

Structural and Functional Insights into the F Plasmid Type IV Secretion System proteins TrbI and TrbB

Arnold Apostol

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

GRADUATE PROGRAM IN CHEMISTRY
YORK UNIVERSITY
TORONTO, ON

December 2023

© Arnold Apostol, 2023

ABSTRACT

Bacteria have evolved elaborate mechanisms to thrive in stressful environments. One mechanism that bacteria utilize are secretion systems that can traverse protective lipid cell membranes and serve as mediators for a diverse set of goals, including the secretion of toxins implicated with target host pathogenesis. F-like plasmids in gram-negative bacteria encode for the multi-protein Type IV Secretion System (T4SS_F) that is functional for bacterial proliferation and adaptation through the process of conjugation. The periplasmic protein TrbB is believed to have a stabilizing chaperone role in the T4SS_F assembly, with TrbB exhibiting disulfide isomerase (DI) activity. In the current report, we demonstrate that residues W57-K181, which include the active thioredoxin motif, are sufficient for DI activity. Moreover, a structural model of GST-TrbB_{WT} based on ColabFold-AlphaFold2 and Small Angle X-Ray Scattering data indicate that TrbB_{WT}'s N-terminus is disordered, and this disordered nature likely contributes to the protein's dynamicity and recalcitrance to crystallization. A truncation construct, TrbB₅₇₋₁₈₁, was designed and found to exhibit higher physicochemical stability using ¹H-¹⁵N Heteronuclear Single Quantum Correlation spectroscopy and Circular Dichroism spectroscopy. Binding studies of TrbB and other T4SS_F proteins TrbI and TraW were performed, and results do not support the inference of a stable complex forming *in vitro*. Comparative studies of TrbB, TraF, and TrbI also provide insights into the structure of these T4SS_F component proteins. Lastly, crystallization trials of GST-TrbB_{WT} and GST-TrbI provide leads for future crystallization campaigns.

For Marilyn and Arnulfo

ACKNOWLEDGEMENTS

This work would not be possible without my supervisor, Dr. Gerald F Audette. I thank him, first and foremost, for giving me an opportunity to do what I love. I started seeing science in a bad light in the late stages of my undergraduate training, but Gerald allowed me to rekindle my excitement by showing me that science can be done in a rigorous and goal-oriented environment that fosters kindness, humour, teamwork, and humility. It wasn't always "bright and sunny" days, but it was a great privilege to go to work every day, doing something which I happened to be passionate about... and I owe this to Gerald for giving me a chance. I thank my lab colleagues Nicholas Bragagnolo and Christina Rodriguez (mother) for their mentorship, friendship, numerous discussions, invaluable insights, laughter... and patience... I like to think of them as forming a holoenzyme, for they have all allowed me to grow as a learner, communicator, scientist, and individual. Unlike enzymes, they didn't supply me with protons (at least not in way that I was aware) but they sure had provided me a ton of support in and outside the lab.

To my LSB friends, Tanvir Chakkal and Alessia Libertucci, thank you. The *violations* we endured together, the chemicals we almost didn't dispose properly, the autoclave mishaps turned entertainment-for-the-day, the NanoPure water we spilled, cleaned up, and pretended didn't happen, the entertaining *teas we sipped*, and the people we perhaps bothered during our tenure at LSB will forever be tattooed on my skin. I appreciated your company and patience and all the other things you taught me that allowed me to grow. Some of the skills that enabled this thesis were initiated because of our friendship—like how I started using GraphPad Prism to make pretty figures and to use it for t-tests; like how I learned about an accessible way to model protein using ColabFold-AlphaFold2 and approach PyMOL as if it were a friend who has answers and not be intimidated by its GUI. To my other friends: Mark, Ivan, and Frangel—thank you for allowing me to take my mind off work, do other things, de-stress, be reminded that I am not a robot (darn it), and go back to the lab with a renewed vigor.

To Dr. Logan Donaldson, Ms. Cristina Lento, Dr. Derek J Wilson, and Dr. Philip Johnson—thank you for your insightful discussions and feedback. Many of the stories in this thesis were inspired by the seeds of ideas that you most certainly planted in my head by being my teacher, advisor, critic, or all of the above. I also most certainly want to express my deep gratitude to Logan for allowing me (as far as I know) to use many of his lab equipment and resources. I thank Cristina for assisting me with the BLI experiment and the analytical SEC experiments described in this thesis. I thank my professors for facilitating me to appreciate foundational scientific principles, soft skills, and allowing me to understand that science is a *way of knowing* rather than *knowing*, and that it is dynamic and therefore exciting: Dr. Katalin Hudak, Dr. Jan Sapp, Dr. Pierre Potvin, and Dr. Tamara Kelly. I also thank my students for providing me opportunities to reflect on, and in some ways re-learn, the concepts, principles, and skills that I had learned to be able to teach them.

To my parents, Marilyn and Arnulfo, I dedicate this hard work to you. Whenever I need to be reminded to be resilient, driven, and grounded, I draw from you. Thank you for believing in me, being proud of my accomplishments, offering counsel when I *trip and fall*, and supporting me every step of the way. To my siblings, Mark and Caryn, thank you for setting great examples for me to emulate.

TABLE OF CONTENTS

ABSTRACT	ii
DEDICATION	iii
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS	v
LIST OF ACRONYMS	viii
LIST OF FIGURES	x
LIST OF TABLES	xi
1. CHAPTER ONE: INTRODUCTION	1
1.1. Rapid genetic recombination is a driver of the rich phenotypic variability in bacteria	1
1.1.1. Bacterial diversity and adaptability	1
1.1.2. Bacteria have a genome subject to mutations	1
1.1.3. Horizontal gene transfer; the phenomena of transformation and conjugation	1
1.1.4. Transduction; the third mechanism of Horizontal Gene Transfer	3
1.1.5. Transposable elements; their role in Horizontal Gene Transfer and genomic plasticity	3
1.2. Bacteria's developed survival mechanisms	4
1.2.1. Secretion systems allow bacteria to thrive, but in turn cause pathogenesis	4
1.2.2. The variability of secretion systems illustrates the rich diversity of bacteria	6
1.2.3. "When you play the game of thrones, you win or you die": Bacterial competition	7
1.3. T4SS-mediated pathological processes	10
1.3.1. T4SS-mediated diseases in plants and humans	10
1.3.2. T4SS _F -mediated conjugation has greatly contributed to the antibiotic resistance crisis	12
1.4. Proteome stability and the role of disulfide isomerases	14
1.4.1. Protein stability is aided by redox-assisted folding	14
1.4.2. Cellular compartments for redox-assisted protein folding	14
1.4.3. Disulfide isomerases in prokaryotes and eukaryotes	15
1.5. The F-like Type IV Secretion System (T4SS_F) Transferosome	15
1.5.1. Comparisons among the F-, P-, and I-like T4SSs, and the process of conjugation	15
1.5.2. Pilus retraction in conjugation and the role of TrbI	20
1.5.3. Pilus extension and the role of TrbB and TraF of the T4SS _F	21
1.6. Research Significance & Objectives	22
2. CHAPTER TWO: MATERIALS & METHODS	23
2.1. Cloning <i>trbI</i> & <i>trbB</i> into pGEX-4T-2	23
2.1.1. pGEX4-T-2 plasmid extraction from <i>E. coli</i> DH5a	23
2.1.2. Phenol: chloroform plasmid extraction from <i>E. coli</i> XK1200	23
2.1.3. Polymerase Chain Reaction (PCR) amplification	24
2.1.4. PCR Purification of PCR amplicon	25
2.1.5. Restriction endonuclease (RE) Digestion	26
2.1.6. DNA Gel Extraction	26
2.1.7. Ligation <i>in vitro</i>	26
2.1.8. Transformation via heat shock & CaCl ₂	26
2.1.9. DNA Detection and Verification	27

2.2.	Expressing TrbI & TrbB in <i>E. coli</i> BL21(DE3)	28
2.2.1.	Large-scale expression using 1 mM IPTG	28
2.2.2.	Cell lysis by sonication	28
2.3.	Protein purification	28
2.3.1.	Affinity GST Sepharose Fast Protein Liquid Chromatography Purification	28
2.3.2.	Thrombin cleavage	29
2.3.3.	Secondary Purification using Size-Exclusion Chromatography	29
2.3.4.	Protein Qualification by SDS-PAGE	29
2.3.5.	Buffer exchange and Quantification	29
2.3.6.	Protein Quantification	30
2.4.	Crystallization trials	30
2.4.1.	General protein preparation	30
2.4.2.	Crystallization trials using the vapour diffusion method	30
2.4.3.	Protein preparation for the Hauptman-Woodward Institute high-throughput crystallization screening	31
2.5.	Assaying the Protein disulfide Isomerase (PDI) activity of TrbB	31
2.5.1.	Protein preparation	31
2.5.2.	Assay specifications	31
2.5.3.	Statistical Analysis	31
2.6.	Circular Dichroism (CD) Spectroscopy	31
2.6.1.	Estimating Protein 2 ^o structures using CD Spectroscopy	31
2.6.2.	Thermal denaturation	32
2.7.	Modelling using ColabFold-AlphaFold2 and PyMOL	32
2.7.1.	ColabFold-AlphaFold2 modelling using amino acid sequence	32
2.7.2.	Analysis using PyMOL v. 2.5.2	33
2.8.	Size-Exclusion Chromatography Multi-Angle-Light Scattering Small-Angle X-Ray Scattering	33
2.8.1.	Protein preparation	33
2.8.2.	Data collection	33
2.8.3.	Data processing	33
2.9.	Bio-Layer Interferometry	34
2.9.1.	Protein preparation	34
2.9.2.	Assay specifics and setup	34
2.10.	Analytical Size-Exclusion Chromatography	34
2.11.	¹H-¹⁵N Heteronuclear Single Quantum Coherence Nuclear Magnetic Resonance Spectroscopy	35
2.11.1.	Protein preparation	35
2.11.2.	Experiment details and HSQC parameters	35
2.12.	Figure Processing and Statistical T-test	35
3.	CHAPTER THREE: RESULTS & DISCUSSION	36
3.1.	Cloning, Expression, and Purification of GST-TrbB and GST-TrbI	36
3.1.1.	Cloning	36
3.1.2.	Expression and Purification	36
3.2.	Sampling of GST-TrbB's and GST-TrbI's chemical crystallization space as guides for future crystallization campaigns	40

3.2.1.	Crystallization trials on GST-TrbI	40
3.2.2.	Crystallization trials on GST-TrbB	41
3.3.	A volume model for GST-TrbB_{WT} illustrates its dynamics	43
3.3.1.	ColabFold-AlphaFold2 Models for GST-tagged cognate T4SS _F proteins	43
3.3.2.	Secondary structures estimations for TrbB _{WT} and TrbB ₅₇₋₁₈₁	45
3.3.3.	SAXS/CF-AF2 model for GST-TrbB _{WT} illustrates its dynamics	45
3.4.	Design of a more stable truncation construct, TrbB₅₇₋₁₈₁	50
3.4.1.	Truncation mutant protein, TrbB ₅₇₋₁₈₁ , design	50
3.4.2.	CD spectroscopy thermal denaturation studies support TrbB ₅₇₋₁₈₁ 's thermal stability	52
3.4.3.	Heteronuclear Single Quantum Coherence supports the stability of TrbB ₅₇₋₁₈₁	55
3.5.	Comparative structural studies on cognate T4SS_F proteins	56
3.5.1.	GST-TrbB is more similar to GST-TrbI at the secondary structure level compared to GST-TraF	56
3.5.2.	GST-TrbB _{WT} vs. GST-TrbI vs. GST-TraF ColabFold-AlphaFold2 models	56
3.5.3.	Comparing CF-AF2 and CD spectroscopy 2° structure estimations	57
3.6.	TrbB functions as a disulfide isomerase but it does not bind its client protein <i>in vitro</i>	57
3.6.1.	TrbB functions as a disulfide isomerase <i>in vitro</i>	57
3.6.2.	The N-terminus of TrbB _{WT} (M1-R56) is not required for enzymatic activity <i>in vitro</i>	58
3.6.3.	TrbB does not bind GST-TrbI <i>in vitro</i> using BLI	61
3.6.4.	Changes in secondary structure level in the presence of TrbB	61
3.6.5.	TrbB does not bind TraW <i>in vitro</i> using SEC, even in the presence of DTT	63
3.6.6.	Preliminary investigations on the chaperone activity of TrbB	68
3.6.7.	Other putative T4SS _F TrbB client proteins and CF-AF2 modelling	69
3.6.8.	Inconclusive findings as indications for the use of less stringent methods	70
4.	CHAPTER FOUR: CONCLUSIONS & FUTURE WORK	72
5.	REFERENCES	74

LIST OF ACRONYMS

HGT	Horizontal Gene Transfer
IS	Insertion Sequences
TE	Transposable Elements
SS	Secretion Systems
T4SS _F	F-like Type IV Secretion System
ssDNA	single stranded DNA
dsDNA	double stranded DNA
O/N	overnight
LB	Luria-Bertani broth
NaOAc	sodium acetate
NH ₄ OAc	ammonium acetate
Amp	ampicillin
EtBr	ethidium bromide
DNA	deoxyribonucleic acid
IPTG	Isopropyl β -D-1-thiogalactopyranoside
nt	nucleotides
FPLC	Fast Protein Liquid Chromatography
PCR	Polymerase Chain Reaction
EtBr	ethidium bromide
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
GST	Glutathione S-Transferase
PBS	Phosphate Buffered Saline
CF-AF2	ColabFold-AlphaFold2
HWI	Hauptman-Woodward Institute
SHG	Simple Harmonic Generation
UV-TPEF	Ultra-Violet Two-photon Excited Fluorescence
CD	Circular Dichroism
DTT	Dithiothreitol
HSQC-NMR	Heteronuclear Single Quantum Correlation Nuclear Magnetic Resonance
MW	molecular weight
MWCO	Molecular weight cut-off
SEC	Size Exclusion Chromatography
MALS	Multi-Angle Light Scattering
SAXS	Small-Angle X-Ray Scattering
V _p	Porod volume
V _c	Volume of correlation
R _g	Radius of gyration
DI	Disulfide Isomerase
BLI	Bio-layer Interferometry
FRET	Fluorescence Resonance Energy Transfer

pTM	predicted Template Modelling
ipTM	interface predicted Template Modelling

LIST OF FIGURES

- Figure 1. Three classical mechanisms of Horizontal Gene Transfer.
- Figure 2. Transposable elements confer genomic dynamics and thus plasticity.
- Figure 3. Prokaryotic Secretion Systems demonstrate the rich phenotypic diversity of bacteria.
- Figure 4. Conserved themes and regulation of secretion systems.
- Figure 5. Different effects of T6SS-mediated injection of effector toxins into target cells.
- Figure 6. The family of T4SS is functionally diverse.
- Figure 7. Global antibiotic consumption classified according to national income.
- Figure 8. Comparison of the F-like, P-like, and I-like Type IV Secretion Systems.
- Figure 9. Differences between F-like, P-like, and I-like Type IV Secretion Systems.
- Figure 10. Model of the F-like type IV secretion system (T4SS) transferosome based on available structural information.
- Figure 11. T4SS_F-mediated conjugation.
- Figure 12. Representative cloning, expression, and purification of GST-TrbI.
- Figure 13. Representative cloning, expression, and purification of GST-TrbB and TrbB_{WT}.
- Figure 14. Representative cloning, expression, and purification of GST-TrbB₅₇₋₁₈₁ and TrbB₅₇₋₁₈₁.
- Figure 15. Representation of the chemical space of crystallization sampled by crystallization trials.
- Figure 16. ColabFold-AlphaFold2 models and primary sequences of three T4SS_F proteins.
- Figure 17. Secondary structure estimations.
- Figure 18. Size Exclusion Chromatography (SEC) Multi Angle Light Scattering (MALS) on protein sample and molecular weight estimations characteristic of a homodimeric GST-TrbB.
- Figure 19. AlphaFold2 model of homodimeric GST-TrbB fitted into *ab initio* 3D reconstruction for homodimeric GST-TrbB from SAXS data.
- Figure 20. ColabFold-AlphaFold2 model for TrbB_{WT} emphasizing the thioredoxin domain and residues T37-R56 deleted in TrbB₅₇₋₁₈₁ construct.
- Figure 21. Thermal denaturation profiles of α -helices in (A) TrbB₅₇₋₁₈₁ and (B) full-length TrbB_{WT}.
- Figure 22. TrbB₅₇₋₁₈₁ residues predicted to form α -helices.
- Figure 23. ¹H-¹⁵N Heteronuclear Single Quantum Correlation (HSQC) Spectra of TrbB_{WT} and TrbB₅₇₋₁₈₁.
- Figure 24. TrbB functions as a disulfide isomerase *in vitro*.
- Figure 25. TrbB does not bind TrbI *in vitro*.
- Figure 26. Observed changes in the secondary structure composition of GST-TrbB/His6TraW compared to the proteins alone.
- Figure 27. Preliminary investigation of the TrbB/His₆TraW putative interaction.
- Figure 28. ColabFold-AlphaFold2 (CF-AF2) heterodimeric models and their observed binding residues.
- Figure 29. TrbB can increase the thermostability of Lysozyme.
- Figure 30. The active CXXC motif is not located at the binding interface of CF-AF2 heterodimeric models aside from TrbC.

LIST OF TABLES

Table 1	Primers used for PCR
Table 2	Thermocycler settings for PCR amplification
Table 3	SEC-MALS-SAXS parameters for homodimeric GST-TrbB
Table 4	Welch's T-test for Figure 14. Analysis was performed using GraphPad Prism v.9.5.1 for macOS.

1. CHAPTER ONE: INTRODUCTION

1.1. Rapid genetic recombination is a driver of the rich phenotypic variability in bacteria

1.1.1. Bacterial diversity and adaptability

The rich diversity of bacterial life is attributable to their rapid reproductive capabilities, and this mechanism has led to their development of elaborate phenotypes to thrive even in the presence of deliberate stressors. These diverse set of phenotypes have allowed bacteria to defend themselves against bacteriophages (one of their main evolutionary predators),¹ survive in eukaryotic tissues protected by a complex immune system², survive when subjected to radiation and other mutagens³, and even develop macromolecular mechanisms to evade antibiotics^{4–6}. Bacteria can reproduce through vertical gene transfer through a process called binary fission, but this mechanism of reproduction is limited with respect to conferring genetic variability, and cannot explain the phenotypic diversity of bacteria, because the resulting daughter cells are genetically identical⁷.

1.1.2. Bacteria have a genome subject to mutations

One key explanation for bacteria's phenotypic diversity was proposed in 1943 by Salvador Luria and Max Delbrück. The consensus of the scientific community at the time was that bacteria is a separate form of life compared to “higher” organisms like humans, fruit flies (*Drosophila*), and maize, and many doubted if bacteria even had genetic material. However, Luria and Delbrück showed, through statistical analyses, that spontaneous mutations influence bacterial genomes^{8,9}, implicating that they, like “higher” organisms, are subject to adaptation by natural selection. Their work jumpstarted the field of bacterial genetics and earned them the 1969 Nobel Prize in Physiology or Medicine, which they shared with Alfred Hershey.

