terminal tagged variant of both should be made and

expressed. If the same phenomenon occurs, it will confirm the inability for the protein to properly fold if no region precedes T497 of TraG*. As well, cleavage of the N-terminal His₆ tag from HisTraG*_Δ is an additional experiment to test the requirement of additional residues prior to T497. However, there will be an additional 17 residues from the remaining pET28a tag prior to the start of the TraG*_Δ sequence, which may alter the result. Alternatively, a plasmid containing a C-terminally His₆-tagged truncation mutant from the highly homologous TraG*_{R100} could be constructed and its expression could be tested in the same manner.

Other unexpected results included some of the non-congruent data in the thermofluor assays, such as TraG* being destabilized by MES pH 6.0 (Section 3.4). Further thermofluor studies must be performed with different buffer and salt concentrations, pHs, and entirely new buffer and salt combinations in order to properly assess solvent interactions with TraG* and protein variants. CD deconvolution by BeStSel did not provide logical secondary structure predictions for spectra provided from urea denaturation experiments (Section 3.6). This was likely due to abnormalities in the spectra from 207-190nm from the UV absorption of urea, however these are vital regions for structural information to be gleaned from CD. A different method should be performed such as heat denaturation or use of a different chemical denaturant such as guanidine hydrochloride (Matsuo et al., 2007; Rashid et al., 2005).

Discrepancies between the predicted size of HisTraG* and its observed size by ESI MS were theorized to be caused by a truncation of the 17 aa tag during ionization (Section 3.7). To confirm this, purified HisTraG* will have its His₆ tag removed using thrombin cleavage, untagged TraG* will be purified by collecting it in the flow through of a Ni²⁺-IMAC purification, the size of tagged and untagged compared by SDS PAGE, and the sample will be analysed through ESI MS. Global HDX MS results effectuated speculation of differences in TraG* mutant stability when

$\text{NH}_4\text{CH}_3\text{COO}$ is used as a buffer. Further replicates of the native MS and global HDX MS will be performed with TraG* Δ to confirm the presence of an unfolded population that may cause an increase in the overall deuterium uptake of this protein. As well, local HDX MS with time resolved ESI (TRESI) can be used to deuterate the protein, quench with acid at various time point, and then use trypsin digestion to generate peptide fragments (Konermann et al., 2011; Oganessian et al., 2018). This will allow for a refined analysis of the residue-by-residue solvent exposure of the protein and will allow for an accurate definition of the N-terminal dynamic region, as well as answering other questions on the protein's dynamics. The global HDX MS experiments showed the rate of deuterium uptake is similar between TraG* and TraG* Δ , indicating similarities in their achievable folds. Using small angle X-ray scattering will confirm changes in accessible conformational modes with the N-terminus deleted, as common folding states can be detected as low-resolution electron density models (Wen et al., 2014). These experiments would aid in assessing the viability of further experiments with TraG* Δ , as maintenance of the native fold is essential in assuring structural solution of this mutant produces an accurate depiction of the region as it would be when attached to full length TraG is in the T4SS complex.

To properly identify the role TraG* plays in the F pilus, its predicted interacting partners must be confirmed to interact. Pulldowns with GST-tagged TraS and TraN peptide fragments on the stationary phase could be used to identify binding of TraG*, as passing the protein through the column would identify which regions have strong affinity for TraG*. Confirmation of these protein interactions would be vital and would allow for the elucidation of regions responsible for performing the complex interaction within the F pilus superstructure. As well, if TraG* Δ abolishes or exhibits lowered affinity for the observed TraG* interactions it will provide an indication for whether the putative linker performs any important interactions or is merely a bridge between the

two domains of the protein. The structural solution resulting from diffraction of TraG* Δ co-crystals with TraS peptides could be achieved to better support the conclusions made in these studies.

The goal in studying the F-T4SS is to develop a universal functional model of the conjugative system such that novel antibiotics could be developed that target all or most pathogenic gram-negative bacterial species, as the F pilus is designated as the canonical gram-negative conjugative secretion system. TraG* $_{R100}$ can be studied to see if its structure is similar to that of TraG* $_F$. If the same linker is present then it is likely that the entry exclusion system of all F-like pili with homologous proteins function in a similar manner and allow for plasmid specificity as discussed in previous work (Audette et al., 2007). It is vital to determine whether all F-like pili function similarly in order to develop antibiotics that will target a wide array of gram-negative MDR pathogens.

TraG and TraS form the Eex system of the F pilus and are heavily relied upon for preventing donor-donor plasmid exchange (Achtman et al., 1980, 1977). As excessive conjugation can lead to lethal zygotis it has been theorized that disruption of Eex systems would be detrimental to bacterial colony survival (Garcillán-Barcia and de la Cruz, 2008). Development of novel antibiotics to disrupt TraG–TraS interactions and cause ceaseless conjugation or completely prohibit conjugation is contingent on the structural knowledge of the system's protein subunits. The identification of a highly dynamic region within TraG advances its structural study and provides insight for further experimentation in the objective of a high-resolution structural solution.