1.1.3. Horizontal gene transfer; the phenomena of transformation and conjugation

Mutations allow bacteria to adapt and evolve, but it cannot explain the rapid rate at which they do so. In 1944, Oswald Avery, Colin McLeod and Maclyn McCarty showed that DNA was the substance that allows an initially non-virulent *Pneumococcus* to adopt a virulent phenotype¹⁰, the

very first evidence for horizontal gene transfer (HGT) in bacteria, through the process now known as transformation. Transformation involves the ability of competent bacteria to uptake DNA from its immediate environment, without the use of a vector (**Fig. 1**). The uptake of DNA occurs through complex energy-requiring processes¹¹.

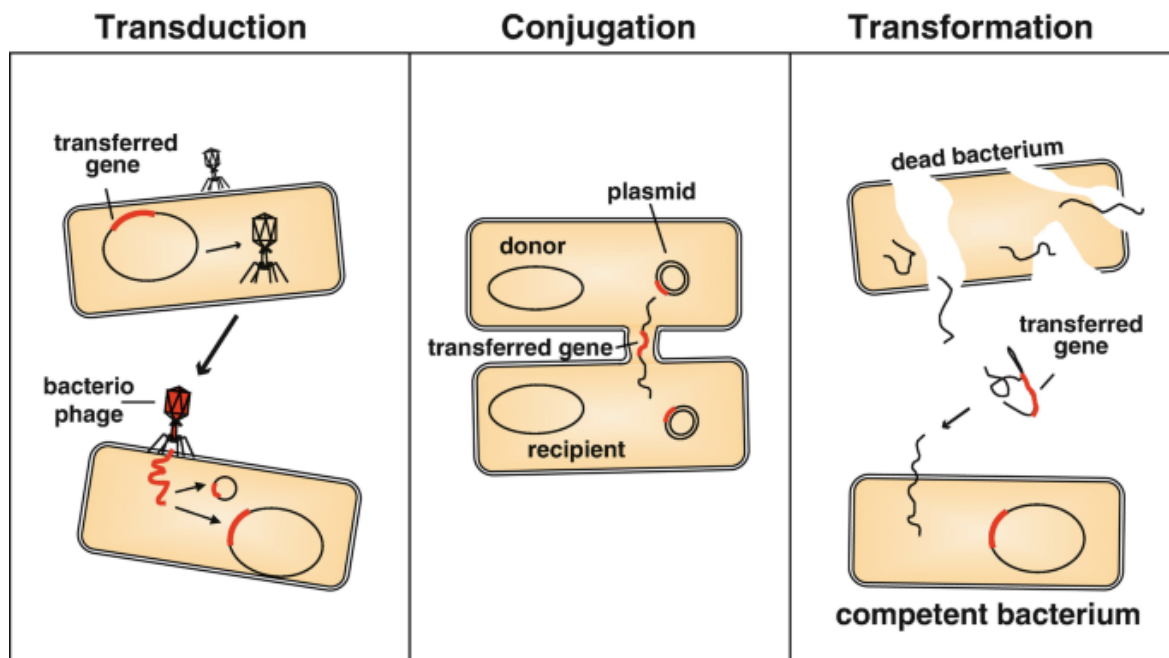


Figure 1. Three classical mechanisms of Horizontal Gene Transfer. Transduction involves a viral vector (usually bacteriophages) for inter-bacterial DNA exchange. Conjugation requires contact between donor (termed F+ for a bacterium bearing the Fertility factor, F) and recipient bacteria (F-) through a specialized tube-like appendage called a pilus. Transformation involves the ability of competent bacteria to uptake DNA from its immediate environment, without the use of a vector. Figure adapted from Blockesch (2016)¹².

In 1946, the work of Joshua Lederberg and Edward Tatum elucidated the second form of HGT¹³, which Lederberg later named “conjugation”, catapulting our ability to understand the vast diversity of bacterial phylogeny¹⁴. Contrary to the dominant belief of their time, they showed that bacteria engage in “sex”, allowing for genetic recombination, plasticity, and increased genetic variability in their populations. Conjugation requires contact between donor (termed F+ for a bacterium bearing the Fertility factor, F) and recipient bacteria (F-) through a specialized tube-like appendage called a pilus (**Fig. 1**). Conjugative pili allow for donor bacteria to share their

extrachromosomal circular DNA, called plasmids (coined by Lederberg), which often encode non-essential traits that aid in the bacterium's survival. Joshua Lederberg's revolutionary work was recognized by the Nobel Prize committee, earning Lederberg, together with George Beadle and Edward Tatum, the 1958 Nobel Prize in Physiology or Medicine.

1.1.4. Transduction; the third mechanism of Horizontal Gene Transfer

Detection of the phage, Lambda, *E. coli* K-12 cultures in 1950 by Esther Lederberg was one of the first genetic observations implicating viruses with genomic plasticity in some way^{15,16}. It was later established by the group of Joshua Lederberg, including his wife Esther Lederberg, that Lambda carry some *E. coli* K-12 genes, particularly those that are near the site where Lambda viral genome integrated. In 1952, they published their discovery of the phenomena of transduction (**Fig. 1**), the reliance on a viral vector (usually bacteriophages) to exchange foreign DNA among each other^{11,17}, using the model bacteria *Salmonella*¹⁸. They found that, even when they prevent cell-cell contact necessary for conjugation by filtering the bacterial medium using a filter with pores smaller than bacteria, genetic recombination still occurred, implicating another horizontal gene transfer mechanism. They ruled out transformation because they observed recombination even when deoxyribonucleases (DNases) were added to the growth medium. Now, transduction is a well characterized classical HGT mechanism.

1.1.5. Transposable elements; their role in Horizontal Gene Transfer and genomic plasticity

Barbara McClintock's discovery of transposable elements (TEs), also dubbed "jumping genes", in the model organism maize¹⁹ elucidated the dynamicity of genes in the genome, implicating TEs with genomic plasticity. McClintock's Nobel-prize-winning work was done in 1950 in the model organism maize¹⁹, but her work ushered the eventual discovery of TEs in prokaryotes, which received research attention during the 1970s²⁰. The dynamics of TEs are in and of itself significant in conferring genomic plasticity, but they can also often flank passenger genes such as antibiotic resistance or virulence genes (**Fig. 2**), allowing these genes to be mobile in the genome^{21,22}. Moreover, TEs can often jump from the bacterial chromosome to plasmids, and vice versa, affording survival-conferring genes mechanisms to evade detection.

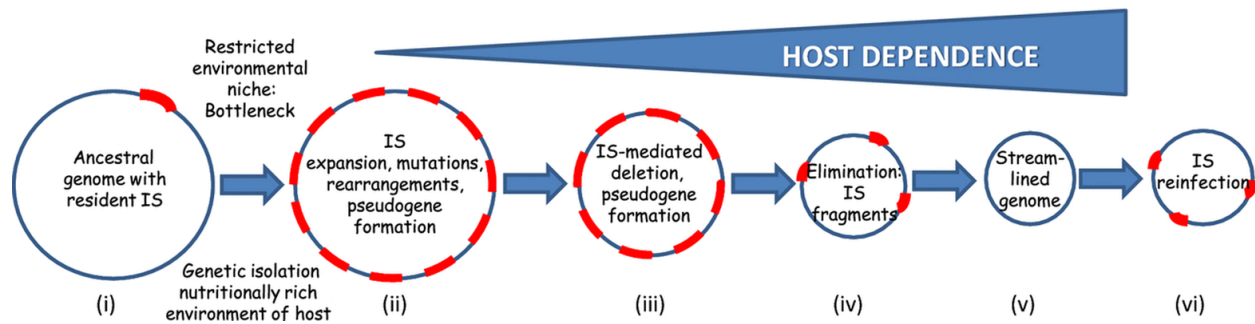


Figure 2. Transposable elements confer genomic dynamics and thus plasticity. Insertion sequences (IS; red) are the smallest and most numerous autonomous transposable elements (TEs) in bacteria. IS expansion (transition from i to ii) can be driven by bacterial population bottlenecks, competition, and even homologous recombination events between identical IS copies. Absence of direct selective pressures over time leads to deletion of ISs with adjacent DNA sequences (ii to iv). Eventually, the IS infection cycle repeats (v to vi). Transposable elements can flank passenger genes, such as those that confer antibiotic resistance, making these passenger genes mobile in the genome. Figure adapted from Siguier et al. (2014)²¹.

1.2. Bacteria's developed survival mechanisms

1.2.1. Secretion systems allow bacteria to thrive, but in turn cause pathogenesis

Bacteria utilize a multitude of mechanisms to survive in stressful and competitive external environments, in many cases harming other cells. Secretion systems (SSs) are large multi-protein complexes expressed by bacteria that help them to survive and thrive through a multitude of functionalities²³. SSs can traverse through a target cell's lipid membrane(s), allowing them to perform acts of host pathogenesis such as invading eukaryotic hosts by damaging tissue sites and weakening immune defenses. SSs mediate the secretion of toxins and virulence genes, enhancing bacterial survivability, attachment to host cells, intoxicating target cells and thus outcompeting other microbes or the host cell for resources in the immediate environmental niche. Among the eleven²⁴ multi-protein SSs identified in total (**Fig. 3**), those associated with gram-negative bacteria are unique compared to those in gram-positive bacteria because they can traverse through at least two lipid membranes (the inner and outer membranes), and often up to three (including target cell's membrane)^{23,25}. However, common to both gram-negative and gram-positive secretion systems is their reliance on feedback mechanisms to up- or downregulate the

expression of SS component proteins in response to replenishment needs or energy conservation such as when no substrates are readily being excreted (**Fig. 4**)²⁶.

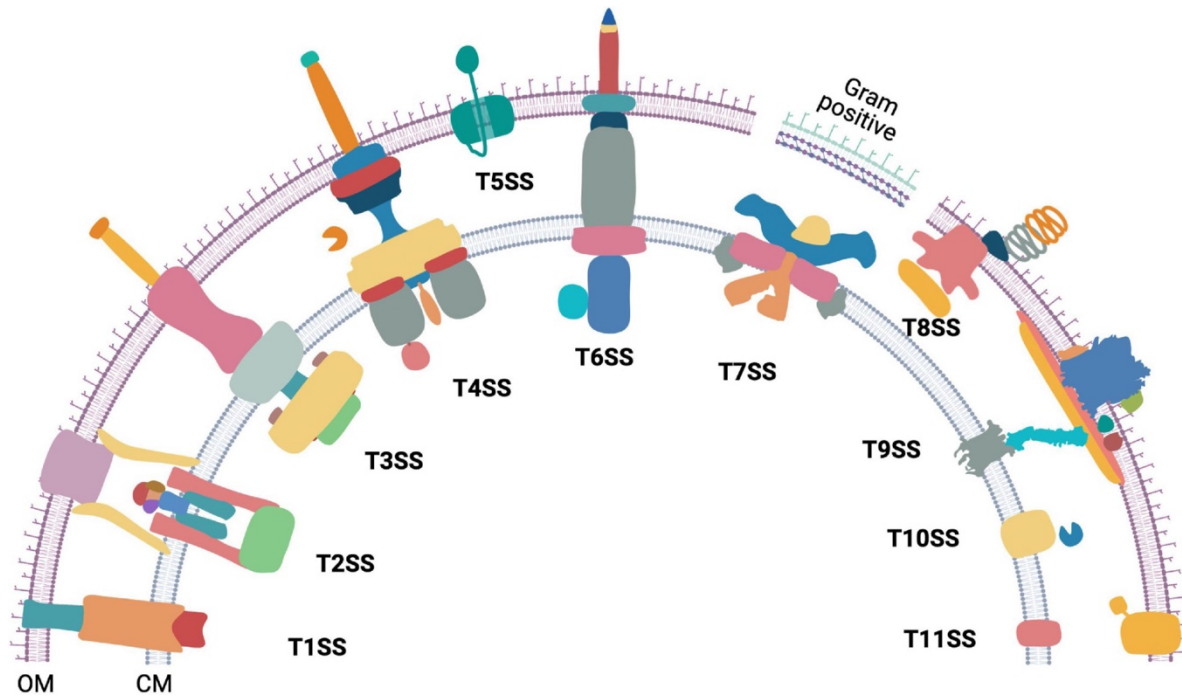
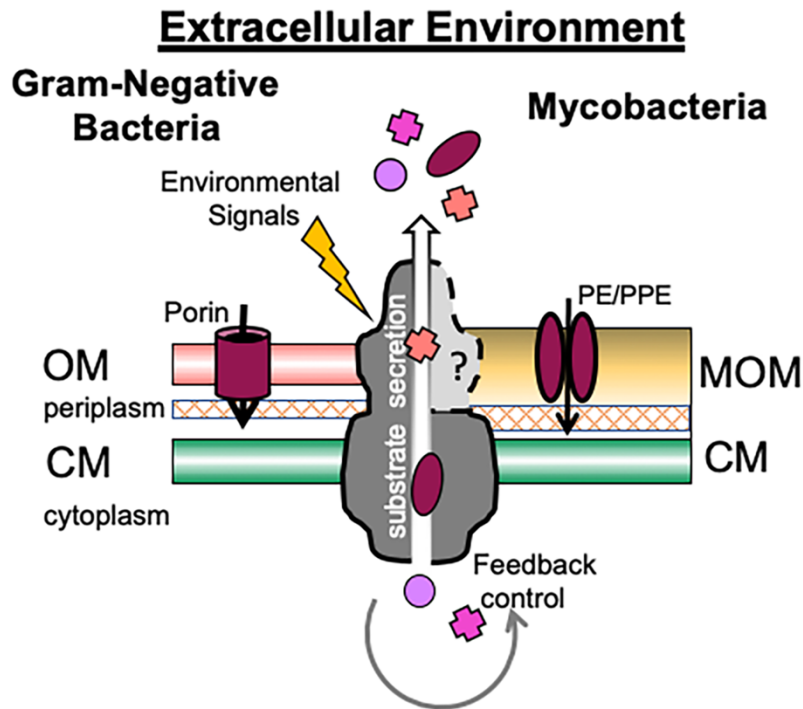


Figure 3. Prokaryotic multi-protein Secretion Systems demonstrate the rich phenotypic diversity of bacteria. The T2SS and T9SS, T3SS and T6SS are functionally similar. T7SS-mediated protein secretion is found within some Gram-positive members of the phylum Actinobacteria that have an outer lipid layer. The T9SS is notable for relying on a proton motive force. The T8SS interestingly plays an important role in biofilm formation. Most of the other secretion systems are functional in increasing bacteria's survival in the context of their multi-cellular host. Figure adapted from Trivedi et al. (2022)²⁴.



Common themes in sensing & regulation

- Some specialized systems are essential
- Some systems function at discrete points during infection
- Some secretion systems serve multiple pathogenic roles
- Secretion systems sense the status of components and substrates and trigger changes in gene expression

Figure 4. Conserved themes and regulation of secretion systems. Only some gram-negative bacterial secretion systems (left) respond to environmental signals, while both gram-negative and *Mycobacteria* (right; gram-positive) rely on feedback controls. Gram-negative bacteria utilize porins for solute transport, while *Mycobacteria* use PE/PPE proteins. Figure adapted from Nicholson et al. (2022)²⁶.

1.2.2. The variability of secretion systems illustrates the rich diversity of bacteria

The 11 bacterial multi-protein secretion systems are all known to be involved in the secretion of proteins, or protein-DNA complexes²⁴, while some are associated with cellular motility^{27–29}, such as Types 2, 3, and 9 (**Fig. 3**). The Type I Secretion System (T1SS) is reported to be involved in the one-step excretion of unfolded proteins into the extracellular space³⁰. T2SS works in concert with the Sec or Tat secretion pathways, which secrete proteins to the periplasm, and the T2SS secretes these proteins in an ATP-hydrolysis-dependent manner from the periplasm to the extracellular environment³¹. The T3SS is functional in excreting proteins from the bacterial cytosol to the

extracellular environment or across the protective cell membrane of eukaryotic cells³². The T4SS exports both proteins and DNA from the cytosol of T4SS-containing bacteria to either a prokaryotic or a eukaryotic host cell, and it is also employed by bacteria to take up extracellular DNA^{5,33}. Interestingly, the autotransporter T5SS is comprised of an outer membrane protein whose β -barrel translocator domain excretes effector proteins with diverse functions such as cell growth inhibition, proteases, lipases, adhesins, and circumventing multi-cellular host's immune system^{34,35}. The T6SS, like some T3SSs, are employed to inject effector proteins from the cytoplasm directly across other bacterial or eukaryotic host membranes^{36,37}. The T7SS is used by pathogenic *Mycobacteria* for host cell immune system evasion^{38,39}, and is therefore subject to active investigations by many groups worldwide as a drug target to mitigate the tuberculosis epidemic⁴⁰. The T8SS is functional in biofilm formation, an aggregation-based community of bacteria to adapt to harsh environments, through the excretion of amyloid protein fibers called curli⁴¹. The T9SS substrates are delivered to the periplasm by Sec transport pathway and T9SS secretes them across the outer membrane, similar functionally to the T2SS which also co-functions with the Sec pathway. The T9SS contains a motor that drives secretion, one of the only three known biological rotary motors driven by a proton motive force²⁴. The T10SS secretes hydrolytic enzymes and toxins, some involved in peptidoglycan modifications⁴². A new secretion system, proposed to be the T11SS, is under active investigation and is thought to be functional in improving bacterial survival in multi-cellular hosts⁴³.

1.2.3. “When you play the game of thrones, you win or you die”: Bacterial competition

Bacterial competition is a highly active area of research and some of the mechanisms bacteria utilize to outcompete other cells in microenvironments with limited resources include (but are not limited to) contact-dependent inhibition (CDI) systems^{44,45}, the T6SS^{36,37}, and more recently the F pilus⁴⁶. CDI has been reported in the genus *Burkholderia*, *Pseudomonas*, and *Acinetobacter* among others, since its first report in 2005 from a uropathogenic strain EC93 of *Escherichia coli*^{44,45}. Significant homology is observed in the CdiA protein among strains of *E. coli* and *Pseudomonas aeruginosa*, with the highly conserved Filamentous Hemagglutinin 1 domain being responsible for delivering the cytoplasm-translocation domain and the C-terminal domain of the

Filamentous Hemagglutinin 2 domain (CdiA-CT) into the cytoplasm of target gram-negative bacteria^{47,48}. The CdiA-CT intoxicates target bacteria through a multitude of molecular mechanisms, such as the formation of pores in the inner membrane of gram-negative bacteria and the degradation of target cell's tRNA and DNA^{47,49–51}. The T6SS delivers toxic effector proteins to competitor cells using a mechanism akin to that of bacteriophages^{36,37}. The T6SS assembly is initiated by the cytoplasmic baseplate-like structure (TssA/-E/-F/-G/-K complex) and membrane component (TssJ, TssL, and TssM) proteins. The second step involves the contraction of the external sheath (composed of TssBC) to propel an inner tube (composed of Hcp proteins), which then punctures the target cell. The puncturing device is composed of PAAR and VgrG, which together deliver the toxins to the target cell and cause diverse effects (**Fig. 5**). The F pilus was also reported to mediate the import of a CDI toxin, CdiA-CT, from uropathogenic strain 536 of *E. coli*⁴⁶. The deletion of *trbI*, the gene implicated with F pilus retraction^{5,52}, in the genome of *E. coli* 536, resulted in normal *E. coli* growth even in the presence of a purified toxic protein in the medium. However, complementation of a plasmid bearing *trbI* in the *E. coli*_{ΔtrbI} liquid culture resulted in growth inhibition in the presence of the purified toxin.


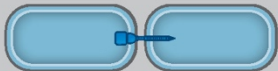
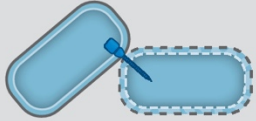
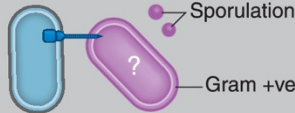
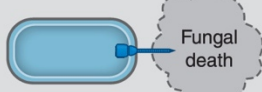
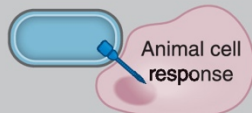
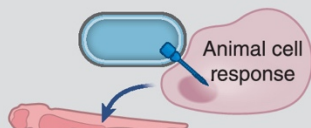
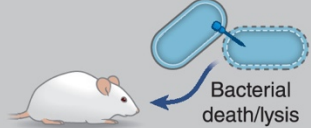
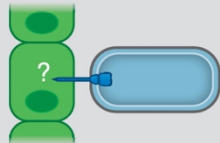
Number	Function	Description	Model	References
I	General secretion	T6SS effector secretion into the environment, e.g. secretion of nutrient scavenging proteins		(Lin et al., 2017)
II	Growth stasis	Effector delivery into competing bacteria for growth inhibition		(Whitney et al., 2015)
III	Bacterial death	Effector delivery resulting in bacterial cell death, e.g. killing of bacteria releases nutrients, DNA or enables niche access. Immunity proteins facilitate kin cell recognition		(Russell et al., 2011)
IV	Gram positive impact	T6SS+ results in higher sporulation of Bacillus. Unknown if this is mediated by effectors		(Molina-Santiago et al., 2019)
V	Fungal death	Effector delivery induces fungal cell death		(Trunk et al., 2018)
VI	Effector driven cell level response	Effectors result in a cell level response, e.g. killing of amoebae or promoting internalisation		(Hachani et al., 2016)
VII	Effector driven organ response	Effector results in organism level response, e.g. VgrG-1 delivery triggers Zebra fish intestinal contractions, dispelling microbiota, enhancing colonization		(Logan et al., 2018)
VIII	Indirect host effects	Indirect modulation of the host via T6SS, e.g. T6SS bacterial killing results in host inflammation and increased colonization		(Fast et al., 2018; Zhao et al., 2018)
IX	Plant cell subversion	Plant host manipulation. Currently unknown if this is mediated by effectors delivery		(Shyntum et al., 2015; L. Zhang et al., 2014)

Figure 5. Different effects of T6SS-mediated injection of toxins into target cells. The T6SS is generally associated with bacterial competition. Figure adapted from Allsopp et al. (2020)³⁷.

1.3. T4SS-mediated pathological processes

1.3.1. T4SS-mediated diseases in plants and humans

T4SS found in pathogenic bacteria perform other functions (**Fig. 6**) including the transfer of toxic effector proteins between bacterium and eukaryotic hosts to aid in their survival². Examples of bacterial species that employ the T4SS to secrete effector proteins include *Legionella pneumophila* (causes Legionnaire's pneumonia in humans), *Bordetella pertussis* (whooping cough), *Bartonella henselae* (cat-scratch fever), and *Helicobacter pylori* (peptic ulcer and gastric cancer). Similarly, *Agrobacterium tumefaciens* is known to inject nucleoprotein complexes into plant cells for increased survivability, a mechanism exploited in biotechnology to induce transient expression of foreign genes in plant models called agroinfiltration⁵³. In addition, conjugative T4SS are key in the intra- and inter-species dissemination of plasmid DNA⁵⁴. From a utilitarian perspective, the goal of conjugative mating is to gain and disseminate genes that confer increased survivability. An example of a trait that is of utmost importance in bacterial survivability, and in human affairs, is antibiotic resistance.

Processes in which T4S systems are involved

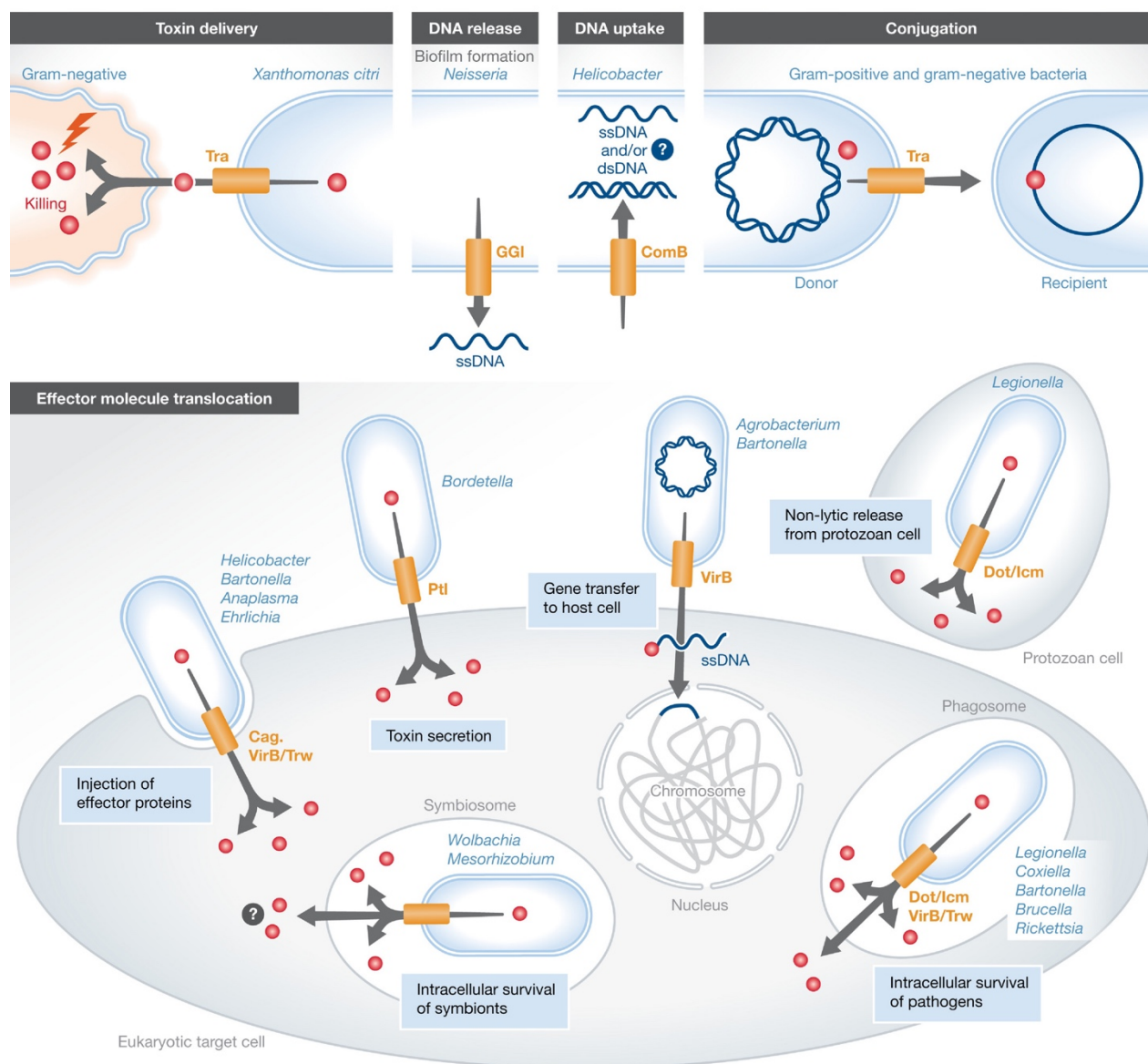


Figure 6. The family of T4SS is functionally diverse. The T4SS is functional in inter-bacterial mating and as mediator of inter-kingdom transfer processes. Figure adapted from Waksman (2019)⁵⁵.

1.3.2. T4SS_F-mediated conjugation has greatly contributed to the antibiotic resistance crisis

Since Alexander Fleming's breakthrough discovery in 1928, antibiotics have become the primary, and in some cases the only, therapeutic choice for managing a multitude of infectious diseases⁵⁶. Fleming published his discovery of penicillin in 1929, but his work received little recognition. It was not until the 1940s during the second world war, when the demand for antibiotics to treat infected soldiers was high, that scientists Howard Florey and Ernst Chain developed methods to mass produce penicillin. Together, the work of Alexander Fleming, Howard Florey, and Ernst Chain was recognized by the Nobel committee, earning them the Nobel Prize in Physiology or Medicine in 1945.

Profound advancements in suppressing infectious diseases in the late 1960s, heralded by antibiotics, minimized the perceived threat of infectious micro-organisms, and US Surgeon General William H. Stewart was quoted as saying: "it is time to close the book on infectious diseases and declare the war against pestilence won" (p.156)⁵⁷. However, this idea was rash due to one key factor: evolution. Indeed, bacteria's ability to adapt quite rapidly has led to one of the twenty-first century's greatest problems; antibiotic resistance. Much emphasis has been placed on developing novel pharmaceutical drugs since the early 1980s, but the number of developed and approved novel antibiotics over the past three decades has steadily decreased⁵⁶. Still, cases of resistance to antibiotics continue. For instance, *Mycobacterium tuberculosis*, known to cause tuberculosis, was classified in 2000 as extensively drug resistant to first-line antibiotics.

The problem of antibiotic resistance is attributable to several, and largely inter-related, socio-political, and economic factors. Widespread use of antibiotics in agriculture is one cause of the resistance problem. Profoundly, 80% of antibiotics sold in the US are used in livestock for preventative measures⁵⁶. That is, livestock are mass treated with antibiotics to make a larger proportion of the livestock population less likely to die from infections. This is done regardless of whether livestock exhibit symptoms of bacterial infections. Furthermore, the use of antibiotics in agriculture also spreads resistance. In fact, up to 90% of the antibiotics given to livestock are excreted in urine and stool, which are then spread through fertilizers and runoff⁵⁶.

Ironically, while the mass production of livestock is efficient in terms of yield, problems like antibiotic resistance are consequences. As a result of the massive herding of livestock,

infections are transmitted easier among the animals that live side-by-side, often in improperly ventilated factories⁵⁶. In the past, the proactive use of antibiotics to control such infections was an effective measure. However, through natural selection, mutant bacteria resistant to antibiotics flourish and proliferate, thus leading to the modern crisis of resistance.

Another chief cause of the crisis is the overuse of antibiotics worldwide. Many countries do not have regulations to control antibiotics. Even in countries that have regulations, overuse and over-prescription is prevalent. Data from Klein et al. (2018)⁵⁸ reveal the global antibiotic consumption in the years 2000-2015. High-income countries have the highest defined daily doses (DDD) per 1000 inhabitants per day, but consumption in other countries have increased from 2000 to 2015 (**Fig. 7A**). In 2015, low- & lower-middle-income countries represented the greatest total antibiotic consumption (**Fig. 7B**). The authors suggest that this is due to the growing population in low- and middle-income countries.

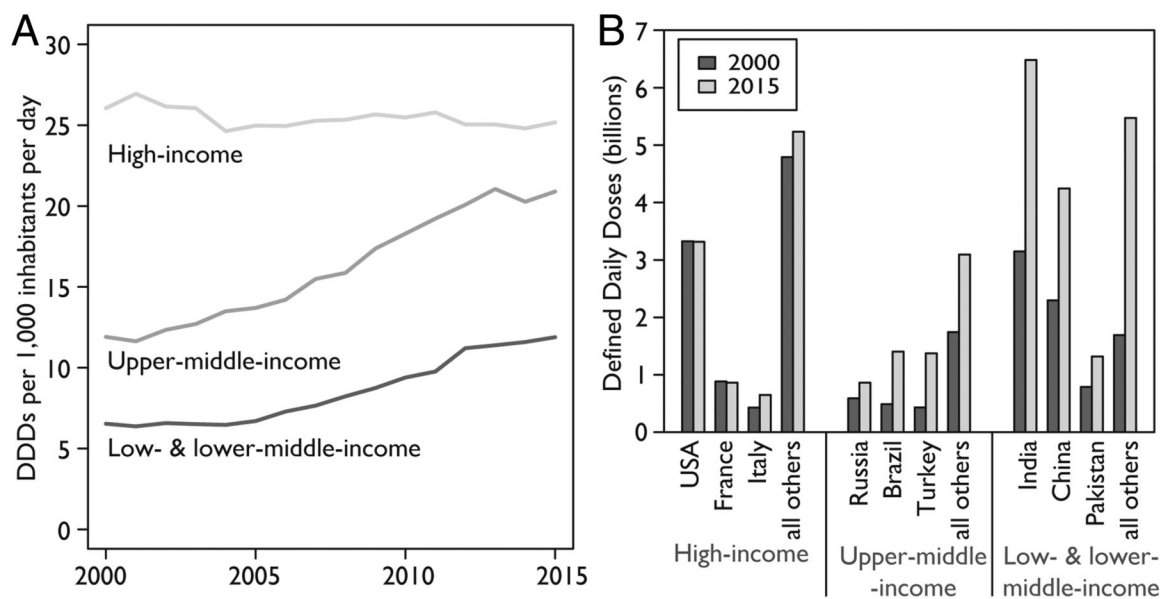


Figure 7. Global antibiotic consumption classified according to national income. (A) Defined daily doses (DDD) per 1000 inhabitants per day reveal that antibiotic consumption has steadily increased in low- & lower-middle countries and upper-middle countries, while that of the high-income countries has remained relatively steady. (B) However, other data suggest that total antibiotic consumption (DDD) in 2015 is greater in low- & lower-middle income countries. Figure adapted from Klein et al. (2018)⁵⁸.

1.4. Proteome stability and the role of disulfide isomerases

1.4.1. Protein stability is aided by redox-assisted folding

Protein stability depends on a plethora of intrinsic and extrinsic conditions. One factor that can significantly confer increased structural and physicochemical stability to a protein is the presence of intramolecular covalent disulfide bonds^{59,60}. Accordingly, both prokaryotes and eukaryotes have evolved elaborate mechanisms to catalyze the formation, rearrangement, and breakage of disulfide bonds within their proteomes^{61,62}. Therefore, a common feature to these gram-negative SSs are disulfide isomerases and protein-folding-assisting pathways. However, these DIs are understudied in the context of SSs.

1.4.2. Cellular compartments for redox-assisted protein folding

While eukaryotes and prokaryotes are significantly distinct from one another, both have designated compartments for redox-assisted protein folding, particularly with respect to the formation, rearrangement, and breakage of disulfide bonds. With respect to cellular compartmentalization, eukaryotes are more complex. Redox-assisted folding occurs in numerous organelles in eukaryotes including, but not limited to, the endoplasmic reticulum^{63,64}, the Golgi apparatus⁶⁵, the mitochondria^{66–68}, and lysosomes⁶⁹. These organelles are not present in prokaryotes; however, similar environments are present *albeit* with less variation. In gram-negative bacteria, redox-assisted protein folding occurs mainly in the periplasm, thought to be the evolutionary precursor of lysosomes⁷⁰. The periplasm offers a more oxidizing environment due to its higher reduction potential of -165 mV⁷¹ compared to that of the cytoplasm, reported to be between -260 and -280 mV^{72–74}. Gram-positive bacteria lack compartmentalization altogether, consisting of a single membrane surrounded by an extensive network of peptidoglycan, posing questions such as where, and if, redox-assisted protein folding occurs in these bacteria. Certainly, it has been found that some gram-positive bacteria (*e.g.*, the *Firmicutes*) do not rely on disulfide bond formation, while some do (*e.g.*, the *Actinobacteria*)^{75–77}. Interestingly, disulfide-bond-forming pathways in gram-negative bacteria are believed to be non-essential for growth⁷⁸, rather they are required for pathogenesis⁷⁹. Therefore, characterization of

proteins involved in these pathways can provide avenues for the design of novel strategies to mitigate bacterial pathogenesis.

1.4.3. Disulfide isomerases in prokaryotes and eukaryotes

The first such eukaryotic enzyme identified was aptly named protein disulfide isomerase (PDI)^{80–82}. Further attention into the importance of disulfide bonds in protein folding and stability led to the identification of the first bacterial PDI counterpart, DsbA, which also functions as a disulfide isomerase (DI)⁷⁸. The identification and characterizations of novel DIs as related to redox-assisted protein folding in eukaryotes has received more attention^{62,65} compared to that of prokaryotes^{83,84}. Understanding the structure and function of DIs in prokaryotes, and comparing them to eukaryotic DIs, will provide further insights into their evolution, the bacterial proteome and the role of DIs in its stability that contributes to bacteria's adaptable resilience to environmental stressors.

1.5. The F-like Type IV Secretion System (T4SS_F) Transferosome

1.5.1. Comparisons among the F-, P-, and I-like T4SSs, and the process of conjugation

Genetic recombination by horizontal gene transfer works in concert with mutations and selective pressures caused by the over-use and misuse of antibiotics to confer genomic plasticity, explaining the vastly rich phenotypic diversity of bacteria. The conjugative T4SS is structurally and functionally diverse, accounting for bacteria's development in different environmental niches. There are functional differences between F-like and P-like T4SS, with P-like systems lacking conserved F-like auxiliary proteins (TraF, -G, -H, -N, -U, -W, and TrbB, -C, and -I)^{85,86} (**Fig 8**), and the key difference between these systems are simply the conjugative plasmids that encode them (**Fig. 9**). In contrast to F-like systems, the conjugative ability of P-like systems is lower in liquid media than on solid media, reflecting the developed phenotypic diversity of bacteria to survive in different ecological niches⁸⁷.