CHAPTER 5 – REFERENCES

- Achtman, M., Kennedy, N., Skurray, R., 1977. Cell-cell interactions in conjugating *Escherichia coli*: role of traT protein in surface exclusion. *Proc. Natl. Acad. Sci.* 74, 5104–5108. <https://doi.org/10.1073/pnas.74.11.5104>
- Achtman, M., Manning, P.A., Kusecek, B., Schwuchow, S., Neil, W., 1980. Genetic analysis of F sex factor cistrons needed for surface exclusion in *Escherichia coli*. *J. Mol. Biol.* 138, 779–795. [https://doi.org/10.1016/0022-2836\(80\)90065-0](https://doi.org/10.1016/0022-2836(80)90065-0)
- Allen, H.K., Donato, J., Wang, H.H., Cloud-Hansen, K.A., Davies, J., Handelsman, J., 2010. Call of the wild: antibiotic resistance genes in natural environments. *Nat. Rev. Microbiol.* 8, 251–9. <https://doi.org/10.1038/nrmicro2312>
- Arutyunov, D., Arenson, B., Manchak, J., Frost, L.S., 2010. F plasmid TraF and TraH are components of an outer membrane complex involved in conjugation. *J. Bacteriol.* 192, 1730–1734. <https://doi.org/10.1128/JB.00726-09>
- Arutyunov, D., Frost, L.S., 2013. F conjugation: Back to the beginning. *Plasmid* 70, 18–32. <https://doi.org/10.1016/j.plasmid.2013.03.010>
- 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. <https://doi.org/10.1099/mic.0.2006/001917-0>
- Bailey, S., Ward, D., Middleton, R., Grossmann, J.G., Zambryski, P.C., 2006. *Agrobacterium tumefaciens* VirB8 structure reveals potential protein-protein interaction sites. *Proc. Natl. Acad. Sci.* 103, 2582–2587. <https://doi.org/10.1073/pnas.0511216103>
- Balasubramaniam, D., Komives, E.A., 2013. Hydrogen-exchange mass spectrometry for the study of intrinsic disorder in proteins. *Biochim. Biophys. Acta - Proteins Proteomics* 1834, 1202–1209. <https://doi.org/10.1016/j.bbapap.2012.10.009>
- Barker, C.S., Meshcheryakova, I. V., Kostyukova, A.S., Freddolino, P.L., Samatey, F.A., 2017. An intrinsically disordered linker controlling the formation and the stability of the bacterial flagellar hook. *BMC Biol.* 15, 1–14. <https://doi.org/10.1186/s12915-017-0438-7>
- Baron, C., 2013. A novel strategy to target bacterial virulence. *Future Microbiol.* 8, 1–3. <https://doi.org/10.2217/fmb.12.120>
- Baron, C., 2010. Antivirulence drugs to target bacterial secretion systems. *Curr. Opin. Microbiol.* 13, 100–105. <https://doi.org/10.1016/j.mib.2009.12.003>
- Bayburt, T.H., Sligar, S.G., 2010. Membrane protein assembly into Nanodiscs. *FEBS Lett.* 584, 1721–7. <https://doi.org/10.1016/j.febslet.2009.10.024>
- Bennett, P.M., 2008. Plasmid encoded antibiotic resistance: Acquisition and transfer of antibiotic resistance genes in bacteria. *Br. J. Pharmacol.* 153, 347–357. <https://doi.org/10.1038/sj.bjp.0707607>
- Beveridge, R., Chappuis, Q., Macphee, C., Barran, P., 2013. Mass spectrometry methods for

- intrinsically disordered proteins. *Analyst* 138, 32–42. <https://doi.org/10.1039/c2an35665a>
- Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L., Tompa, P., Fuxreiter, M., 2018. Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* 28, 420–435. <https://doi.org/10.1016/j.tcb.2018.02.004>
- Bradley, D.E., 1980. Morphological and serological relationships of conjugative pili. *Top. Catal.* 4, 155–169. [https://doi.org/10.1016/0147-619X\(80\)90005-0](https://doi.org/10.1016/0147-619X(80)90005-0)
- Cabezón, E., de la Cruz, F., Arechaga, I., 2017. Conjugation inhibitors and their potential use to prevent dissemination of antibiotic resistance genes in bacteria. *Front. Microbiol.* 8, 1–7. <https://doi.org/10.3389/fmicb.2017.02329>
- Carattoli, A., Bertini, A., Villa, L., Falbo, V., Hopkins, K.L., Threlfall, E.J., 2005. Identification of plasmids by PCR-based replicon typing. *J. Microbiol. Methods* 63, 219–228. <https://doi.org/10.1016/j.mimet.2005.03.018>
- Casu, B., Mary, C., Sverzhinsky, A., Fouillen, A., Nanci, A., Baron, C., 2018. VirB8 homolog TraE from plasmid pKM101 forms a hexameric ring structure and interacts with the VirB6 homolog TraD. *Proc. Natl. Acad. Sci. U. S. A.* 115, 5950–5955. <https://doi.org/10.1073/pnas.1802501115>
- Chang, Q., Wang, W., Regev-Yochay, G., Lipsitch, M., Hanage, W.P., 2015. Antibiotics in agriculture and the risk to human health: How worried should we be? *Evol. Appl.* 8, 240–247. <https://doi.org/10.1111/eva.12185>
- Cheng, J., Randall, A.Z., Sweredoski, M.J., Baldi, P., 2005. SCRATCH: A protein structure and structural feature prediction server. *Nucleic Acids Res.* 33, 72–76. <https://doi.org/10.1093/nar/gki396>
- Chhetri, G., Kalita, P., Tripathi, T., 2015. An efficient protocol to enhance recombinant protein expression using ethanol in *Escherichia coli*. *MethodsX* 2, 385–391. <https://doi.org/10.1016/j.mex.2015.09.005>
- Christie, P.J., 2016. The Mosaic Type IV Secretion Systems. *EcoSal Plus* 7, 1–34. <https://doi.org/10.1128/ecosalplus.ESP-0020-2015.The>
- Christie, P.J., 2004. Type IV secretion: The *Agrobacterium* VirB/D4 and related conjugation systems. *Biochim. Biophys. Acta - Mol. Cell Res.* 1694, 219–234. <https://doi.org/10.1016/j.bbamcr.2004.02.013>
- Costa, T.R.D., Felisberto-Rodrigues, C., Meir, A., Prevost, M.S., Redzej, A., Trokter, M., Waksman, G., 2015. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat. Rev. Microbiol.* 13, 343–359. <https://doi.org/10.1038/nrmicro3456>
- Costa, T.R.D., Ilangovan, A., Ukleja, M., Redzej, A., Santini, J.M., Smith, T.K., Egelman, E.H., Waksman, G., 2016. Structure of the Bacterial Sex F Pilus Reveals an Assembly of a Stoichiometric Protein-Phospholipid Complex. *Cell* 166, 1436–1444.e10. <https://doi.org/10.1016/j.cell.2016.08.025>