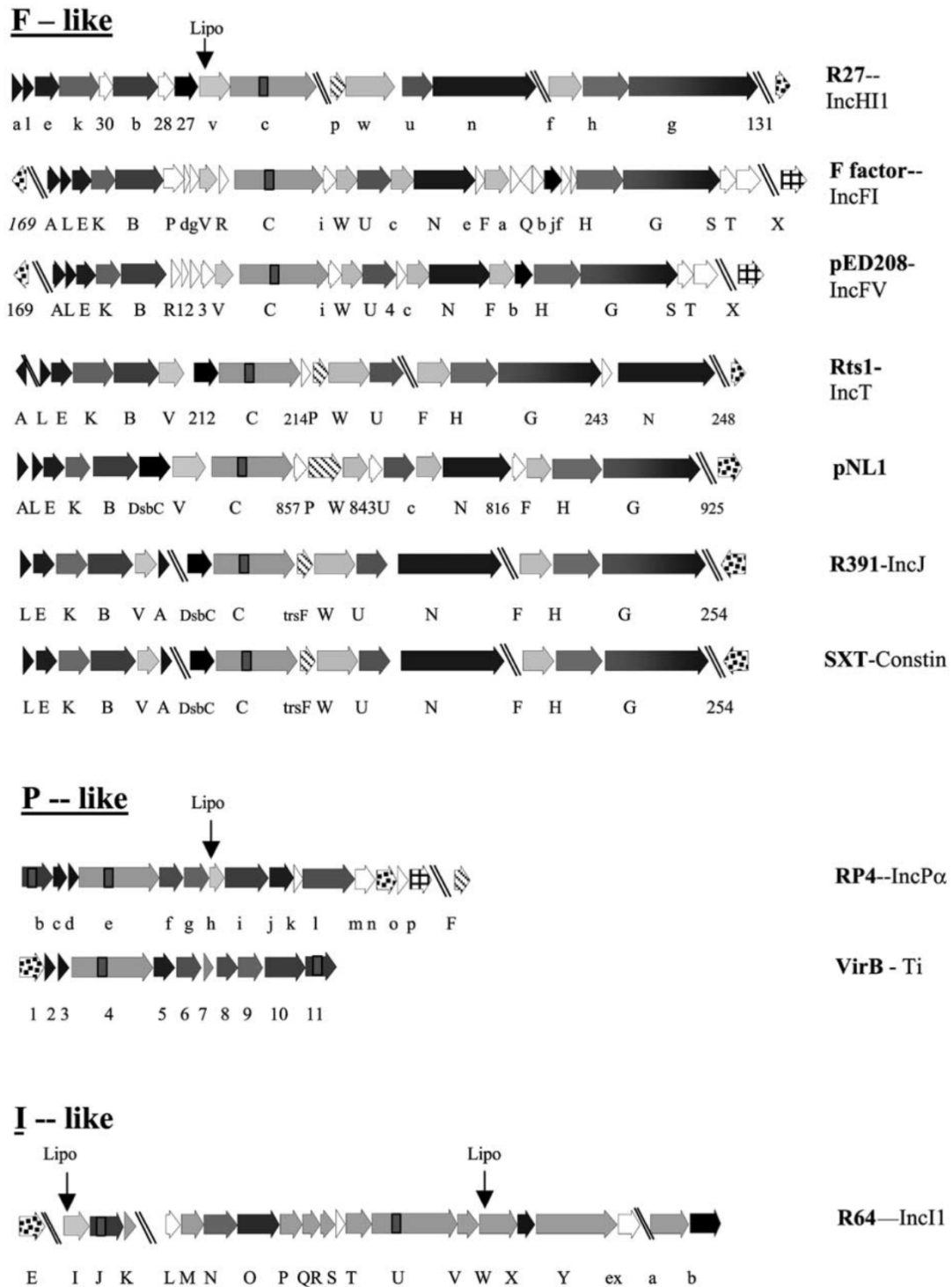
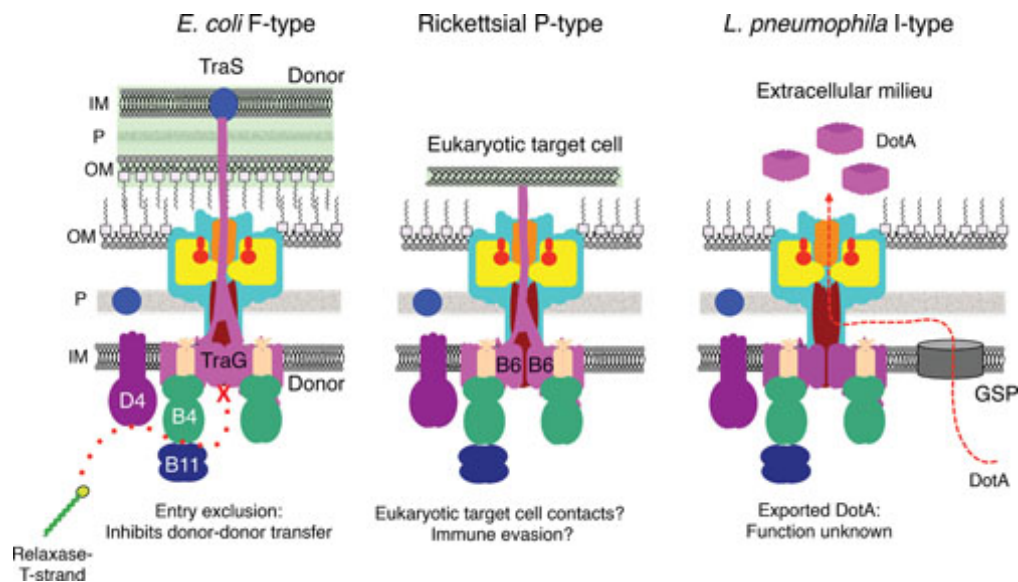


Figure 8. Comparison of the F-like, P-like, and I-like Type IV Secretion Systems. Transfer genes are shown in colour/pattern and homologues conserved across and within F-, P-, and I-like T4SS have the same colour/pattern. Light gray genes have no homology to other T4SS family. Upper case gene names represent Tra genes; lower case represent Trb (F, pNL1 and RP4) or Trh (R27). Figure adapted from Lawley et al. (2003) ⁸⁶.



Plasmid	Inc group	Organism
F-like		
F	FI	<i>Escherichia coli</i>
R100	FII	<i>Salmonella enterica</i> serovar Typhimurium
pSLT		<i>Salmonella enterica</i> serovar Typhimurium
pED208	FV	<i>Salmonella enterica</i> serovar Typhi
pYJ016		<i>Vibrio vulnificus</i>
H-like		
pNL1		<i>Novosphingomonas aromaticivorans</i>
Rts1	T	<i>Proteus vulgaris</i>
R391	J	<i>Providencia rettgeri</i>
SXT		<i>Vibrio cholerae</i>
pCAR1		<i>Pseudomonas resinovorans</i>
pHCM1	HI1	<i>Salmonella enterica</i> serovar Typhi
R27	HI1	<i>Salmonella enterica</i> serovar Typhi
R478	H2	<i>Serratia marcescens</i>
I-like		
R64	I1	<i>Salmonella enterica</i> serovar Typhimurium
ColIb-P9	I1	<i>Shigella sonnei</i>
pEL60		<i>Erwinia amylovora</i>
pCTX-M3		<i>Citrobacter freundii</i>
None	None	<i>Escherichia coli</i>
None	None	<i>Escherichia coli</i>
None	None	<i>Escherichia coli</i>

Figure 9. Differences between F-like, P-like, and I-like Type IV Secretion Systems. Architectural differences between F-, P-, or I-like T4SS [top; Christie (2006)⁸⁸], and some of the plasmids encoding them [bottom; Elton et al. (2005)⁸⁹].

In *Escherichia coli*, and in other gram-negative bacteria, conjugation is mediated by the F⁺-like Type IV Secretion System (T4SS_F). The T4SS_F transferosome is a multi-protein system encoded by the *tra* operon (**Fig. 10**). The T4SS_F transferosome regulates the formation of a mating bridge between a F⁺ bacterium (bacterium bearing the F plasmid) and a F⁻ bacteria (without the F plasmid)⁹⁰. The cell-to-cell contact is initiated by the F pilus (TraA oligomer), which retracts to bring the two cells together (**Fig. 11**). Single-stranded plasmid DNA (ssDNA) is transferred by a relaxosome, composed of Tral, TraM, and TraY (and forming a DNA-protein complex), from the F⁺ bacterium to the F⁻ bacterium through an opening mediated by a Type IV Coupling Protein. The newly transferred ssDNA replicates through a rolling-circle mechanism, forming a double-stranded extrachromosomal plasmid DNA, and making the F⁻ cell now F⁺.

The F transfer region of the F plasmid consists of *tra* genes with 18 involved in the construction of the transferosome, involved in pilus synthesis and DNA transfer among other functions⁹¹. Eight of the *tra* gene products are widely conserved members of the diverse T4SS family. These include TraA, -B, -C, -E, -G, -K, -L, and -V. Another nine are involved in F-specific T4SS; these are TraF, -G, -H, -N, -U, -W, and TrbB, -C, and -I⁸⁵. Proteins that function in pilus assembly and extension are TraE, -L, -C -W, -F, and TrbC, -B.

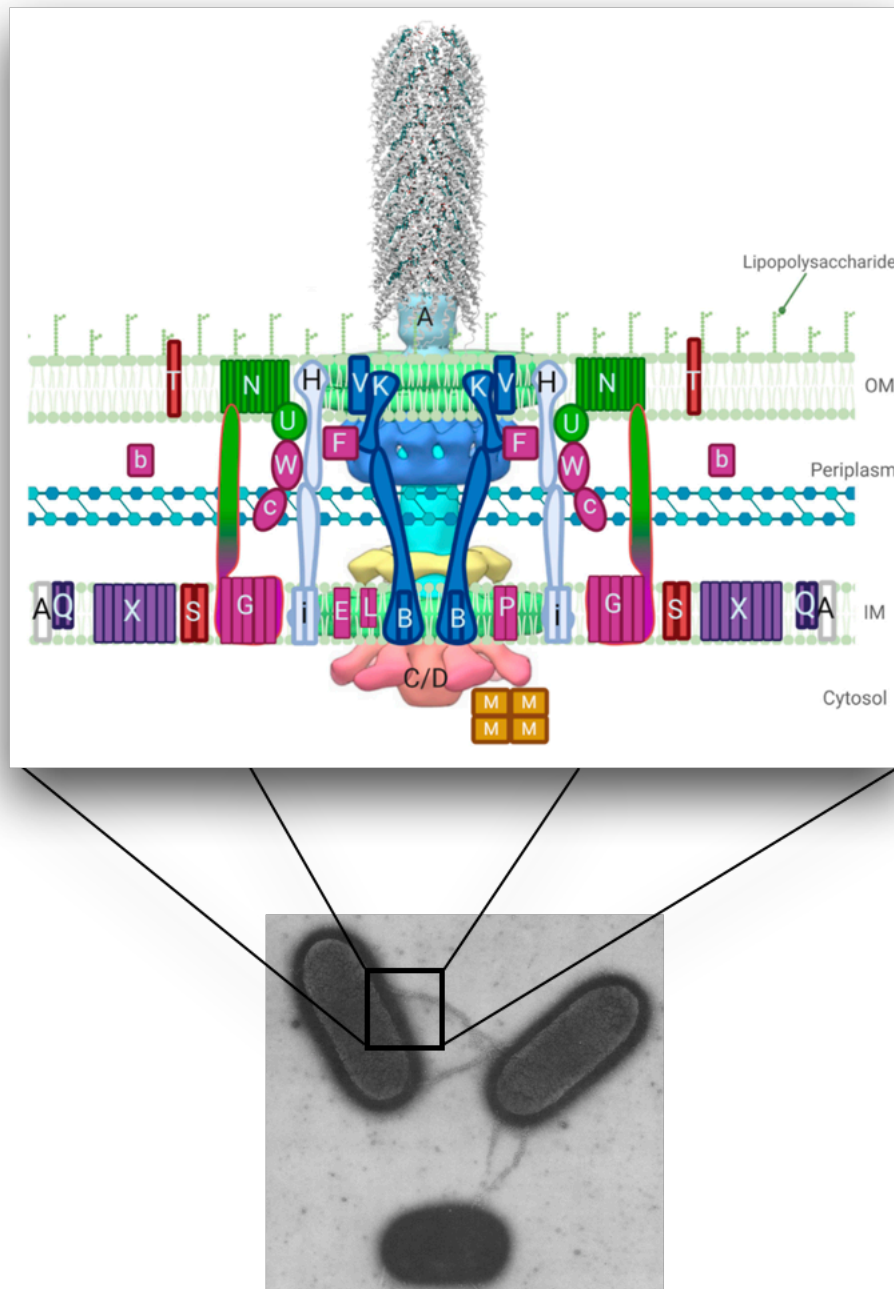


Figure 10. Model of the F-like type IV secretion system (T4SS) transferosome based on available structural information. Tra proteins are indicated with upper case letters, while Trb proteins are indicated with lower case letters. Proteins are coloured based on function; pilin (TraA; white), pilin processing (purple), pilus assembly/extension (fuchsia), core complex (dark blue), pilus retraction (light blue), mating pair stabilization (green), and entry exclusion (red). TrbB, TraF, and TraW (functional in pilus assembly/extension) are periplasmic proteins. TrbI (pilus retraction) is an integral inner-membrane protein. Top adapted from Bragagnolo et al. (2020)⁵. Electron micrograph (bottom) adapted from Curtiss et al. (1969)⁹².

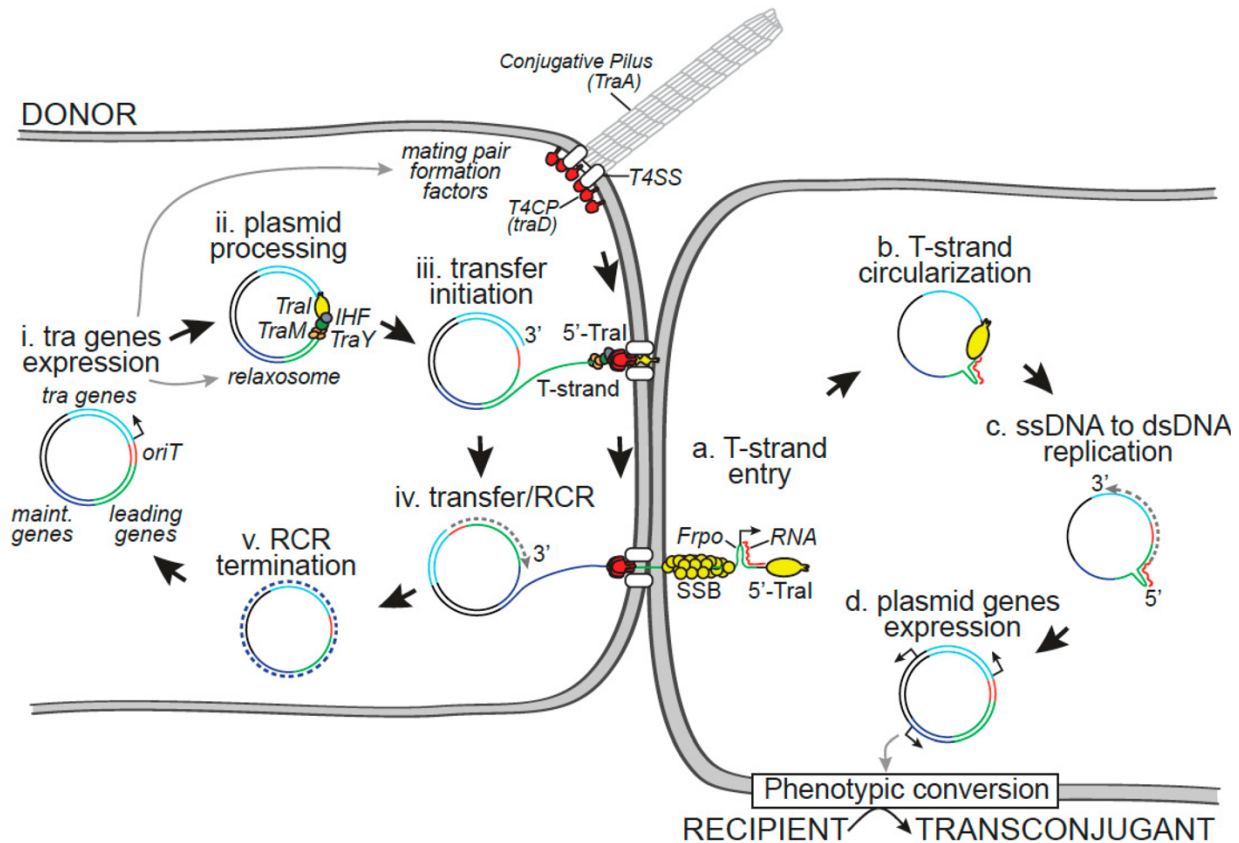


Figure 11. T4SS_F-mediated conjugation. The process is shown; (i) the expression of *tra* genes forms the relaxosome and transferosome (mating pair factors) formation, (ii) relaxosome-mediated plasmid processing, (iii) transfer initiation, (iv) transfer of ssDNA from the F⁻ bacterium to the F⁺ via a rolling circle replication, and (v) termination. The transferred ssDNA is then replicated to form a ssDNA plasmid. Figure adapted from Virolle et al. (2020)⁹⁰.

1.5.2. Pilus retraction in conjugation and the role of TrbI

Following cell-cell contact initiated by the F pilus, the cells are brought together by the retraction of the pilus to allow for plasmid ssDNA exchange through the pore of the Type IV Coupling Protein⁹⁰. Much is unknown about retraction, but previous research has established that it is energy-independent⁹³ and occurs at an average of 15.8 nm/s⁹⁴. Moreover, F pilus retraction has been shown to be regulated by two genes, *trbI* and *traH*⁵.

TrbI is an integral inner membrane-associated protein. Yeast-two hybrid analyses have shown that TrbI self-dimerizes at a hydrophobic segment that begins from W18 to V40 around the protein's N-terminus^{91,95}. This dimer directly interacts with a periplasmic T4SS_F protein, TraH,

at conserved residues, which in turn forms a network of interaction with other periplasmic proteins. Further, mutations in *trbI* were reported to have no effect on conjugative DNA transfer efficiency. Cells with mutated *trbI* were, instead, observed to express elongated pili. These mutational data have led to the hypothesis that TrbI is primarily involved in pilus retraction^{5,52}.

1.5.3. Pilus extension and the role of TrbB and TraF of the T4SS_F

In F-like Type IV Secretion Systems (T4SS_F), the importance of DIs is evident. As protein expression is an energy-expensive task, the conservation of proteins through an organism's evolutionary history is indicative of the paramount importance of those proteins. In T4SSs encoded by F-like plasmids (*i.e.*, F, R100, pSLT, pED208, and pYJ016), the *trbB* and *traF* genes are two of six genes that have no orthologues in P- or I-like subfamilies (**Figs. 8 and 9**), and they contain a thioredoxin and thioredoxin-like motif, respectively, known to be important in the maintenance of cellular redox balance through disulfide exchange and ensuring proper disulfide bond formation^{5,89,91,96–99}. The TraF thioredoxin-like domain does not contain the characteristic CXXC active site (where C is Cys and X is any other proteinogenic residue). TraF's function in T4SS_F pilus extension, independent of redox activity, remains unclear^{5,89,100}. On the other hand, TrbB contains an active thioredoxin domain having amino acid sequence homology to the thioredoxin superfamily which consists of a minimum of three α -helices flanking a four-stranded antiparallel β -sheet, and a CXXC active site^{89,91,100,101}.

TrbB has been shown to function as a DI *in vitro* and *in vivo*^{89,91,96,102}, suggesting that TrbB may ensure proper protein folding of T4SS_F proteins. Several T4SS_F proteins are remarkable for their high cysteine content and these proteins are putative client proteins for TrbB; TraN, TraU, and TraH each have 22, 11, and 6 cysteines, respectively^{91,102}. Conjugative T4SSs express TrbB alone, or TraF/TrbB; non-redox active TraF alone is not observed^{5,89}. For instance, the F and pED208 plasmids contain both TraF and TrbB while the R27 has only a redox active TraF, also suggesting different roles for TraF and TrbB within the T4SS_F conjugative apparatus.

1.6. Research Significance & Objectives

Solving a high-resolution 3D structure of TrbB remains an ongoing effort and will provide high-resolution insights into the structure of TrbB. A high-resolution structure can lead to two applications. Firstly, it can provide overt direction for the design of novel strategies to mitigate T4SS_F-mediated conjugation by inhibiting the protein central to its stability. Secondly, it can provide a fundamental understanding of bacterial phylogenetics and characterize a protein with a thioredoxin fold that both prokaryotes & eukaryotes utilize for proteome stabilizing mechanisms.

Here, we report findings that advance our structural and functional insights into TrbB, providing guidance for high-resolution structure solution, but also in and of itself provide novel insights into TrbB's physicochemical stability, disulfide isomerase activity, solution dynamics, and its interaction with other T4SS_F proteins. We also compare TrbB to cognate T4SS_F proteins TrbI's and TraF and offer preliminary insights into TrbI's chemical space of crystallization. The following research questions are answered by my research:

- what are some of the structural features of TrbI, TraF, and TrbB and how are they similar or different (Sections 3.3, 3.5)
- how does TrbB_{WT}'s structure make it recalcitrant to high-resolution structural analyses (Sections 3.3, 3.4), and
- can we design a protein construct that is more amenable to high-resolution structural analyses (Section 3.4)
- what conditions can facilitate the TrbI and TrbB to crystallize (Section 3.2)
- does TrbB function as a disulfide isomerase enzyme in vitro (Section 3.6), and
- what domain of TrbB is necessary and sufficient for its enzymatic function (Section 3.6)
- does TrbB bind other T4SS_F proteins (Section 3.6)
- does TrbB function as a chaperone for the T4SS_F (Section 3.6)

2. CHAPTER TWO: MATERIALS & METHODS

2.1. Cloning *trbI* & *trbB* into pGEX-4T-2

2.1.1. pGEX4-T-2 plasmid extraction from *E. coli* DH5a

An overnight (O/N) culture of pGEX-4T-2::amp *E. coli* strain DH5 α was prepared by inoculating 5 mL of LB supplemented with 100 μ g/mL of Ampicillin (Amp). The mixture was incubated at 37°C with shaking at 200 rpm overnight (~16 h). Following overnight incubation and storing of glycerol stock (500 μ L O/N culture and 500 μ L 50% glycerol) in -80°C, pGEX-4T-2 plasmid was extracted from the pelleted cells using GeneJET Plasmid Miniprep Kit (ThermoFisher) following the manufacturer's protocol. Briefly, cells were resuspended in 250 μ L of Resuspension solution containing 2 mg/mL lysozyme and 10 mg/mL RNase A to weaken the cell wall and degrade RNA, respectively. The cells were then lysed using 250 μ L of Lysis solution containing SDS and NaOH, then neutralized using 350 μ L Neutralizing solution for the optimal binding of the plasmid on the silica membrane in the spin column. After a 5 min centrifugation, the supernatant was transferred to a GeneJET spin column. The column was centrifuged for 1 min before 500 μ L of Wash solution (96% ethanol) was added. The column was centrifuged for another 1 min. Adding of Wash solution and subsequent centrifugation was performed twice. Plasmid DNA was eluted with 30-50 μ L of elution buffer (10 mM Tris-HCl pH 8.5) from the GeneJET spin column in one or two centrifugation steps. DNA concentration was determined using NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fisher). All centrifugations were performed at 13200 x g.

2.1.2. Phenol: chloroform plasmid extraction from *E. coli* XK1200

An O/N culture of *E. coli* strain XK1200 was prepared by inoculating 5 mL of LB supplemented with 100 μ g/mL Kanamycin (Km). The mixture was incubated at 37°C with shaking at 200 rpm overnight (~16 h). Following O/N incubation, and storing of glycerol stock, the O/N culture was centrifuged (5000 x g, 5 min, 4°C) to pellet the cells.

The T4SS_F-encoding pOX38 plasmid¹⁰³ was extracted from the pelleted cell culture using phenol: chloroform extraction. First, pelleted cells were resuspended in 150 μ L solution 1 (50 mM Dextrose, 25 mM Tris pH 8.0, and 10 mM EDTA). Next, cells were lysed using 300 μ L solution 2

(0.2 M NaOH and 1% SDS). Resulting lysate was then neutralized using 225 μ L of 3M sodium acetate (NaOAc) pH 4.8 and incubated on ice for 5 min. The principle of the above steps is like that of the GeneJET Plasmid Miniprep (Thermo Fisher).

Subsequently, the mixture was centrifuged for 10 min at 4°C. The supernatant was transferred to a new tube prior to the addition of 450 μ L of 1:1 phenol: chloroform. The solution was vortexed for 10 s and subsequently centrifuged for 2 min. The top layer of the resulting suspension was isolated and 450 μ L of chloroform was added before vortexing (10s) and centrifugation (2 min) as before. The top layer was again transferred to a clean tube, prior to adding 1 mL of 95% ethanol pre-chilled at -20°C. The solution was incubated in dry ice for 5 min. The tube was centrifuged for 10 min at 4°C and the ethanol was decanted. The remaining pellet was allowed to air dry before it was dissolved in 30 μ L of 10 mM Tris pH 8.0.

The pellet was then added with 0.5 μ L 5 mg/mL RNase A and incubated at 37°C for 5 min to degrade RNA. The volume was brought to 400 μ L with solution 1 (10 mM Tris pH 8, 1 mM EDTA) prior to the addition of 10 μ L 4 M ammonium acetate (NH₄OAc), which selectively precipitates DNA. 1 mL of 95% ethanol was again added to precipitate DNA, centrifuged (10 min, 4°C) and the pellet was resuspended in 30 μ L of 10 mM Tris pH 8.0. Resulting pOX38 plasmid DNA was stored in -20°C after determining the concentration using NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fisher). All centrifugations were performed at 13200 x g.

2.1.3. Polymerase Chain Reaction (PCR) amplification

Primer Design. Primers were designed according to the following guidelines by Addgene¹⁰⁴. The GC content of the primers was designed to be in the range of 40-60%, the primers were designed to have melting temperatures (T_M) within 55-65°C, and the forward and reverse primers were designed to have a T_M difference within 5°C. Table 1 provides details on the primers used. Interfering interactions (*i.e.*, formation of hairpins & primer dimers) were also minimized using OligoAnalyzer by Integrated DNA Technologies (IDT). The primers were made longer, relative to the 18-25 nt guideline, to reduce non-specific binding of the primers.

Table 1. Primers used for PCR.

	Sequence (5'-3')	T _M (°C)	# nt
<i>trbI</i> _{WT} Forward	GTACTGAATTCacATGAGTTCAACGCAGAAACCCGC	63	36
<i>trbI</i> _{WT} Reverse	GTCTTGTCGACTCATGGTCCGCCCTCATTCGC	62	33
<i>trbB</i> _{WT} Forward	GTACTGAATTCatgtctctcactaaatcactgctgttcaccc	63	42
<i>trbB</i> _{Δ168} Forward	GTACTGAATTCaatgggtccgtctcagtaatggcagg	63	37
<i>trbB</i> _{WT} & <i>trbB</i> _{Δ168} Reverse	GTCTTGTCGACTtatttcgcacctttttctccgtacatctgc	63	45

Restriction endonuclease cleavage sites are underlined; *EcoRI* in forward primers and *Sall* in reverse.

PCR on pOX38 to amplify *trbB* or *trbI*. The PCR reaction was set up by using 250 ng of pOX38 DNA. The reaction was set to a total of 50 µL by adding 10 µL of 5X Phusion HF buffer, 1 µL 10 mM dNTPs, 1.5 mL DMSO (final conc. of 3%), 0.5 µL of DNA polymerase (Phusion), and 32.5 µL ddH₂O. 1 µL of 20 mM forward or reverse primer was added to the reaction. The thermocycler was set to an annealing temperature of 53°C using the first set of primers and 61°C for the second set of primers, both for 10 s. The annealing temperature was set to 67°C or 69°C for the third set following the suggestion of NEB T_M calculator. Additionally, only the hybridization sequence was considered in determining the T_M. Table 2 details the PCR parameters used.

Table 2. Thermocycler settings for PCR amplification.

Cycle step	Temperature (°C)	Time	No. of cycles
Initial denaturation	98	30 seconds	1
Denaturation	98	10 seconds	30
Annealing	60* or 59 [§]	20 seconds	30
Extension	72	10 seconds	30
Final Extension	72	10 minutes	1
Hold	4		1

*for *trbB*_{WT} & *trbB*_{Δ168}; [§]for *trbI*_{WT}

2.1.4. PCR Purification of PCR amplicon

A 1:1 volume of PCR product to binding buffer (containing guanidium hydrochloride, a protein denaturant) mixture was transferred to a GeneJET spin column. Samples were centrifuged for 1 min at 13200 x g, washed with 700 µL wash buffer and centrifuged again (using the same settings). Before eluting with 30-50 µL of elution buffer (10 mM Tris-HCl pH 8.5), the column was centrifuged to assure that all residual flow-through were removed. DNA concentration was determined using NanoDrop 2000 UV-Vis Spectrometer (Thermo Fisher).

2.1.5. Restriction endonuclease (RE) Digestion

The PCR amplicon and pGEX-4T-2 vector plasmid were both double-digested with 20,000 units *EcoRI* and 20,000 units *Sall*. One µg of DNA was used in the reaction. Insert and vector DNA were double digested with the same REs to facilitate ligation. 5 µL of 10X NEB Buffer 2.1 and ddH₂O were added to bring the total volume to 50 µL. Double digested pGEX-4T-2 was treated with 20,000 units calf intestinal phosphatase (CIP) to reduce self-ligation of the plasmid, added simultaneously with the other components. The reaction was incubated at 37°C overnight to ensure complete DNA digestion.

2.1.6. DNA Gel Extraction

The appropriately sized DNA band was excised from the agarose gel using a razor blade. 1:1 (volume: weight) of Binding Buffer was added to the 1.5 mL microtube containing the gel slice. The mixture was incubated at 55- 60°C to melt the agarose gel. 1 gel volume of 100% isopropanol was added to the mixture prior to transferring to a GeneJET column. The column was added 700 µL of Wash Buffer. Following two cycles of centrifugation (1 min, 13200 x g) and discarding of flow-through, DNA was eluted from the column with 30 -50 µL of Elution Buffer (10 mM Tris-HCl, pH 8.5).

2.1.7. Ligation *in vitro*

The following insert to vector molar ratios were used: 3:1, 5:1, and 7:1. The vector DNA mass used for all experiments was 75 ng. Insert DNA mass was determined using NEBioCalculator®; the formula is:

$$\left(\frac{mol_{insert}}{mol_{vector}}\right) * mass_{vector} * \left(\frac{length_{insert}}{length_{vector}}\right)$$

T4 DNA ligase was used for the *in vitro* ligation. Reactions were overnight at 4°C. Subsequently, the reaction was heat inactivated at 65°C for 20 min.

2.1.8. Transformation via heat shock & CaCl₂

The ligation solution was mixed with 100 µL chemically competent DH5α (for high-copy number replication of construct plasmid) or BL21(DE3) (for large-scale expression). The mixture was

incubated on ice for 20 min, at 42°C for 90 s, and again on ice for 5 min. 900 µL of LB was added and the mixture was incubated at 37°C for 1 h. Cells were then pelleted at 5000 x g for 5 min and 900 µL of LB was withdrawn. Resuspended cells were subsequently plated on LB-Agar plates supplemented with 100 µg/mL of ampicillin and incubated overnight at 37°C. To verify transformation of the correct plasmid construct, two steps were performed. First, transformants were grown in LB supplemented with 100 µg/mL ampicillin O/N at 37°C and plasmids were mini-prepped from the cells, double-digested with 20,000 units *EcoRI* and *Sall*, and double-digested DNA were electrophoresed in 1.2% agarose. Second, transformants were further picked and streaked onto another LB-Agar-Amp plate and were allowed to grow overnight at 37°C. Proliferated colonies were analyzed using colony PCR to verify presence of insert DNA.

2.1.9. DNA Detection and Verification

Detection by agarose gel electrophoresis. DNA samples intercalated by Ethidium Bromide (EtBr) were electrophoresed in 1.2% agarose at 100 V for 30-42 min following: (1) PCR amplification (to qualify PCR amplicon); (2) Restriction Endonuclease Digestion of pGEX-4T-2 and *trbB* or *trbI*; (3) double-digestion of plasmid constructs from *E. coli* transformants, and (4) colony PCR. Agarose gels were prepared by mixing 1.2% (w/v) of agarose powder into 1X Tris-Acetate-EDTA buffer.

Verification and detection pre-DNA sequencing. Following observation of *E. coli* colonies on LB-Agar-Amp plates, these colonies were (1) picked and grown in liquid LB supplemented with 100 µg/mL ampicillin overnight at 37°C at 200 rpm. Plasmids were mini-prepped from the grown colonies and double-digested with 20,000 units *EcoRI/Sall* for at least 3 h at 37°C, intercalated with EtBr and electrophoresed in 1.2% agarose. (2) Colonies were also streaked onto another fresh LB-Agar-Amp plate, grown overnight at 37°C, and were analyzed using colony PCR. Amplicons were intercalated with EtBr and electrophoresed in 1.2% agarose for detection.

DNA Sequencing. Following the two-fold verification, and observation of the gene in the agarose gels, construct plasmids were sent to The Centre for Applied Genomics (Toronto, ON) for sequencing using the primers 5'-GGGCTGGCAAGCCACGTTTGGTG-3' and 5'-

CCGGGAGCTGCATGTGTCAGAGG-3'. The institution's plasmid preparation instructions were followed; *i.e.*, providing 200-300 ng of plasmid DNA in 7 μ L. DNA concentrations were quantified using NanoDrop2000 spectrophotometer (Thermo Fisher). DNA sequencing data was verified against the known sequence of *trbI* or *trbB* using Nucleotide BLAST (National Center for Biotechnology Information)¹⁰⁵.

2.2. Expressing TrbI & TrbB in *E. coli* BL21(DE3)

2.2.1. Large-scale expression using 1 mM IPTG

A litre of sterile LB supplemented with 100 μ g/mL ampicillin and 1 mM dextrose was inoculated with *E. coli* BL21(DE3) containing pGEX4T2::*trbI* or pGEX4T2::*trbB* (transformant colony grown in LB O/N at 37°C at 200 rpm). Once the OD₆₀₀ ~0.4-0.7 at the mid-log growth phase, expression was induced by the addition of 1 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG). The cells were incubated at 16-18°C at 200 rpm overnight (~16 h). Cells were then pelleted at 5000 x g, 4°C for 40 min.

2.2.2. Cell lysis by sonication

Cells were re-suspended in Lysis/Loading Buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA) supplemented with protease inhibitors (1.25 mM phenylmethylsulfonyl fluoride, 5 mM benzamidinium HCl, 5 mM aminocaproic acid). To liberate proteins, cells were sonicated at 25% amplitude for a total of 5 min, in a cycle of 15s on pulse and 30s pulse off. The lysate was centrifuged at 25000 x g for 40 min at 4°C to separate the soluble fraction (the supernatant) from other cellular debris (pellet).

2.3. Protein purification

2.3.1. Affinity GST Sepharose Fast Protein Liquid Chromatography Purification

Glutathione S-Transferase (GST)-affinity chromatography was performed using an ÄKTA Purifier 10S Fast Protein Liquid Chromatography (FPLC) system. The column's stationary phase was GST Sepharose beads. Lysis/loading buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA) was used

to wash the column of non-column-binding proteins. GST fusion protein was eluted using an Elution buffer (20 mM Tris pH 8.0, 20 mM glutathione).

2.3.2. Thrombin cleavage

GST-tagged protein samples were incubated with Thrombin (5-10 units per mg of fusion protein) of in Phosphate Buffered Saline (PBS) at 4°C overnight, with no rocking to avoid protein aggregation.

2.3.3. Secondary Purification using Size-Exclusion Chromatography

To isolate the untagged proteins, secondary purification using Size-Exclusion Chromatography (SEC) through an ÄKTA Purifier 10S FPLC system (GE Healthcare) was also performed using HiPrep 16/60 Sephacryl S-100 HR (Cytiva).

2.3.4. Protein Qualification by SDS-PAGE

Samples were verified in 12.5% polyacrylamide separating gel and 4% polyacrylamide stacking gel. Prior to loading into the gel lanes, protein samples were mixed with 1X SDS buffer solution (160 mM Tris pH 6.8, 4% SDS, 20% glycerol, 0.0012% bromophenol blue, 20% β -mercaptoethanol) and boiled at 95°C for 5-10 min to denature the proteins. Samples were electrophoresed at 180 V for 60-65 min. Gels were then stained with Coomassie Blue solution, washed with ddH₂O, and rocked with a destain solution (80% v/v methanol, 20% v/v acetic acid).

2.3.5. Buffer exchange and Quantification

After qualification by Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) with Coomassie solution staining, relevant fractions were pooled, and buffer exchanged into the relevant buffer solution to (a) remove the GSH and (b) prepare for the downstream analyses. Buffer exchange was performed using: [1] a 10 or 30 kDa MWCO centrifugal concentrator column (Corning, Millipore) at 3500 x g at 4°C; [2] dialysis tube in 4 L buffer solution for 3h at 4°C and then in a 4 L buffer solution overnight 4°C.

2.3.6. Protein Quantification

Protein concentrations were routinely quantified by using [1] a colorimetric method using bicinchoninic acid (BCA), following the manufacturer's (Thermo Scientific) protocol; or [2] absorbance of the protein solution at 280 nm ($A_{280\text{ nm}}$) and using the Beer-Lambert law ($A = \epsilon Cl$) and solving for C . Absorbance measurements, corrected for background buffer solution absorbance, were determined using DU730 UV-Vis Spectrophotometer (Beckman Coulter) or using NanoDrop 2000 Spectrophotometer (Thermo Fisher).

2.4. Crystallization trials

2.4.1. General protein preparation

Purified GST-TrbI or GST-TrbB was concentrated to 5.5, 6.7, or 10 mg/mL and buffer exchanged into 10 mM HEPES pH 7.5 or pH 7.14 (the isoelectric point of GST-TrbI), 50 mM NaCl, 5% (v/v) glycerol or 10mM MES pH 6.72, 50 mM NaCl, 5% (v/v) glycerol prior to crystallization trials.

2.4.2. Crystallization trials using the vapour diffusion method

A 1 μL droplet of a protein sample (GST-TrbI or GST-TrbB) was transferred onto the sitting drop wells of a 96 well plate or a cover slip (for hanging drop method). A mixture of buffer/precipitant/additive at a volume of 90 μL (for sitting drop) or 1000 μL (for hanging drop) was added to the reservoir well, and 1 μL of the buffer from the reservoir was mixed with the protein in the corresponding sitting well or cover slip. Plates for 96-well sitting drop crystallization trials were sealed with tape to prevent equilibration with the atmosphere and allow vapor diffusion to occur exclusively within wells. Each well of 24-well hanging drop crystallization trial plates were sealed with the cover slip bearing the protein/reservoir buffer sample, with the drop being inverted. Vacuum seal grease was used to seal the wells, preventing each well's exposure to the external atmosphere. Plates were incubated at room temperature or at 4°C. Crystal growth was monitored periodically using a brightfield microscope.