- Couturier, M., Bex, F., Bergquist, P.L., Maas, W.K., 1988. Identification and classification of bacterial plasmids. *Microbiol. Rev.* 52, 375–95.
- Dimaio, F., Terwilliger, T.C., Read, R.J., Wlodawer, A., Oberdorfer, G., Wagner, U., Valkov, E., Alon, A., Fass, D., Axelrod, H.L., Das, D., Vorobiev, S.M., Iwai, H., Pokkuluri, P.R., Baker, D., 2011. Improved molecular replacement by density- and energy-guided protein structure optimization. *Nature* 473, 540–543. <https://doi.org/10.1038/nature09964>
- Donnelly, D.P., Rawlins, C.M., DeHart, C.J., Fornelli, L., Schachner, L.F., Lin, Z., Lippens, J.L., Aluri, K.C., Sarin, R., Chen, B., Lantz, C., Jung, W., Johnson, K.R., Koller, A., Wolff, J.J., Campuzano, I.D.G., Auclair, J.R., Ivanov, A.R., Whitelegge, J.P., Paša-Tolić, L., Chamot-Rooke, J., Danis, P.O., Smith, L.M., Tsybin, Y.O., Loo, J.A., Ge, Y., Kelleher, N.L., Agar, J.N., 2019. Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat. Methods* 16, 587–594. <https://doi.org/10.1038/s41592-019-0457-0>
- Dosztányi, Z., Mészáros, B., Simon, I., 2009. ANCHOR: Web server for predicting protein binding regions in disordered proteins. *Bioinformatics* 25, 2745–2746. <https://doi.org/10.1093/bioinformatics/btp518>
- Dubnau, E., Maas, W.K., 1968. Inhibition of replication of an F'lac episome in Hfr cells of *Escherichia coli*. *J. Bacteriol.* 95, 531–539.
- Dumetz, A.C., Snellinger-O'Brien, A.M., Kaler, E.W., Lenhoff, A.M., 2007. Patterns of protein-protein interactions in salt solutions and implications for protein crystallization. *Protein Sci.* 16, 1867–1877. <https://doi.org/10.1110/ps.072957907>
- Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M., Obradović, Z., 2002. Intrinsic disorder and protein function. *Biochemistry* 41, 6573–82. <https://doi.org/10.1021/bi012159+>
- Dyson, H.J., Wright, P.E., 2005. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197–208. <https://doi.org/10.1038/nrm1589>
- Edwards, S., Meynell, G.G., 1968. General method for isolating de-repressed bacterial sex factors. *Nature* 219, 869–870. <https://doi.org/10.1038/219869a0>
- Efimov, A. V., Brazhnikov, E. V., 2003. Relationship between intramolecular hydrogen bonding and solvent accessibility of side-chain donors and acceptors in proteins. *FEBS Lett.* 554, 389–393. [https://doi.org/10.1016/S0014-5793\(03\)01189-X](https://doi.org/10.1016/S0014-5793(03)01189-X)
- Ergodan, F., 2016. Structural and Functional Studies of TraG from the F Plasmid and the Major Pilin from *Coxiella Burnetii*. MSc. York University.
- Ericsson, U.B., Hallberg, B.M., DeTitta, G.T., Dekker, N., Nordlund, P., 2006. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* 357, 289–298. <https://doi.org/10.1016/j.ab.2006.07.027>
- Estrozi, L.F., Boehringer, D., Shan, S.O., Ban, N., Schaffitzel, C., 2011. Cryo-EM structure of the *E. coli* translating ribosome in complex with SRP and its receptor. *Nat. Struct. Mol. Biol.* 18, 88–90. <https://doi.org/10.1038/nsmb.1952>
- Exner, M., Bhattacharya, S., Christiansen, B., Gebel, J., Goroncy-Bermes, P., Hartemann, P.,

- Heeg, P., Ilschner, C., Kramer, A., Larson, E., Merkens, W., Mielke, M., Oltmanns, P., Ross, B., Rotter, M., Schmithausen, R.M., Sonntag, H.-G., Trautmann, M., 2017. Antibiotic resistance: What is so special about multidrug-resistant Gram-negative bacteria? *GMS Hyg. Infect. Control* 12, 1–24. <https://doi.org/10.3205/dgkh000290>
- Fercher, C., Probst, I., Kohler, V., Goessweiner-Mohr, N., Arends, K., Grohmann, E., Zangger, K., Meyer, N.H., Keller, W., 2016. VirB8-like protein TraH is crucial for DNA transfer in *Enterococcus faecalis*. *Sci. Rep.* 6, 1–13. <https://doi.org/10.1038/srep24643>
- Fernandez-Lopez, R., de Toro, M., Moncalian, G., Garcillan-Barcia, M.P., de la Cruz, F., 2016. Comparative Genomics of the Conjugation Region of F-like Plasmids: Five Shades of F. *Front. Mol. Biosci.* 3. <https://doi.org/10.3389/fmolb.2016.00071>
- Firth, N., Ippen-ihler, K., Skurray, R.A., 1996. Structure and Function of the F Factor and Mechanism of Conjugation. *Escherichia coli Salmonella Cell. Mol. Biol. Second Ed.* 2377–2401.
- Firth, N., Skurray, R., 1992. Characterization of the F plasmid bifunctional conjugation gene, traG. *MGG Mol. Gen. Genet.* 232, 145–153. <https://doi.org/10.1007/BF00299147>
- Focia, P.J., Shepotinovskaya, I. V., Seidler, J.A., Freymann, D.M., 2004. Heterodimeric GTPase Core of the SRP Targeting Complex. *Science (80-)*. 303, 373–377. <https://doi.org/10.1126/science.1090827>
- Frost, Laura S, Ippen-ihler, K., Skurray, R.A., 1994. Analysis of the Sequence and Gene Products of the Transfer Region of the F Sex Factor. *Microbiol. Rev.* 5, 162–210.
- Frost, L S, Ippen-Ihler, K., Skurray, R.A., 1994. Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev.* 58, 162–210.
- Garcillán-Barcia, M.P., de la Cruz, F., 2008. Why is entry exclusion an essential feature of conjugative plasmids? *Plasmid* 60, 1–18. <https://doi.org/10.1016/j.plasmid.2008.03.002>
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handb. Protein Identif. Anal. Tools ExPASy Serv.* 571–607. <https://doi.org/10.1385/1592598900>
- Geetha, V., 1996. Distortions in protein helices. *Int. J. Biol. Macromol.* 19, 81–89. [https://doi.org/10.1016/0141-8130\(96\)01106-3](https://doi.org/10.1016/0141-8130(96)01106-3)
- Gillespie, J.J., Phan, I.Q.H., Scheib, H., Subramanian, S., Edwards, T.E., Lehman, S.S., Piitulainen, H., Sayeedur Rahman, M., Rennoll-Bankert, K.E., Staker, B.L., Taira, S., Stacy, R., Myler, P.J., Azad, A.F., Pulliainen, A.T., 2015. Structural Insight into How Bacteria Prevent Interference between Multiple Divergent Type IV Secretion Systems. *MBio* 6, 1–17. <https://doi.org/10.1128/mbio.01867-15>
- Giske, C.G., Monnet, D.L., Cars, O., Carmeli, Y., 2008. Clinical and economic impact of common multidrug-resistant gram-negative bacilli. *Antimicrob. Agents Chemother.* 52, 813–821. <https://doi.org/10.1128/AAC.01169-07>
- Glover, K., Mei, Y., Sinha, S.C., 2016. Identifying intrinsically disordered protein regions likely to undergo binding-induced helical transitions. *Biochim. Biophys. Acta - Proteins*

- Proteomics 1864, 1455–1463. <https://doi.org/10.1016/j.bbapap.2016.05.005>
- Gogarten, J.P., Townsend, J.P., 2005. Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* 3, 679–687. <https://doi.org/10.1038/nrmicro1204>
- Greenfield, N.J., 2007. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* 1, 2876–2890. <https://doi.org/10.1038/nprot.2006.202>
- Gregersen, T., 1978. Method for the distinction of gramnegative from grampositive bacteria. *Eur. J. Appl. Microbiol.* 5, 123–127. <https://doi.org/10.1007/BF01385437>
- Griko, Y., Sreerama, N., Osumi-Davis, P., Woody, R.W., Woody, A.Y., 2001. Thermal and urea-induced unfolding in T7 RNA polymerase: calorimetry, circular dichroism and fluorescence study. *Protein Sci.* 10, 845–53. <https://doi.org/10.1110/ps.39701>
- Grohmann, E., Muth, G., Espinosa, M., 2003. Conjugative plasmid transfer in gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* 67, 277–301, table of contents. <https://doi.org/10.1128/membr.67.2.277-301.2003>
- Gross, J.D., Moerke, N.J., von der Haar, T., Lugovskoy, A.A., Sachs, A.B., Mccarthy, J.E.G., Wagner, G., 2003. Ribosome Loading onto the mRNA Cap Is Driven by Conformational Coupling between eIF4G and eIF4E University of California at Berkeley. *Cell* 115, 739–750.
- 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 Res.* 42, 5715–5727. <https://doi.org/10.1093/nar/gku194>
- Gyles, C., Boerlin, P., 2014. Horizontally Transferred Genetic Elements and Their Role in Pathogenesis of Bacterial Disease. *Vet. Pathol.* 51, 328–340. <https://doi.org/10.1177/0300985813511131>
- Haugh, E.F., Hirschfelder, J.O., 1955. Pi-electron forces between conjugated double bond molecules. *J. Chem. Phys.* 23, 1778–1796. <https://doi.org/10.1063/1.1740579>
- Hendrickson, W.A., Horton, J.R., LeMaster, D.M., 1990. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* 9, 1665–1672. <https://doi.org/10.1002/j.1460-2075.1990.tb08287.x>
- Houde, D., Arndt, J., Domeier, W., Berkowitz, S., Engen, J.R., 2009. Characterization of IgG1 Conformation and conformational dynamics by hydrogen/deuterium exchange mass spectrometry. *Anal. Chem.* 81, 2644–2651. <https://doi.org/10.1021/ac802575y>
- Hu, B., Khara, P., Christie, P.J., 2019. Structural bases for F plasmid conjugation and F pilus biogenesis in *Escherichia coli*. *Proc. Natl. Acad. Sci.* 201904428. <https://doi.org/10.1073/pnas.1904428116>
- Jolley, K.A., Maiden, M.C.J., 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11, 595. <https://doi.org/10.1186/1471-2105-11-595>

- Kaltashov, I.A., Bobst, C.E., Abzalimov, R.R., 2013. Mass spectrometry-based methods to study protein architecture and dynamics. *Protein Sci.* 22, 530–544.
<https://doi.org/10.1002/pro.2238>
- Kantardjieff, K.A., Rupp, B., 2004. Protein isoelectric point as a predictor for increased crystallization screening efficiency. *Bioinformatics* 20, 2162–2168.
<https://doi.org/10.1093/bioinformatics/bth066>
- Katta, V., Chait, B.T., 1991. Conformational changes in proteins probed by hydrogen-exchange. *Rapid Commun Mass Spectrom* 5, 214–217.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858.
<https://doi.org/10.1038/nprot.2015.053>
- Kelly, S.M., Jess, T.J., Price, N.C., 2005. How to study proteins by circular dichroism. *Biochim. Biophys. Acta - Proteins Proteomics* 1751, 119–139.
<https://doi.org/10.1016/j.bbapap.2005.06.005>
- Kerr, J.E., Christie, P.J., 2010. Evidence for VirB4-mediated dislocation of membrane-integrated VirB2 pilin during biogenesis of the *Agrobacterium* VirB/VirD4 type IV secretion system. *J. Bacteriol.* 192, 4923–4934. <https://doi.org/10.1128/JB.00557-10>
- Kertesz, M., 2019. Pancake Bonding: An Unusual Pi-Stacking Interaction. *Chem. - A Eur. J.* 25, 400–416. <https://doi.org/10.1002/chem.201802385>
- Kim, S., Huang, J., Lee, Y., Dutta, S., Yoo, H.Y., Jung, Y.M., Jho, Y., Zeng, H., Hwang, D.S., 2016. Complexation and coacervation of like-charged polyelectrolytes inspired by mussels. *Proc. Natl. Acad. Sci.* 113, E847–E853. <https://doi.org/10.1073/pnas.1521521113>
- Kirkwood, J., Hargreaves, D., O’Keefe, S., Wilson, J., 2015. Using isoelectric point to determine the pH for initial protein crystallization trials. *Bioinformatics* 31, 1444–1451.
<https://doi.org/10.1093/bioinformatics/btv011>
- Klimke, W.A., Frost, L.S., 1998. Genetic analysis of the role of the transfer gene, traN, of the F and R100-1 plasmids in mating pair stabilization during conjugation. *J. Bacteriol.* 180, 4036–4043.
- 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, 3527–3540. <https://doi.org/10.1099/mic.0.28025-0>
- Kohler, V., Keller, W., Grohmann, E., 2019. Regulation of Gram-Positive Conjugation. *Front. Microbiol.* 10. <https://doi.org/10.3389/fmicb.2019.01134>
- Konermann, L., Pan, J., Liu, Y.H., 2011. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* 40, 1224–1234.
<https://doi.org/10.1039/c0cs00113a>
- Koraimann, G., Teferle, K., Markolin, G., Woger, W., Högenauer, G., 1996. The FinOP repressor system of plasmid R1: analysis of the antisense RNA control of traJ expression