2.4.3. Protein preparation for the Hauptman-Woodward Institute high-throughput crystallization screening

GST-TrbI and GST-TrbB at a concentration of 10 mg/mL in 10 mM MES pH 6.72, 50 mM NaCl, 5% (v/v) glycerol was sent to the Hauptman-Woodward Institute (Buffalo, NY, USA) for an extensive screening against 1,536 buffer-precipitant conditions utilizing microbatch-under-oil method. Brightfield (visible light) images, Ultraviolet Two-Photon Excited Fluorescence (UV-TPEF) and Second Harmonic Generation (SHG) signals were analyzed using the institute's software package MARCO Polo¹⁰⁶.

2.5. Assaying the Protein disulfide Isomerase (PDI) activity of TrbB

2.5.1. Protein preparation

TrbB_{WT} and TrbB₅₇₋₁₈₁ were buffer exchanged into the kit's PDI assay buffer supplied by the manufacturer prior to analyses, using a 5 kDa MWCO concentrator column (Sartorius).

2.5.2. Assay specifications

PDI activity was assessed using a fluorometric protein disulfide isomerase quenched-fluorophore kit from Abcam (Boston, USA; cat ab273337). Signal detection was performed using Synergy H4 Microplate reader (Agilent BioTek) with excitation and emission wavelengths of 490 nm and 580 nm, respectively, at 25°C. Data was collected in triplicate (n = 3) from two independently expressed and purified protein samples with final concentrations 50 µM in each well.

2.5.3. Statistical Analysis

T-test comparing 50 µM Bovine Serum Albumin (BSA) and 50 µM TrbB was performed using GraphPad Prism for macOS v.9.1.1.

2.6. Circular Dichroism (CD) Spectroscopy

2.6.1. Estimating Protein 2° structures using CD Spectroscopy

CD spectra were acquired between 200-260 nm using a Jasco J810 CD spectrometer equipped with a six-position temperature-regulated cell holder with measurements at 22°C and a protein

concentration of 5 μ M in 10 mM KH_2PO_4 pH 7.5, 100 mM KF, 5% (v/v) glycerol was included to minimize noise¹⁰⁷. Following acquisition, CD spectra were input into BeStSel¹⁰⁸ for secondary structure content prediction. Input parameters included units of measured ellipticity (mdeg) at 5 μ M, entering the number of residues (TrbB₅₇₋₁₈₁: 125, TrbB_{WT}: 181, GST-TrbB: 410, His₆TraW: 199, GST-TrbB/His₆TraW mixture: 610, GST-TrbI: 358, GST-TraF: 486), and the pathlength (0.1 cm). CD spectra is presented as mean residue ellipticity (MRE) computed using the equation below; observed CD signal (mdegrees) multiplied by mean residue weight (MRW; protein molecular weight normalized by the number of residues minus 1), normalized by the pathlength (0.1 cm), concentration (g/mL), corrected by a factor of 10.

$$MRE = \frac{MRW * \theta_{observed}}{10 * l * c}$$

2.6.2. Thermal denaturation

Thermal denaturation measurements were performed by sampling from a single wavelength (222 nm for α -helices) as a function of temperature (30-90°C) at a rate of 1°C/min. Experiments were performed in triplicate.

2.7. Modelling using ColabFold-AlphaFold2 and PyMOL

2.7.1. ColabFold-AlphaFold2 modelling using amino acid sequence

Computational 3D protein structure models were generated using the primary sequence of the relevant protein as input into ColabFold-AlphaFold2 (CF-AF2)¹⁰⁹ site (<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>) using default settings. These default settings include MSA mode (set to mmseqs2_uniref_env), pair mode (set to unpaired_paired), model type (set to alphafold2_ptm or alphafold2_multimer_v3, used to model heterodimeric binding TrbB/TraW, TrbB/TraU), number of recycles was set to 3, recycle_early_stop_tolerance (auto), pairing strategy (greedy); sample settings included max msa (auto) and number of seeds (1). Predicted structure models were visualized in PyMOL v2.5.2 (Schrödinger Inc.). Indicated confidence metric, in the range [0,1], is given by $0.8 * ipTM + 0.2 * pTM$, weighing the metric for the reliability of binding interfaces (interface predicted Template Modelling score, ipTM) more compared to pTM score¹¹⁰.

2.7.2. Analysis using PyMOL v. 2.5.2

3D modelling and basic structural analyses were performed using PyMOL v2.5.2 (Schrödinger, Inc.)¹¹¹.

2.8. Size-Exclusion Chromatography Multi-Angle-Light Scattering Small-Angle X-Ray Scattering

2.8.1. Protein preparation

GST-TrbB was prepared without GST cleavage as per Sections 2.2-2.3 and was buffer exchanged into a 10X dilution of a buffer stock (20 mM HEPES pH 7.0, 100 mM NaCl, 5% (v/v) glycerol, 0.05% NP40) using a 30 kDa MWCO concentrator column. The 5.5 mg/mL GST-TrbB sample (as quantified via the Edelhoch method) and the matching buffer were sent to the BioCAT facility where the buffer was diluted 10X and used as the running buffer for the Size-Exclusion Chromatography Multi-Angle-Light Scattering Small-Angle X-Ray Scattering (SEC-MALS-SAXS) experiment.

2.8.2. Data collection

SEC-MALS-SAXS data for GST-TrbB was collected at the BioCAT 18ID beamline (Advanced Photon Source, Argonne National Laboratory, USA). Additional purification was performed by the facility; in-line size-exclusion chromatography (SEC) SAS was employed by injecting 350 μ L of the protein sample into a Superdex 200 Increase 10/300 SEC column at 0.6 mL/min, and the sample underwent sequential multi-angle light scattering (MALS) analysis using a Wyatt DAWN Heleos II MALS system. Briefly, data acquisition was acquired using an Eiger2 XE 9M detector at a sample-detector distance of 3.7 m and at $\lambda = 0.1033$ nm.

2.8.3. Data processing

Data processing was performed using BioXTAS RAW v.2.1.1 and ATSAS packages^{112,113}. *Ab initio* reconstruction was performed using DAMMIF with DAMAVER averaging and refinement using ColabFold-AlphaFold2 (CF-AF2) homodimeric GST-TrbB model, and clustering using DAMCLUST.

Bead model was generated using PyMOL v2.5.2 (Schrödinger Inc.)¹¹¹; correcting bead radius to 5.4.

2.9. Bio-Layer Interferometry

2.9.1. Protein preparation

Protein samples (GST-TrbI and TrbB) were expressed and purified according to Section 3.2 and 3.3, buffer exchanged to Phosphate Buffered Saline (PBS) using 10kDa MWCO centrifugal concentrator columns (Corning, Millipore) for Bio-layer Interferometry (BLI) analysis.

2.9.2. Assay specifics and setup

Streptavidin biosensor tips (Sartorius) were loaded with biotinylated GST-TrbI at a concentration of 10 μ M. The tips were then washed with PBS, associated with 10 μ M TrbB, and disassociated in PBS. The tips were regenerated using 10 mM glycine, pH 1.6. The assay protocol was as follows: equilibration (60 seconds), immobilization (600 seconds), baseline (120 seconds), association (300 seconds), dissociation (300 seconds), regeneration and neutralization (30 seconds). The experiment was performed using Octet BLI Discovery Software on an Octet R4 (Sartorius).

2.10. Analytical Size-Exclusion Chromatography

TrbB_{WT}, TrbB₅₇₋₁₈₁, His₆TraW, and His₆ Δ TraW at 19 μ M in 20 mM HEPES pH 7.5, 150 mM NaCl, with or without 2 mM DTT, were analyzed using a Zenix SEC-150 column (Sepax Tech. Inc.) at a rate of 1.0 mL/min in an H-class Ultra Performance Liquid Chromatography (UPLC) system (Waters Acquity) with A₂₈₀ sample detection. The column was washed with 1 column volume of buffer (15 mL of 20 mM HEPES pH 7.5, 150 mM NaCl) prior to every protein analysis. The TrbB/His₆TraW, TrbB/His₆ Δ TraW, TrbB₅₇₋₁₈₁/His₆TraW mixtures were equilibrated on ice for at least 30 min prior to SEC analysis to allow for any complex formation.

2.11. ^1H - ^{15}N Heteronuclear Single Quantum Coherence Nuclear Magnetic Resonance Spectroscopy

2.11.1. Protein preparation

TrbB₅₇₋₁₈₁ and TrbB_{WT} were expressed from *E. coli* BL21(DE3) grown in M9 minimal media (6 g of Na₂HPO₄, 3 g of KH₂PO₄, 1 g of $^{15}\text{NH}_4\text{Cl}$, 0.5 g of NaCl, and 10 g of glucose in 1 L of water supplemented with 1 mM CaCl₂, 1 mM MgSO₄, 50 µg/mL kanamycin, and a trace mineral mix). Large-scale protein expression and purification was performed as described in Sections 2.2-2.3.

2.11.2. Experiment details and HSQC parameters

Purified protein samples were concentrated to 0.2 mM and supplemented with 10% (v/v) D₂O in 10 mM HEPES pH 6.0, 50 mM NaCl. HSQC spectra were acquired on a Bruker DRX 600 NMR spectrometer operating at a ^1H frequency of 599.80 MHz at 21°C.

2.12. Figure Processing and Statistical T-test

Figure preparation, and statistical T-test comparison involving PDI assay data, were performed using GraphPad Prism for macOS v9.1.1 or v9.5.1.

3. CHAPTER THREE: RESULTS & DISCUSSION

3.1. Cloning, Expression, and Purification of GST-TrbB and GST-TrbI

3.1.1. Cloning

The genes *trbB*_{WT}, *trbB*_{Δ168} and *trbI*_{WT} were PCR-amplified from the T4SS_F-containing pOX38 plasmid¹⁰³ using designed primers (**Table 1**). These genes were inserted into an *EcoRI/SalI* double-digested vector plasmid, pGEX-4T-2, and transformed into *E. coli* DH5α for high copy number replication. Following ligation-dependent cloning, three-fold verifications were performed. First, colonies grown overnight at 37°C in LB-Agar petri dishes supplemented with 100 µg/mL ampicillin were picked and subjected to colony PCR. The size of *trbI*, *trbB*_{WT}, and *trbB*_{Δ168} are 387 bp, 546 bp, 378 bp, respectively, and the signals for these are evident in the agarose gels (**Figs. 12A, 13A, 14A**), providing initial assurance that the genes were successfully inserted into plasmids and these DNA constructs are intact in the picked colonies. Second, the colonies which yielded a positive colony PCR result were grown overnight in LB at 37°C, mini-prepped to purify construct plasmid DNA (pGEX-4T-2::*trbI*, pGEX-4T-2::*trbB*_{WT}, and pGEX-4T-2::*trbB*_{Δ168}), double-digested with *EcoRI/SalI*, electrophoresed in 1.2% agarose stained with ethidium bromide (EtBr), and visualized under UV light. Evidence of faint bands about 400 bp are shown (**Figs. 12B, 13B, 14B**), further supporting successful cloning. Lastly, purified construct plasmid DNA were sent to The Centre for Applied Genomics (TCAG; The Hospital for Sick Children) for Sanger DNA sequencing (data not shown) definitively confirming successful cloning.

3.1.2. Expression and Purification

Construct plasmid DNA (pGEX-4T-2::*trbI*, pGEX-4T-2::*trbB*_{WT}, and pGEX-4T-2::*trbB*_{Δ168}) were transformed into *E. coli* BL21(DE3) for large-scale expression overnight at 18°C, induced by 1 mM IPTG. Following cell lysis by sonication (25% amplitude), and centrifugation (25000 x g, 40min, 4°C) to separate soluble proteins and insoluble cellular debris, water-soluble proteins were purified using an Affinity GST-Sepharose column on an ÄKTA Purifier 10S FPLC system (chromatograms shown in **Figs. 12C, 13C, 14C**) and eluted GST-TrbI, GST-TrbB, and GST-TrbB₅₇₋₁₈₁ were qualified by SDS-PAGE (**Figs. 12D, 13D, 14D**) to verify protein identity (based on MW) and

relative purity. To purify untagged TrbB_{WT} and TrbB₅₇₋₁₈₁ following GST-tag cleavage (5-10 units Thrombin/mg fusion protein, 4°C, overnight), secondary Size-exclusion chromatography (SEC) purifications (**Figs. 13E, 14E**), and SDS-PAGE were performed to ascertain which SEC fractions the protein-of-interest eluted (**Figs. 13F, 14F**).

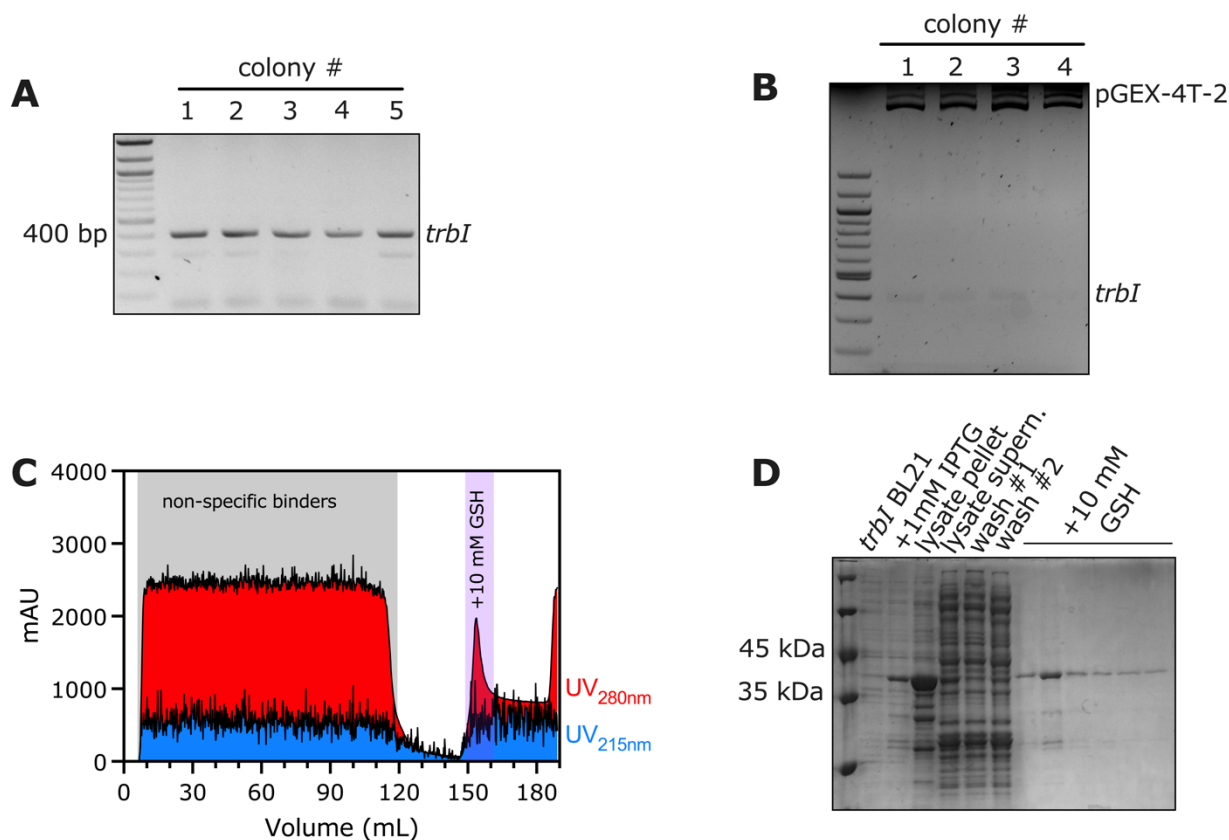


Figure 12. Representative cloning, expression, and purification of GST-TrbI. (A) Colony PCR verification of pGEX-4T-2::*trbI* BL21(DE3) colonies. Colonies were picked from LB-Agar plates with 100 ug/mL Ampicillin, PCR amplified, and electrophoresed in 1.2% agarose gel with EtBr, and visualized under UV light. (B) Colonies which yielded a positive colony PCR result were grown in liquid LB for 16 h, mini-prepped, double-digested with *EcoRI* and *SaI* for 3 h, and electrophoresed in agarose gel for detection under UV light. Faint signals about 400 bp are evident for all four colonies, while signals corresponding to pGEX-4T-2 at about 5000 bp are strongly evident. Multiple pGEX-4T-2 bands in each lane corresponds to the different topological forms of the plasmid. (C) Primary GST affinity purification following large-scale expression and (D) SDS-PAGE qualification to assess for protein identity and relative purity.