- and conjugative DNA transfer. *Mol. Microbiol.* 21, 811–21.
- Koraimann, G., Wagner, M.A., 2014. Social behavior and decision making in bacterial conjugation. *Front. Cell. Infect. Microbiol.* 4, 1–7. <https://doi.org/10.3389/fcimb.2014.00054>
- Lavinder, J.J., Hari, S.B., Sullivan, B.J., Magliery, T.J., 2009. High-throughput thermal scanning: A general, rapid dye-binding thermal shift screen for protein engineering. *J. Am. Chem. Soc.* 131, 3794–3795. <https://doi.org/10.1021/ja8049063>
- Lawley, T.D., Klimke, W.A., Gubbins, M.J., Frost, L.S., 2003. F factor conjugation is a true type IV secretion system. *FEMS Microbiol. Lett.* 224, 1–15. [https://doi.org/10.1016/S0378-1097\(03\)00430-0](https://doi.org/10.1016/S0378-1097(03)00430-0)
- Leahy, D.J., Erickson, H.P., Aukhil, I., Joshi, P., Hendrickson, W.A., 1994. Crystallization of a fragment of human fibronectin: Introduction of methionine by site-directed mutagenesis to allow phasing via selenomethionine. *Proteins Struct. Funct. Bioinforma.* 19, 48–54. <https://doi.org/10.1002/prot.340190107>
- Lederberg, J., Cavalli, L.L., Lederberg, R.M., 1952. Sex compatibility in *Escherichia coli*. *Genetics* 720–730.
- 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 Lett.* 590, 376–386. <https://doi.org/10.1002/1873-3468.12066>
- Lerminiaux, N.A., Cameron, A.D.S., 2018. Horizontal transfer of antibiotic resistance genes in clinical environments. *Can. J. Microbiol.* 65, 34–44. <https://doi.org/10.1139/cjm-2018-0275>
- Li, P., Banjade, S., Cheng, H.C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J. V., King, D.S., Banani, S.F., Russo, P.S., Jiang, Q.X., Nixon, B.T., Rosen, M.K., 2012. Phase transitions in the assembly of multivalent signalling proteins. *Nature* 483, 336–340. <https://doi.org/10.1038/nature10879>
- Lin, Y.H., Forman-Kay, J.D., Chan, H.S., 2016. Sequence-Specific Polyampholyte Phase Separation in Membraneless Organelles. *Phys. Rev. Lett.* 117, 1–6. <https://doi.org/10.1103/PhysRevLett.117.178101>
- Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., Russell, R.B., 2003. Protein disorder prediction: Implications for structural proteomics. *Structure* 11, 1453–1459. <https://doi.org/10.1016/j.str.2003.10.002>
- Lu, J., Wong, J.J.W., Edwards, R.A., Manchak, J., Frost, L.S., Glover, J.N.M., 2008. Structural basis of specific TraD-TraM recognition during F plasmid-mediated bacterial conjugation. *Mol. Microbiol.* 70, 89–99. <https://doi.org/10.1111/j.1365-2958.2008.06391.x>
- 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. *Proc. Natl. Acad. Sci.* 104, 12282–12287. <https://doi.org/10.1073/pnas.0702760104>
- Magnan, C.N., Baldi, P., 2014. SSpro/ACCpro 5: Almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural

- similarity. *Bioinformatics* 30, 2592–2597. <https://doi.org/10.1093/bioinformatics/btu352>
- Manwaring, N.P., Skurray, R.A., Firth, N., 1999. Nucleotide sequence of the F plasmid leading region. *Plasmid* 41, 219–225. <https://doi.org/10.1006/plas.1999.1390>
- Martinez-Yamout, M., Evans, R.M., Dyson, H.J., Wright, P.E., Chung, J., Demarest, S.J., Xu, W., Chen, H., 2002. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* 415, 549–553. <https://doi.org/10.1038/415549a>
- Matson, S.W., Ragonese, H., 2005. The F-plasmid Tral protein contains three functional domains required for conjugative DNA strand transfer. *J. Bacteriol.* 187, 697–706. <https://doi.org/10.1128/JB.187.2.697-706.2005>
- Matsuo, K., Sakurada, Y., Yonehara, R., Kataoka, M., Gekko, K., 2007. Secondary-structure analysis of denatured proteins by vacuum-ultraviolet circular dichroism spectroscopy. *Biophys. J.* 92, 4088–4096. <https://doi.org/10.1529/biophysj.106.103515>
- McGaughey, G.B., Gagné, M., Rappé, A.K., 1998. π -Stacking Interactions. *J. Biol. Chem.* 273, 15458–15463. <https://doi.org/10.1074/jbc.273.25.15458>
- Mészáros, B., Erdős, G., Dosztányi, Z., 2018. IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* 46, W329–W337. <https://doi.org/10.1093/nar/gky384>
- Meynell, E., Ewins, A., 1973. Effect on exclusion of alterations to the sex pilus. *J. Bacteriol.* 113, 71–75.
- Micsonai, A., Wien, F., Kernya, L., Lee, Y.-H., Goto, Y., Réfrégiers, M., Kardos, J., 2015. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Natl. Acad. Sci.* 112, E3095–E3103. <https://doi.org/10.1073/pnas.1500851112>
- Mitreá, D.M., Kriwacki, R.W., 2016. Phase separation in biology; Functional organization of a higher order Short linear motifs - The unexplored frontier of the eukaryotic proteome. *Cell Commun. Signal.* 14, 1–20. <https://doi.org/10.1186/s12964-015-0125-7>
- Mittag, T., Kay, L.E., Forman-Kaya, J.D., 2010. Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recognit.* 23, 105–116. <https://doi.org/10.1002/jmr.961>
- Mizianty, M.J., Peng, Z., Kurgan, L., 2013. MFDp2. *Intrinsically Disord. Proteins* 1, e24428. <https://doi.org/10.4161/idp.24428>
- Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D., Baldwin, A.J., 2015. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell* 57, 936–947. <https://doi.org/10.1016/j.molcel.2015.01.013>
- Novick, R.P., 1987. Plasmid incompatibility. *Microbiol. Rev.* 51, 381–95.
- Novick, R.P., 1969. Extrachromosomal inheritance in bacteria. *Bacteriol. Rev.* 33, 210–63.
- Oganesyan, I., Lento, C., Wilson, D.J., 2018. Contemporary hydrogen deuterium exchange mass spectrometry. *Methods* 144, 27–42. <https://doi.org/10.1016/j.ymeth.2018.04.023>

- Ou, J.T., 1980. Role of surface exclusion genes in lethal zygosis in *Escherichia coli* K12 mating. *MGG Mol. Gen. Genet.* 178, 573–581. <https://doi.org/10.1007/BF00337863>
- Pelat, C., Kardaś-Słoma, L., Birgand, G., Ruppé, E., Schwarzingler, M., Andremont, A., Lucet, J.C., Yazdanpanah, Y., 2015. Hand hygiene, cohorting, or antibiotic restriction to control outbreaks of multidrug-resistant enterobacteriaceae. *Infect. Control Hosp. Epidemiol.* 37, 272–280. <https://doi.org/10.1017/ice.2015.284>
- Peng, Y., Lu, J., Wong, J.J.W., Edwards, R.A., Frost, L.S., Glover, J.N.M., 2014. Mechanistic basis of plasmid-specific DNA binding of the F plasmid regulatory protein, TraM. *J. Mol. Biol.* 426, 3783–3795. <https://doi.org/10.1016/j.jmb.2014.09.018>
- Phillips, K., de la Peña, A.H., 2011. The combined use of the ThermoFluor assay and ThermoQ analytical software for the determination of protein stability and buffer optimization as an aid in protein crystallization. *Curr. Protoc. Mol. Biol.* 2011, 1–15. <https://doi.org/10.1002/0471142727.mb1028s94>
- Provencher, S.W., Glöckner, J., 1981. Estimation of Globular Protein Secondary Structure from Circular Dichroism. *Biochemistry* 20, 33–37. <https://doi.org/10.1021/bi00504a006>
- Rashid, F., Sharma, S., Bano, B., 2005. Comparison of guanidine hydrochloride (GdnHCl) and urea denaturation on inactivation and unfolding of human placental cystatin (HPC). *Protein J.* 24, 283–292. <https://doi.org/10.1007/s10930-005-6749-5>
- Reddy Chichili, V.P., Kumar, V., Sivaraman, J., 2013. Linkers in the structural biology of protein-protein interactions. *Protein Sci.* 22, 153–167. <https://doi.org/10.1002/pro.2206>
- Resetca, D., Wilson, D.J., 2013. Characterizing rapid, activity-linked conformational transitions in proteins via sub-second hydrogen deuterium exchange mass spectrometry. *FEBS J.* 280, 5616–5625. <https://doi.org/10.1111/febs.12332>
- Ribeiro, E.A., Ramos, C.H.I., 2005. Circular permutation and deletion studies of myoglobin indicate that the correct position of its N-terminus is required for native stability and solubility but not for native-like heme binding and folding. *Biochemistry* 44, 4699–4709. <https://doi.org/10.1021/bi047908c>
- Riley, K., Tran, K.-A., 2017. Strength and Character of R–X··· π Interactions Involving Aromatic Amino Acid Sidechains in Protein-Ligand Complexes Derived from Crystal Structures in the Protein Data Bank. *Crystals* 7, 273. <https://doi.org/10.3390/cryst7090273>
- Roberts, A.P., Chandler, M., Courvalin, P., Guédon, G., Mullany, P., Pembroke, T., Rood, J.I., Jeffery Smith, C., Summers, A.O., Tsuda, M., Berg, D.E., 2008. Revised nomenclature for transposable genetic elements. *Plasmid* 60, 167–173. <https://doi.org/10.1016/j.plasmid.2008.08.001>
- Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5, 725–738. <https://doi.org/10.1038/nprot.2010.5>
- Schep, D.G., Zhao, J., Rubinstein, J.L., 2016. Models for the α subunits of the *Thermus thermophilus* V/A-ATPase and *Saccharomyces cerevisiae* V-ATPase enzymes by cryo-EM and evolutionary covariance. *Proc. Natl. Acad. Sci.* 113, 3245–3250.

<https://doi.org/10.1073/pnas.1521990113>

- Schröder, G., Lanka, E., 2005. The mating pair formation system of conjugative plasmids - A versatile secretion machinery for transfer of proteins and DNA. *Plasmid* 54, 1–25. <https://doi.org/10.1016/j.plasmid.2005.02.001>
- Schrödinger, L., 2015. The PyMol Molecular Graphics System.
- Seabrook, S.A., Newman, J., 2013. High-throughput thermal scanning for protein stability: Making a good technique more robust. *ACS Comb. Sci.* 15, 387–392. <https://doi.org/10.1021/co400013v>
- Sfakianos, A.P., Whitmarsh, A.J., Ashe, M.P., 2016. Ribonucleoprotein bodies are phased in. *Biochem. Soc. Trans.* 44, 1411–1416. <https://doi.org/10.1042/bst20160117>
- Shala-Lawrence, A., Bragagnolo, N., Nowroozi-Dayeni, R., Kheyson, S., Audette, G.F., 2018. The interaction of TraW and TrbC is required to facilitate conjugation in F-like plasmids. *Biochem. Biophys. Res. Commun.* 503, 2386–2392. <https://doi.org/10.1016/j.bbrc.2018.06.166>
- Silhavy, T.J., Kahne, D., Walker, S., 2010. The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* 2, a000414. <https://doi.org/10.1101/cshperspect.a000414>
- Skurray, R.A., Reeves, P., 1974. F factor mediated immunity to lethal zygotism in *Escherichia coli* K 12. *J. Bacteriol.* 117, 100–106.
- Skurray, R.A., Reeves, P., 1973. Characterization of lethal zygotism associated with conjugation in *Escherichia coli* K-12. *J. Bacteriol.* 113, 58–70.
- Smith, M.A., Coinon, M., Paschos, A., Jolicoeur, B., Lavallée, P., Sygusch, J., Baron, C., 2012. Identification of the binding site of *Brucella* VirB8 interaction inhibitors. *Chem. Biol.* 19, 1041–1048. <https://doi.org/10.1016/j.chembiol.2012.07.007>
- Su, X., Ditlev, J.A., Hui, E., Xing, W., Banjade, S., Okrut, J., King, D.S., Taunton, J., Rosen, M.K., Vale, R.D., 2016. Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* 352, 595–9. <https://doi.org/10.1126/science.aad9964>
- Sultan, I., Rahman, S., Jan, A.T., Siddiqui, M.T., Mondal, A.H., Haq, Q.M.R., 2018. Antibiotics, resistome and resistance mechanisms: A bacterial perspective. *Front. Microbiol.* 9. <https://doi.org/10.3389/fmicb.2018.02066>
- Tatum, E.L., Lederberg, J., 1947. Gene Recombination in the Bacterium *Escherichia coli*. *J. Bacteriol.* 53, 673–84.
- Taylor, J.P., Brown, R.H., Cleveland, D.W., 2016. Decoding ALS: From genes to mechanism. *Nature* 539, 197–206. <https://doi.org/10.1038/nature20413>
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–80. <https://doi.org/10.1093/nar/22.22.4673>
- Thomson, J.M., Bonomo, R.A., 2005. The threat of antibiotic resistance in Gram-negative