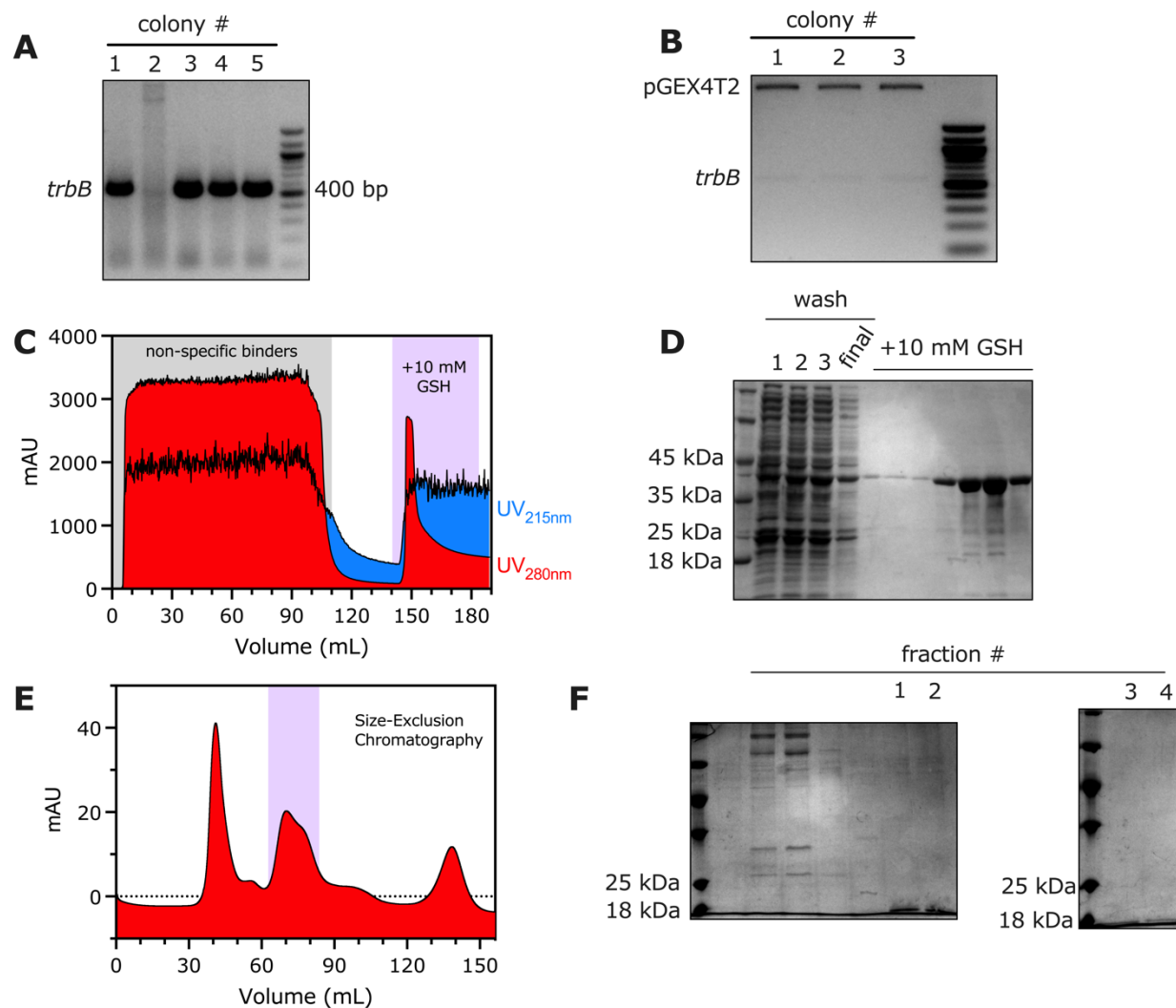


Figure 13. Representative cloning, expression, and purification of GST-TrbB and TrbB_{WT}. (A) Colony PCR verification of pGEX-4T-2::*trbB*_{WT} BL21(DE3) colonies. Colonies were picked from LB-Agar plates with 100 ug/mL Ampicillin, PCR amplified, electrophoresed in 1.2% agarose gel with EtBr, and visualized under UV light. (B) Colonies which yielded a positive colony PCR result (lanes 1, 3-5) were grown in liquid LB for 16 h, mini-prepped, double-digested with *Eco*RI and *Sal*I for 3 h, and electrophoresed in agarose gel for detection under UV light. Faint signals about 400 bp are evident for all four colonies, while signals corresponding to pGEX-4T-2 at about 5000 bp are strongly evident. (C) Primary GST affinity purification following large-scale expression and (D) SDS-PAGE qualification to assess for protein identity and relative purity. (E) Secondary Size-exclusion chromatography purification of the untagged protein following GST-tag cleavage. Peaks were verified for quality by SDS-PAGE and the pure protein of interest from the purple peak of the chromatogram is labeled 1-4 (F).

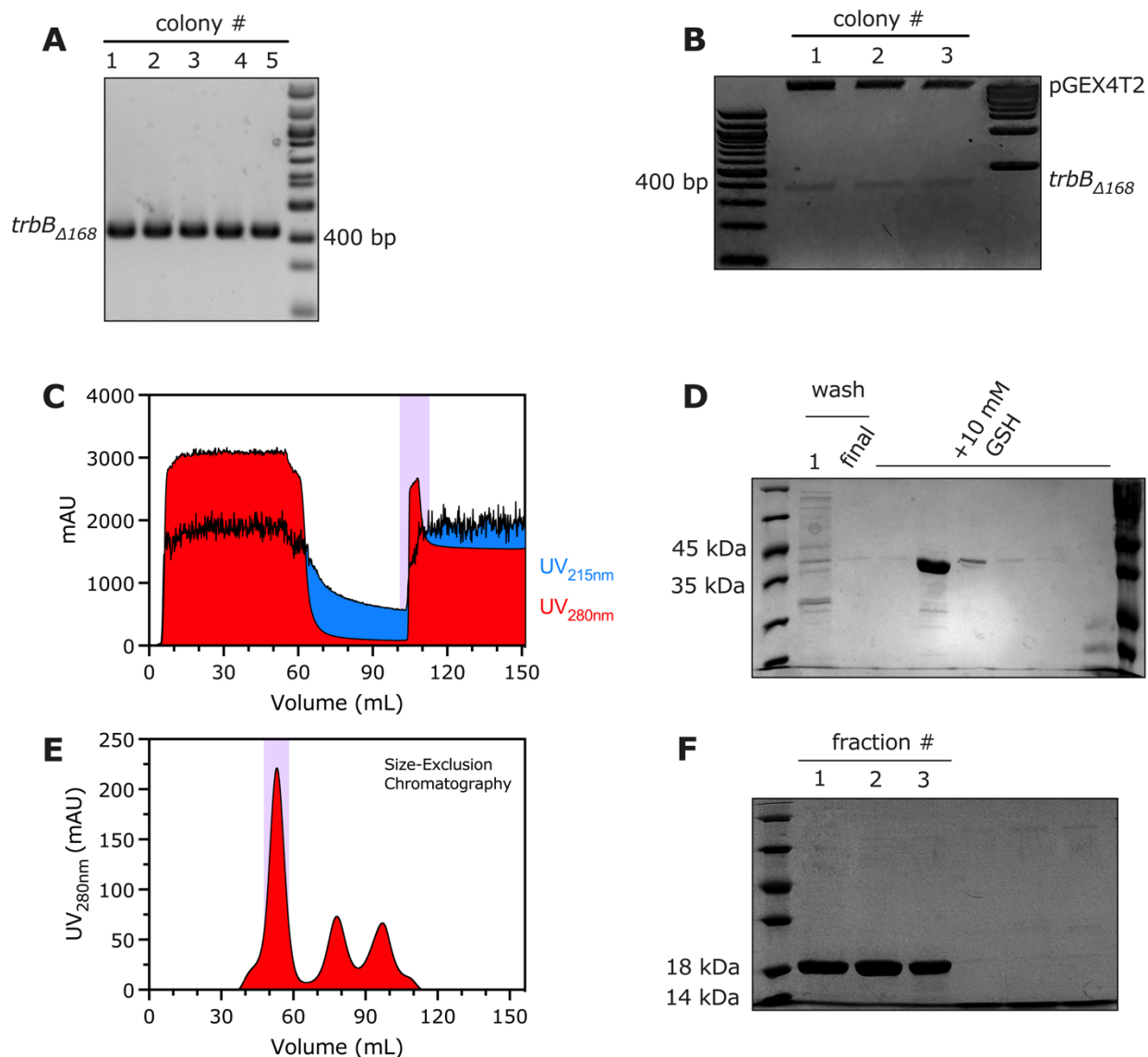


Figure 14. Representative cloning, expression, and purification of GST-TrbB₅₇₋₁₈₁ and TrbB₅₇₋₁₈₁. (A) Colony PCR verification of pGEX-4T-2::*trbB*_{Δ168} BL21(DE3) colonies. Colonies were picked from LB-Agar plates with 100 ug/mL Ampicillin, PCR amplified, electrophoresed in 1.2% agarose gel with EtBr, and visualized under UV light. (B) Colonies which yielded a positive colony PCR result were grown in liquid LB for 16 h, mini-prepped, double-digested with *Eco*RI and *Sal*I for 3 h, and electrophoresed in agarose gel for detection under UV light. Faint signals about 400 bp are evident for all four colonies, while signals corresponding to pGEX-4T-2 at about 5000 bp are strongly evident. (C) Primary GST affinity purification and (D) SDS-PAGE qualification to assess for protein identity and relative purity. (E) Secondary Size-exclusion chromatography purification of the untagged protein following GST-tag cleavage. Peaks were verified for quality by SDS-PAGE and the pure protein of interest from the purple peak of the chromatogram is labeled 1-3 (F).

3.2. Sampling of GST-TrbB's and GST-TrbI's chemical crystallization space as guides for future crystallization campaigns

3.2.1. Crystallization trials on GST-TrbI

The chemical space of protein crystallization is often vast, making crystallization the main bottleneck in protein crystallographic analysis¹¹⁴. Accordingly, GST-TrbB and GST-TrbI protein samples (10 mg/mL in 10 mM MES pH 6.72, 50 mM NaCl, 5% v/v glycerol) were sent for high-throughput crystallization screening at the Hauptman-Woodward Institute⁵ to sample each protein's crystallization space considerably. Both proteins were screened against 1,536 buffer/precipitant/additive conditions and monitored for six weeks. Both proteins appeared to have had a similar crystallization outcome distribution, with most crystallization drops resulting in precipitation; over 800 conditions for GST-TrbB (**Fig. 15A**) and around 800 conditions for GST-TrbI (**Fig. 15C**).

A less stringent (and more optimistic) initial survey of the outcomes for GST-TrbI resulted in the classification of over a hundred crystal hits (**Fig. 15C**). Of these, only two conditions indicated justifiable protein crystals. The first promising condition is shown in **Figure 15D**. The crystal had formed after six weeks, progressing from a tiny speck (see image of the drop after 2 weeks). No Simple Harmonic Generation (SHG) signal is observed, providing evidence that the observed structure under the brightfield microscope may not be a crystal^{114,115}. Still, a modest Ultraviolet Two-Photon Excited Fluorescence (UV-TPEF) signal in the same location of the drop is evident, suggesting that the observed structure is that of a protein. Despite the inconsistency with the SHG signal, the brightfield image shows that the structure is birefringent and that it can be morphologically classified as a lattice crystal—enough evidence to motivate the replication attempt of the crystallization condition.

The second promising condition is shown in **Figure 15E**. The crystal was first observed after 5 weeks. Again, there appears to be no observable SHG signal, and a modest UV-TPEF signal can be distinguished. Interestingly, both crystal hits for GST-TrbI (**Fig. 15D & E**) were both facilitated by a similar type of buffer (*i.e.*, both Tris and HEPES are zwitterionic buffers), with the exact same concentration of 0.1 M. Additionally, both conditions utilized a sulfate salt at 0.8 M.

Furthermore, the pH of both conditions is 7.0, which is around the isoelectric point of GST-TrbI (pI 7.14) and 2 pH units below that of TrbI (pI 9.42).

3.2.2. Crystallization trials on GST-TrbB

Notably, 69 conditions for GST-TrbB (**Fig. 15A**) provided initial promise. However, after a more rigorous inspection, only one condition showed convincing evidence of a crystallization outcome. The condition of this drop, along with the appearance of a crystal under a brightfield microscope after 14 days, is shown in **Figure 15B**. From a kinetics point-of-view, the formation of a crystal in an intermediate timespan of weeks—in contrast to timespan of days—provides some assurance that the observed crystal is not of a salt^{114,116}. Further evidence in support of the existence of a protein crystal is shown by the SHG signal, which indicates the presence of a lattice crystal. Moreover, the UV-TPEF signal, while modest, supports the presence of a protein. It does not escape our notice that the pH of this condition is 10, significantly higher than that of GST-TrbB (pI 6.27) and almost 2 pH units higher than that of TrbB (pI 8.62). In this pH, the population of GST-TrbB in solution is expected to be predominantly deprotonated and anionic.

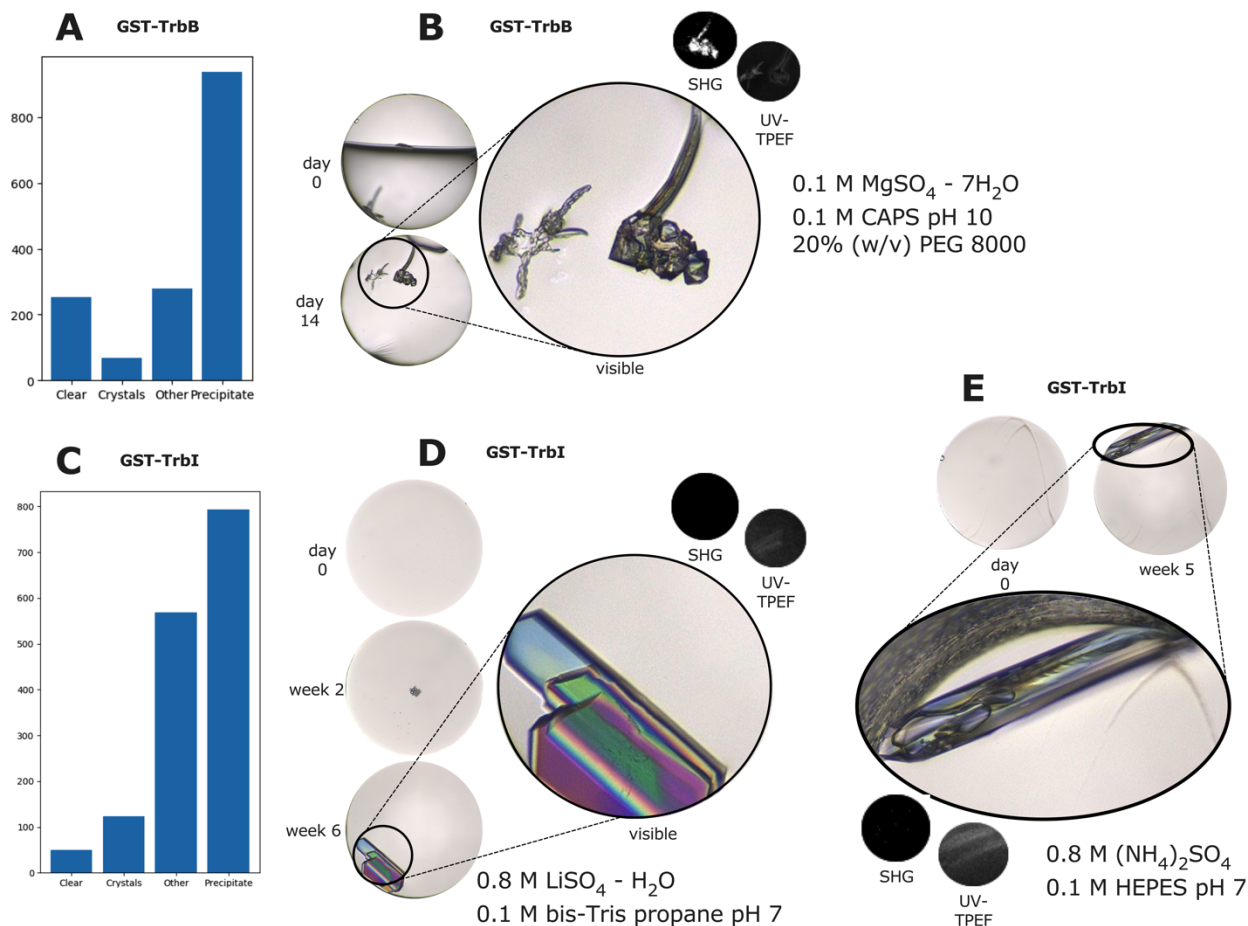


Figure 15. Representation of the chemical space of crystallization sampled by crystallization trials. Outcomes of crystallization trials for (A) GST-TrbB_{WT} or (C) GST-TrbI from 1,536 buffer/precipitant conditions. Promising crystal hits from screens for (B) GST-TrbB_{WT} or (D-E) GST-TrbI. Reservoir buffer-precipitant solutions are shown beside each drop images. Crystal hits are evident under visible light (brightfield) microscopy (see magnified views). A positive SHG signal (white) indicates the presence of a lattice crystal. A positive UV-TPEF (white) is suggestive of a protein. Data from high-throughput screens from the Hauptman-Woodward crystallization Screening Centre¹¹⁵.

3.3. A volume model for GST-TrbB_{WT} illustrates its dynamics

3.3.1. ColabFold-AlphaFold2 Models for GST-tagged cognate T4SS_F proteins

Recent advances in computational structural biology have paved the way for high-throughput 3D protein structure modelling^{109,117}. Protein structures determined using computational methods provide some insights into the organization and structure of proteins in three-dimensional space, a good starting point for structural and functional studies on proteins whose structures are still yet to be solved. ColabFold-AlphaFold2 (CF-AF2) was utilized to generate 3D models for GST-TrbB, GST-TrbI, and GST-TraF (**Fig. 16**), primarily for comparison with empirical CD data (Section 3.5).

Still, these 3D models in and of themselves can be instructive when compared to previous research. The thioredoxin domain (minimum of three α -helices flanking a four-stranded antiparallel β -sheet^{5,100}) of TrbB_{WT} can be observed (**Figs. 16A, 20**). The active CXXC motif is shown in the primary sequence (**Fig. 16A**; see C81-PY-C84) and emphasized in **Figure 30** (dark grey sphere representation). The thioredoxin-like fold of TraF can be observed and the absence of the active CXXC motif is evident in the primary sequence (**Fig. 16B**). TrbI has been previously reported to be a bitopic protein, with its H17-V40 residues spanning the inner membrane (**Fig. 16C** sequence at right) and the remaining 88 hydrophilic residues located in the periplasm⁵. CF-AF2 predicts that H17-V40 forms a α -helical structure, and this is supported by previous research.

These models are important visualization tools, especially for proteins that are recalcitrant to high-resolution structural analyses, but they must be considered from a more nuanced viewpoint. The predicted Template Modelling (pTM) scores of each CF-AF2 is reported beside each model (**Fig. 16**) and they are all lower (GST-TrbB: 0.58; GST-TraF: 0.51; GST-TrbI: 0.62) than the proposed threshold established by Yin and colleagues (2022)¹¹⁸ to be characteristic of a reliable model (pTM = 0.8).