- pathogenic bacteria: β -lactams in peril! *Curr. Opin. Microbiol.* 8, 518–524.
<https://doi.org/10.1016/j.mib.2005.08.014>
- Uversky, V.N., 2013. A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Sci.* 22, 693–724. <https://doi.org/10.1002/pro.2261>
- Uversky, V.N., 2002. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11, 739–56. <https://doi.org/10.1110/ps.4210102>
- Uversky, V.N., Oldfield, C.J., Dunker, A.K., 2008. Intrinsically Disordered Proteins in Human Diseases: Introducing the D 2 Concept . *Annu. Rev. Biophys.* 37, 215–246.
<https://doi.org/10.1146/annurev.biophys.37.032807.125924>
- Valvekens, D., Van Montagu, M., Van Lijsebettens, M., 1988. Agrobacterium tumefaciens-mediated transformation of Arabidopsis thaliana root explants by using kanamycin selection. *Proc. Natl. Acad. Sci. U. S. A.* 85, 5536–40.
<https://doi.org/10.1073/pnas.85.15.5536>
- Van Der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., Kim, P.M., Kriwacki, R.W., Oldfield, C.J., Pappu, R. V., Tompa, P., Uversky, V.N., Wright, P.E., Babu, M.M., 2014. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* 114, 6589–6631.
<https://doi.org/10.1021/cr400525m>
- Velappan, N., Sblattero, D., Chasteen, L., Pavlik, P., Bradbury, A.R.M., 2007. Plasmid incompatibility: More compatible than previously thought? *Protein Eng. Des. Sel.* 20, 309–313. <https://doi.org/10.1093/protein/gzm005>
- Vernon, R.M., Chong, P.A., Tsang, B., Kim, T.H., Bah, A., Farber, P., Lin, H., Forman-Kay, J.D., 2018. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife* 7, 1–48. <https://doi.org/10.7554/elife.31486>
- Viljanen, P., 1987. Polycations which disorganize the outer membrane inhibit conjugation in Escherichia coli. *J. Antibiot. (Tokyo).* 40, 882–6.
- Villa, L., García-Fernández, A., Fortini, D., Carattoli, A., 2010. Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. *J. Antimicrob. Chemother.* 65, 2518–2529. <https://doi.org/10.1093/jac/dkq347>
- Vogan, A.A., Higgs, P.G., 2011. The advantages and disadvantages of horizontal gene transfer and the emergence of the first species. *Biol. Direct* 6, 1. <https://doi.org/10.1186/1745-6150-6-1>
- Waksman, G., 2019. From conjugation to T4S systems in Gram-negative bacteria: a mechanistic biology perspective. *EMBO Rep.* 20, e47012. <https://doi.org/10.15252/embr.201847012>
- Wang, L., Pan, H., Smith, D.L., 2002. Hydrogen exchange-mass spectrometry: optimization of digestion conditions. *Mol. Cell. Proteomics* 1, 132–8.
<https://doi.org/10.1074/mcp.m100009-mcp200>
- Watanabe, T., 1963. Infective heredity of multiple drug resistance in bacteria. *Bacteriol. Rev.* 27, 87–115.

- Watanabe, T., Sakaizumi, S., Furuse, C., 1968. Superinfection with R factors by transduction in *Escherichia coli* and *Salmonella typhimurium*. *J. Bacteriol.* 96, 1796–1802.
- Wawrzyniak, P., Plucienniczak, G., Bartosik, D., 2017. The different faces of rolling-circle replication and its multifunctional initiator proteins. *Front. Microbiol.* 8, 1–13. <https://doi.org/10.3389/fmicb.2017.02353>
- Wen, B., Peng, J., Zuo, X., Gong, Q., Zhang, Z., 2014. Characterization of protein flexibility using small-angle x-ray scattering and amplified collective motion simulations. *Biophys. J.* 107, 956–964. <https://doi.org/10.1016/j.bpj.2014.07.005>
- Willetts, N., Achtman, M., 1972. Genetic analysis of transfer by the *Escherichia coli* sex factor F, using P1 transductional complementation. *J. Bacteriol.* 110, 843–851.
- Williams, J.J., Hergenrother, P.J., 2008. Exposing plasmids as the Achilles' heel of drug-resistant bacteria. *Curr. Opin. Chem. Biol.* 12, 389–399. <https://doi.org/10.1016/j.cbpa.2008.06.015>
- Wilm, M., 2011. Principles of electrospray ionization. *Mol. Cell. Proteomics* 10, 1–8. <https://doi.org/10.1074/mcp.M111.009407>
- Wilman, H.R., Shi, J., Deane, C.M., 2014. Helix kinks are equally prevalent in soluble and membrane proteins. *Proteins Struct. Funct. Bioinforma.* 82, 1960–1970. <https://doi.org/10.1002/prot.24550>
- Winkler, R., 2010. ESIprot: A universal tool for charge state determination and molecular weight calculation of proteins from electrospray ionization mass spectrometry data. *Rapid Commun. Mass Spectrom.* 24, 285–294. <https://doi.org/10.1002/rcm.4384>
- Wootton, J.C., Drummond, M.H., 1989. The Q-linker: a class of interdomain sequences found in bacterial multidomain regulatory proteins. *Protein Eng.* 2, 535–43. <https://doi.org/10.1093/protein/2.7.535>
- Xue, B., Dunker, A.K., Uversky, V.N., 2012. Orderly order in protein intrinsic disorder distribution: Disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.* 30, 137–149. <https://doi.org/10.1080/07391102.2012.675145>
- Yeo, G.C., Keeley, F.W., Weiss, A.S., 2011. Coacervation of tropoelastin. *Adv. Colloid Interface Sci.* 167, 94–103. <https://doi.org/10.1016/j.cis.2010.10.003>
- Zhang, Y., 2009. Protein structure prediction: when is it useful? *Curr. Opin. Struct. Biol.* 19, 145–155. <https://doi.org/10.1016/j.sbi.2009.02.005>
- Zhang, Z., Smith, D.L., 1993. Determination of amide hydrogen exchange by mass spectrometry: A new tool for protein structure elucidation. *Protein Sci.* 2, 522–531. <https://doi.org/10.1002/pro.5560020404>

APPENDICES

Appendix A – Relevant Amino Acid Sequences for Proteins used in this Thesis

TraG*His

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
AGSVVDGNYS	YGNMQTENVN	GFSWSTNSTT	SFGQMMYQTG	SGATATQTRD	GNMVMDASGA
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
MSRLPVGINA	TRQIAAAQQE	MAREASNRAE	SALHGFSSSI	ASAWNTLSQF	GSNRGSSDSV
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
TGGADSTMSA	QDSMMASRM	SAVESYAKAH	NISNEQATRE	LASRSTNASL	GLYGDAYAKG
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
HLGISVLGNG	GGVGLQAGAK	ASIDGSDLDS	HEASSGSRAS	HDARHDIDAR	ATQDFKEASD
<u>250</u>	<u>260</u>	<u>270</u>	<u>280</u>	<u>290</u>	<u>300</u>
YFTSRKVSES	GSHTDNNADS	RVDQLSAALN	SAKQSYDQYT	TNMTRSHEYA	EMASRTESMS
<u>310</u>	<u>320</u>	<u>330</u>	<u>340</u>	<u>350</u>	<u>360</u>
GQMSDLSQQ	FAQYVMKNAP	QDVEAILTNT	SSPEIAERRR	AMAWSFVQEQ	VQPGVDNTWR
<u>370</u>	<u>380</u>	<u>390</u>	<u>400</u>	<u>410</u>	<u>420</u>
ESRRDIGKGM	ESVPSGGGSQ	DIIADHQGHQ	AIIEQRTQDS	NIRNDVKHQV	DNMVTEYRGN
<u>430</u>	<u>440</u>	<u>450</u>	<u>460</u>	<u>470</u>	<u>480</u>
IGDTQNSIRG	EENIVKGQYS	ELQNHKTEA	LTQNNKYNEE	KLAQERIPGA	DSPKELLEKA
<u>490</u>	<u>500</u>				
KSYQHKEKLA	AALHHHHHH				

Number of amino acids: 500

Molecular weight: 54238.70 Da

Theoretical pI: 5.61

HisTraG*

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MGSSHHHHH	SSGLVPRGSH	MASMTGGQQM	GRGSAGSVVD	GNYSYGNMQT	ENVNGFSWST
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
NSTTSFGQMM	YQTGSGATAT	QTRDGNMVMD	ASGAMSRLPV	GINATRQIAA	AQQEMAREAS
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
NRAESALHGF	SSSIASAWNT	LSQFGSNRGS	SDSVTGGADS	TMSAQDSMMA	SRMRSAVESY
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
AKAHNISNEQ	ATRELASRST	NASLGLYGDA	YAKHGLGISV	LGNGGGVGLQ	AGAKASIDGS
<u>250</u>	<u>260</u>	<u>270</u>	<u>280</u>	<u>290</u>	<u>300</u>
DLDSHEASSG	SRASHDARHD	IDARATQDFK	EASDYFTSRK	VSESGSHTDN	NADSRVDQLS
<u>310</u>	<u>320</u>	<u>330</u>	<u>340</u>	<u>350</u>	<u>360</u>
AALNSAKQSY	DQYTTNMTRS	HEYAEMASRT	ESMSGQMSD	LSQQFAQYVM	KNAPQDVEAI
<u>370</u>	<u>380</u>	<u>390</u>	<u>400</u>	<u>410</u>	<u>420</u>
LTNTSSPEIA	ERRRAMAWSF	VQEQVQPGVD	NTWRESRRDI	GKGMESVPSG	GGSQDIIADH
<u>430</u>	<u>440</u>	<u>450</u>	<u>460</u>	<u>470</u>	<u>480</u>
QGHQAIIEQR	TQDSNIRNDV	KHQVDNMVTE	YRGNIGDTQN	SIRGEENIVK	GQYSELQNH
<u>490</u>	<u>500</u>	<u>510</u>	<u>520</u>		
KTEALTQNNK	YNEEKLAQER	IPGADSPKEL	LEKAKSYQHK	E	

Number of amino acids: 521

Molecular weight: 56262.90 Da

Theoretical pI: 5.77

TraG* Δ His

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
TQTRDGNMVM	DASGAMSRLP	VGINATRQIA	AAQQEMAREA	SNRAESALHG	FSSSIASAWN
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
TLSQFGSNRG	SSDSVTGGAD	STMSAQDSMM	ASRMRSAVES	YAKAHNISNE	QATRELASRS
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
TNASLGLYGD	AYAKGHLGIS	VLGNNGGVGL	QAGAKASIDG	SDLDSHEASS	GSRASHDARH
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
DIDARATQDF	KEASDYFTSR	KVSESGSHTD	NNADSRVDQL	SAALNSAKQS	YDQYTTNMTR
<u>250</u>	<u>260</u>	<u>270</u>	<u>280</u>	<u>290</u>	<u>300</u>
SHEYAEMASR	TESMSGQMSE	DLSQQFAQYV	MKNAPQDVEA	ILTNTSSPEI	AERRRAMAWS
<u>310</u>	<u>320</u>	<u>330</u>	<u>340</u>	<u>350</u>	<u>360</u>
FVQEQVQPGV	DNTWRESRRD	IGKGMESVPS	GGGSQDIIAD	HQGHQAIIEQ	RTQDSNIRND
<u>370</u>	<u>380</u>	<u>390</u>	<u>400</u>	<u>410</u>	<u>420</u>
VKHQVDNMVT	EYRGNIGDTQ	NSIRGEENIV	KGOYSELQNH	HKTEALTQNN	KYNEEKLAQE
<u>430</u>	<u>440</u>	<u>450</u>			
RIPGADSPKE	LLEKAKSYQH	KEKLAAALEH	HHHHH		

Number of amino acids: 455

Molecular weight: 49549.73 Da

Theoretical pI: 5.72

HisTraG* Δ

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MGSSHHHHHH	SSGLVPRGSH	MASMTGGQQM	GRGSTQTRDG	NMVMDASGAM	SRLPVGINAT
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
RQIAAAQOEM	AREASNRAES	ALHGFSSSIA	SAWNTLSQFG	SNRGSSDSVT	GGADSTMSAQ
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
DSMMASRMRS	AVESYAKAHN	ISNEQATREL	ASRSTNASLG	LYGDAYAKGH	LGISVLGNGG
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
GVGLQAGAKA	SIDGSDLDSH	EASSGSRASH	DARHDIDARA	TQDFKEASDY	FTSRKVSESG
<u>250</u>	<u>260</u>	<u>270</u>	<u>280</u>	<u>290</u>	<u>300</u>
SHTDNNADSR	VDQLSAALNS	AKQSYDQYTT	NMTRSHEYAE	MASRTEMSG	QMSEDLSQQF
<u>310</u>	<u>320</u>	<u>330</u>	<u>340</u>	<u>350</u>	<u>360</u>
AQYVMKNAPQ	DVEAILTNTS	SPEIAERRRA	MAWSFVQEQV	QPGVDNTWRE	SRRDIGKGME
<u>370</u>	<u>380</u>	<u>390</u>	<u>400</u>	<u>410</u>	<u>420</u>
SVPSGGGSQD	IIADHQGHQA	IIEQRTQDSN	IRNDVKHQVD	NMVTEYRGNI	GDTQNSIRGE
<u>430</u>	<u>440</u>	<u>450</u>	<u>460</u>	<u>470</u>	
ENIVKGQYSE	LQNHKTEAL	TQNNKYNEEK	LAQERIPGAD	SPKELLEKAK	SYQHKE

Number of amino acids: 476

Molecular weight: 51573.94 Da

Theoretical pI: 5.89

Reference: (Gasteiger et al., 2005)

Appendix B – Generation of Competent *E. coli* Cells

Overnight cultures of DH5 α cells were prepared in LB broth (37°C, >150rpm). The overnight culture was used to inoculate sterile LB broth in a 1:100 ratio. Incubation was performed at 37 °C, >150 rpm until O.D.₆₀₀ reached the beginning of the logarithmic growth phase (0.4-0.6). The cells were incubated on ice for 10 minutes before centrifugation was performed for 10 minutes at 2,700 x g, 4 °C. LB was decanted from the cell pellet and the pellet was re-suspended with 1.6 mL ice cold, filtered and autoclaved 100 mM CaCl₂. Cells were then incubated on ice for 30 minutes before being centrifuged for 10 minutes at 2,700 x g, 4 °C. Again, the supernatant was completely removed, and the pellet was re-suspended with 1.6mL 100 mM CaCl₂. The cells were incubated on ice for 20 minutes to equilibrate. 0.25 mL of ice-cold autoclaved 80% glycerol was slowly added to the 1.6mL of cells and mixed gently until homogeneous. The competent cells were then stored at -80 °C for up to 6 months. The same procedure was followed for the generation of competent BL21 (DE3) cells.