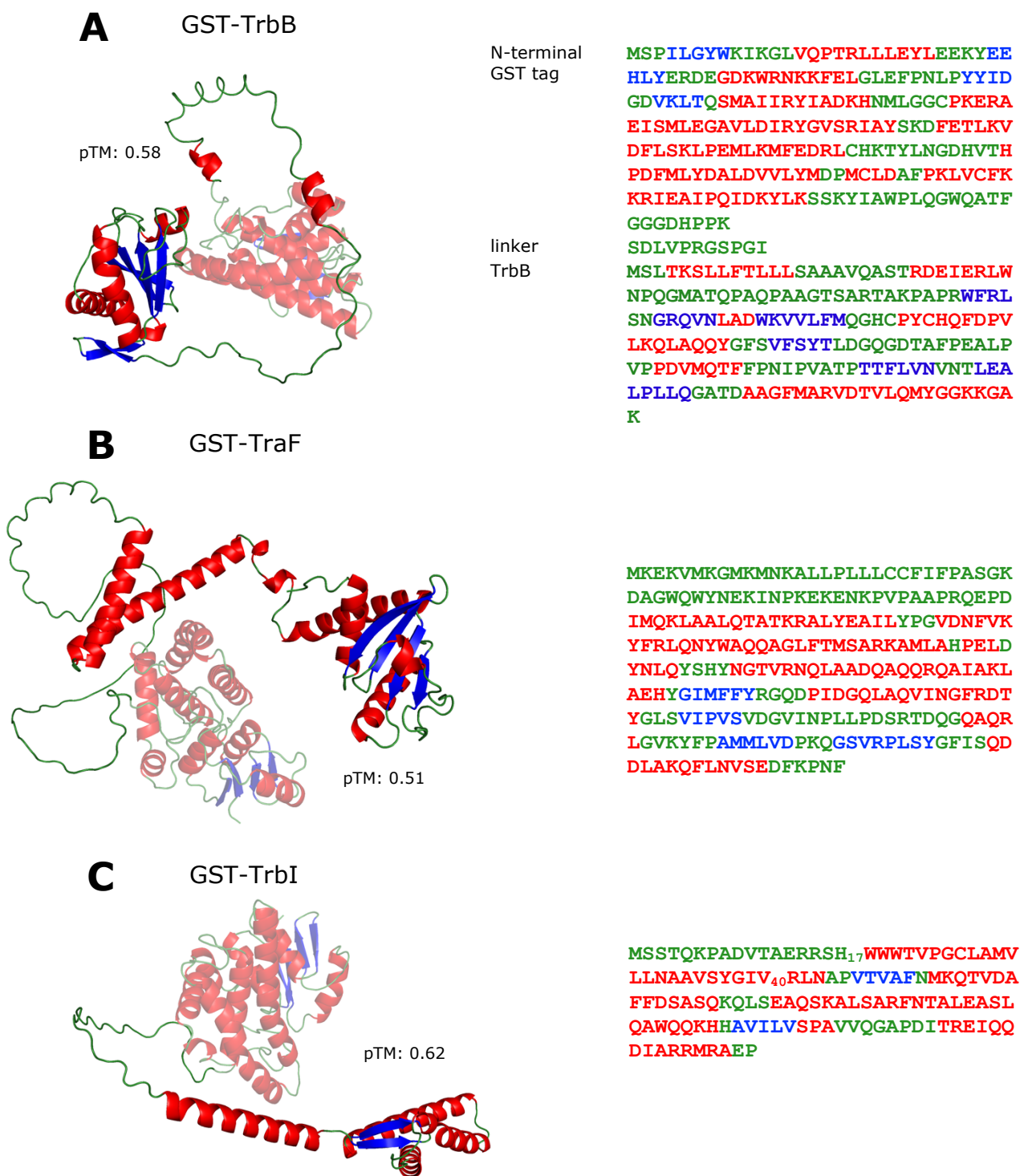


Figure 16. ColabFold-AlphaFold2 models and primary sequences of three T4SS_F proteins. (A) GST-TrbB, (B) GST-TraF, and (C) GST-TrbI, with H17-V40 emphasized. Residues are colour-coded according to CF-AF2's secondary structure prediction. N-terminal GST-tags for each protein are set to 50% transparency. The primary sequence of GST and the linker is common among all three models. GST-TrbB has a total of 410 residues; GST-TraF: 486 residues, and GST-TrbI: 357 residues.

3.3.2. Secondary structures estimations for TrbB_{WT} and TrbB₅₇₋₁₈₁

Still, the need for empirically derived protein structures remains due to errors often inherent in computational models. We investigated whether there are differences between empirical and computational methods with respect to estimating secondary structures for TrbB. Comparing computational data from ColabFold-AlphaFold2 (CF-AF2) and empirical data from Circular Dichroism (CD) spectroscopy, differences in percentage composition are evident (**Fig. 17B, C**). The CF-AF2 predicted model of TrbB_{WT} indicates a secondary structure composition of 26.5% α -helix and 19.3% β -sheets (**Fig. 17C**). In contrast, CD spectroscopy empirically determined the secondary structure content at 6.2% α -helix and 35.4% β -sheets (**Fig. 17B**). In the TrbB₅₇₋₁₈₁ truncation mutant, CF-AF2 predicts protein composition of 38.4% α -helix and 28.0% β -sheets (**Fig. 17C**); CD spectroscopy indicates 6.1% α -helix and 33.7% β -sheets (**Fig. 17B**).

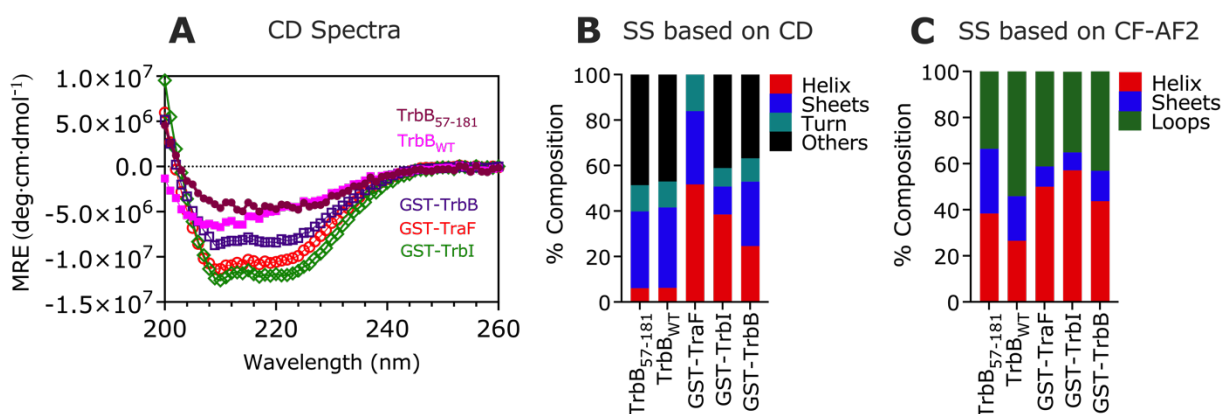


Figure 17. Secondary structure estimations. (A-B) Empirical estimation by Circular Dichroism spectra analyzed using BeStSel¹⁰⁸. CD measurements were collected in triplicates from 5 μ M protein samples at 22°C, expressed as Mean Residue Ellipticity (MRE). (C) Quantitative estimation from ColabFold-AlphaFold2 models. For comparison to empirical CD data, quantitation was performed by normalizing number of residues predicted to form α -helices, β -sheets, or loops, by the total number of residues in the protein construct. Secondary structures classified as “others” include 3₁₀ helix, π -helix, β -bridge, bend, loops, irregular and/or disordered regions¹⁰⁸.

3.3.3. SAXS/CF-AF2 model for GST-TrbB_{WT} illustrates its dynamics

In a multi-methods approach, GST-TrbB_{WT} was further analyzed using SEC-MALS-SAXS. Following SAXS data reduction, an *ab initio* reconstruction was performed to obtain a low-resolution model for GST-TrbB_{WT} from empirical Small Angle X-Ray Scattering (SAXS) data, while also fitting

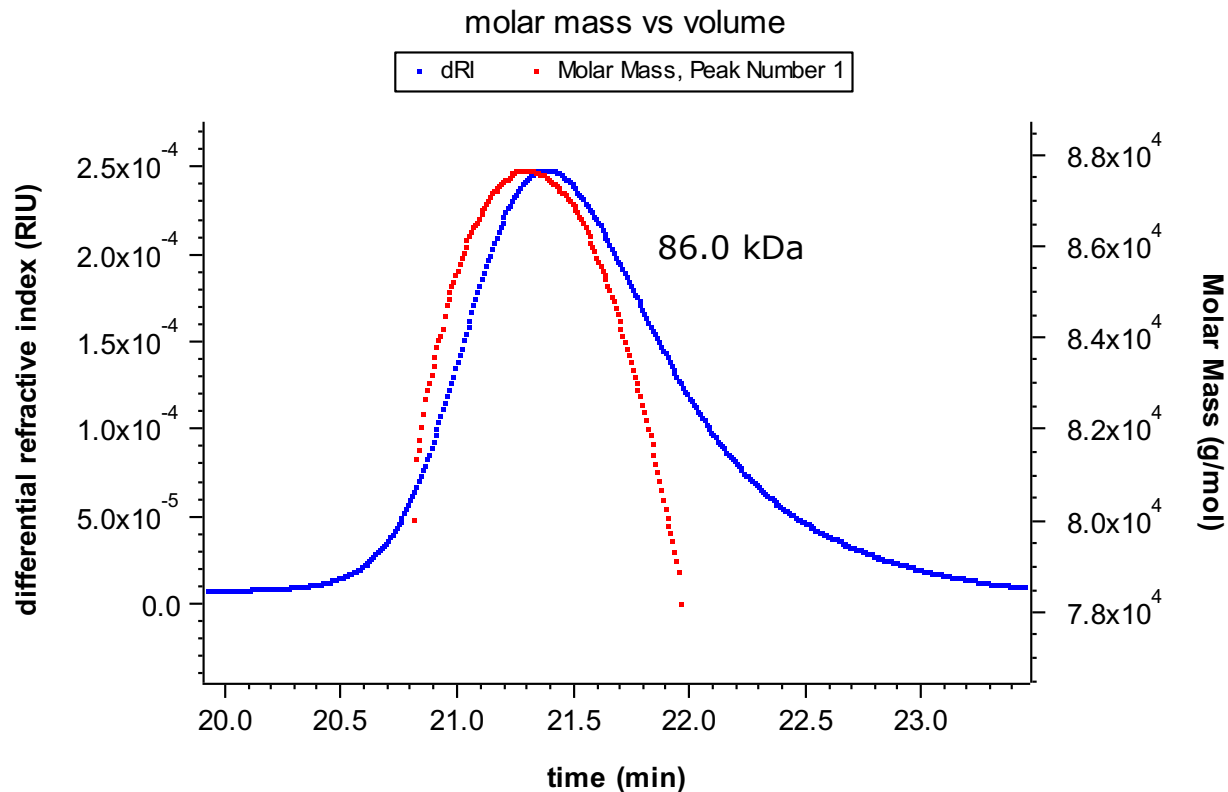
computational CF-AF2 model to the data. Notably, the SEC-MALS data (86.0 kDa) and molecular weight analyses from SAXS (Vd: 119.9 kDa; Vc: 102.6; Shape and Size: 106.4 kDa; and Bayes: 109.1 kDa) were characteristic of a homodimeric GST-TrbB_{WT} (**Fig. 18**), which is reflected as two GST-TrbB_{WT} monomers fit into the SAXS volume bead model. There has been no empirical evidence to suggest that TrbB forms a stable dimer, but the presence of C81 and C84 residues in its primary sequence (**Fig. 16**) leaves room for the possibility that it can dimerize through intermolecular disulfide bonds^{119–121}, albeit at a decreased likelihood because one TrbB molecule can only form one cystine. On the other hand, the GST protein is known to form stable homodimers^{122–124} and the dimerization of the GST moieties can explain the observed homodimeric GST-TrbB_{WT}.

At first inspection, it is evident that there is not a perfect fit between the CF-AF2 predicted GST-TrbB_{WT} model and the SAXS bead model (**Fig. 19E**). However, it is important to note that SAXS provides a volume model generated from a signal-averaged light scattering of the protein as it freely diffuses and occupies 3D space in solution, suggesting that the bead model represents some of the dynamics of the protein in solution^{125,126}. Deviations from the coordinates (1.732, 1.104) in the Kratky plot (**Fig. 19C**, marked by red-dashed cross) indicate structural disorder¹²⁷. A partially disordered protein is often indicated by a bell-shaped Gaussian peak that gradually returns to the baseline¹²⁸, and this is evident in the presented Kratky plot (**Fig. 19C**). Accordingly, SAXS bead models are often utilized in concert with high-resolution models, such as that solved by X-ray crystallography or NMR, to determine whether the protein adopts a wider range of conformations that high-resolution methods cannot determine^{129,130}. TrbB_{WT} has evaded high-resolution structure characterization, but the SAXS/CF-AF2 model provided meaningful insights, indicating that TrbB_{WT} is dynamic (blobby model; **Fig. 19E-F**), partially disordered (Kratky plot, **Fig. 19C**), and that the disordered region is likely the N-terminus (**Fig. 19E-F** CF-AF2 model fit into bead model; **Fig. 20**). These insights informed the design of the truncation mutant, TrbB₅₇₋₁₈₁.

The reliability of the SAXS data is backed by well-fit Guinier (**Fig. 19B**, see r^2 close to 1) and GNOM (**Fig. 19D**, see χ^2 close to 1) analyses. Further, the radius of gyration (Rg) values from the Guinier (43.14 ± 0.27 Å) and GNOM (40.41 ± 0.14 Å) analyses are within an agreeable range (**Table 3** structural parameters). The Rg is the weighted root mean square of the intramolecular distances with respect to the centroid of the electron density, effectively quantitating the size of

the macromolecule in solution¹²⁷. The size of a macromolecule is different in solution as it dynamically moves and interacts with solvent molecules, compared to when it is static in an ordered lattice crystal, and TrbB_{WT}'s dynamic N-terminal region certainly augments the observed Rg. Other proteins within a similar MW range as homodimeric GST-TrbB_{WT} (SEC-MALS: 86.0 kDa; Vd: 119.9 kDa; Vp: 102.6; Shape and Size: 106.4 kDa; and Bayes: 109.1 kDa; **Fig. 18**) have Rg values that are lower compared to GST-TrbB_{WT} (homodimeric yeast enolase, MW: 79.4 kDa, Rg: 27.6 Å; homodimeric rabbit enolase, MW: 86.4 kDa, Rg: 28.3 Å; monomeric transferrin, MW: 76.9 kDa, Rg: 31.1 Å; and homodimeric BSA, MW: 137 kDa, Rg: 36.2 Å; ref¹³¹).

The normalized pair distribution function $P(r)$ (**Fig. 19D**) is characteristic of an elongated macromolecule shape as opposed to a globular (bell-shaped Gaussian), dumbbell (bimodal), or a core-shell (leading asymmetric peak close to the Dmax)¹²⁷. GST-TrbB_{WT} could, in principle, adopt globular and dumbbell shapes if the GST and TrbB moieties are packed by intermolecular interactions into a dense compact structure or if the GST and TrbB_{WT} are arranged linearly by the linker region (comprised of the residues that link GST and TrbB_{WT} and the disordered N-terminus of TrbB_{WT}), respectively. However, these are not indicated by the $P(r)$ curve (**Fig. 19D**), which is observably a tailing asymmetric peak marked by a modest second peak mid r , and gradually approaches 0 at high r , suggesting an elongated shape.



Method	Estimated molecular weight (kDa)
amino acid sequence ^{132*}	92.8 (homodimer)
Porod Volume (Vp) ¹³³	119.6
Volume of correlation (Vc) ¹³⁴	102.6
Shape & Size ¹³⁵	106.4
Bayes ¹³⁶	109.1

*determined computationally by Expasy ProtParam; other methods are experimental based on SAXS

Figure 18. Size Exclusion Chromatography (SEC) Multi Angle Light Scattering (MALS) on protein sample and molecular weight estimations characteristic of a homodimeric GST-TrbB_{WT}. SEC-MALS was collected by the BioCAT 18ID beamline facility (Advanced Photon Source, Argonne National Laboratory, USA) using a GE Superdex 200 Increase 10/300 column at a flow rate of 0.6 mL/min coupled to a Wyatt DAWN Heleos II MALS system. Molecular weight estimations were based on SAXS data determined using BioXTAS RAW v.2.2.1¹¹².

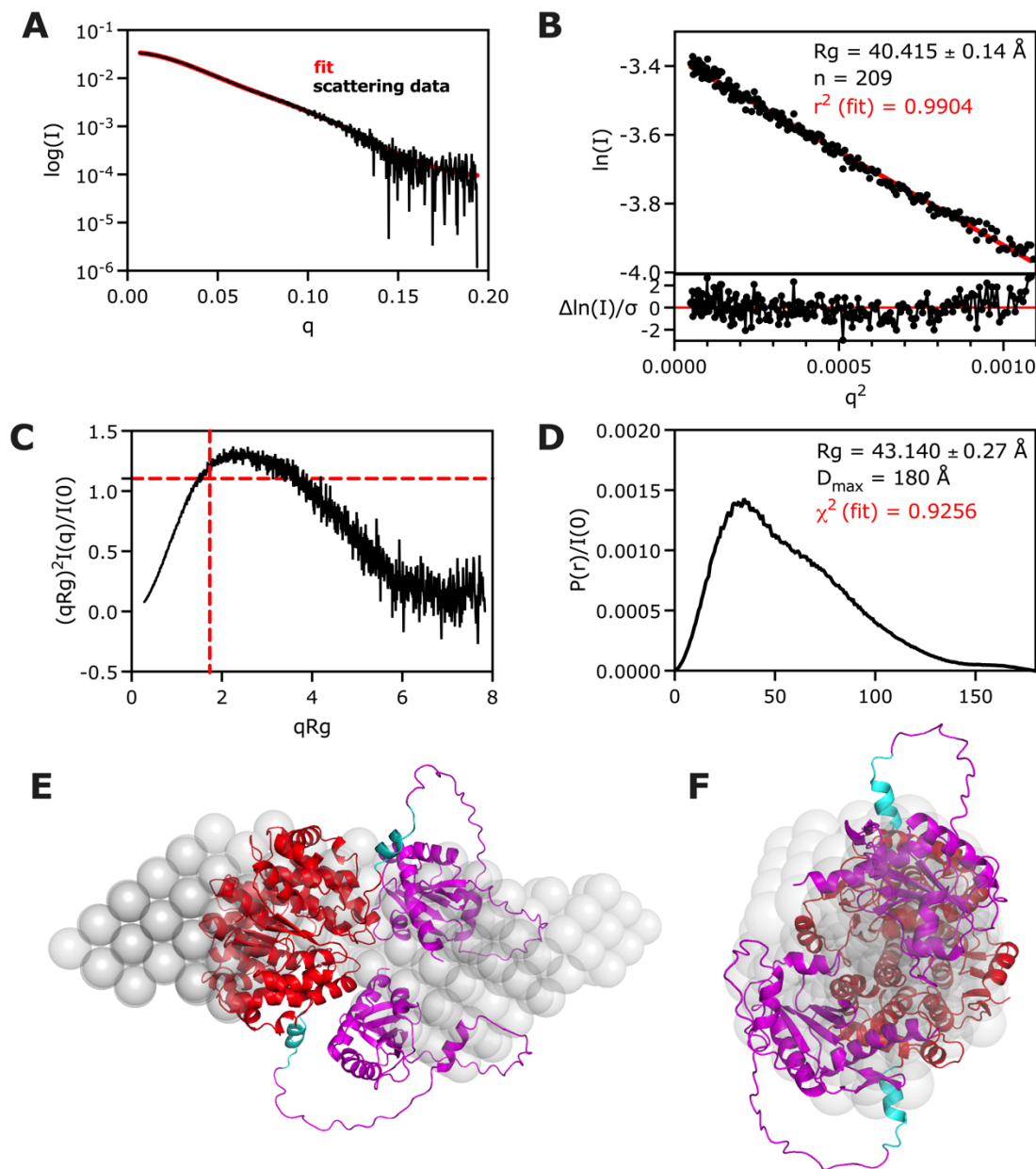


Figure 19. AlphaFold2 model of homodimeric GST-Trb_{WT} fitted into *ab initio* 3D reconstruction for homodimeric GST-Trb_{WT} from SAXS data. (A) Scattering profile. (B) Guinier fit analysis of 5.5 mg/mL GST-Trb_{WT}. (C) Kratky plot. (D) GNOM analysis. (E-F) ColabFold-AlphaFold2 model for GST-Trb_{WT} fitted into the SAXS volume model. (E) Side-view and (F) cross-sectional view are both shown. The CF-AF2 model for GST-Trb_{WT} was generated from its amino acid sequence. GST moiety is coloured in red, the linker in cyan, and Trb_{WT} is magenta.

Table 3. SEC-MALS-SAXS parameters for homodimeric GST-TrbB_{WT}

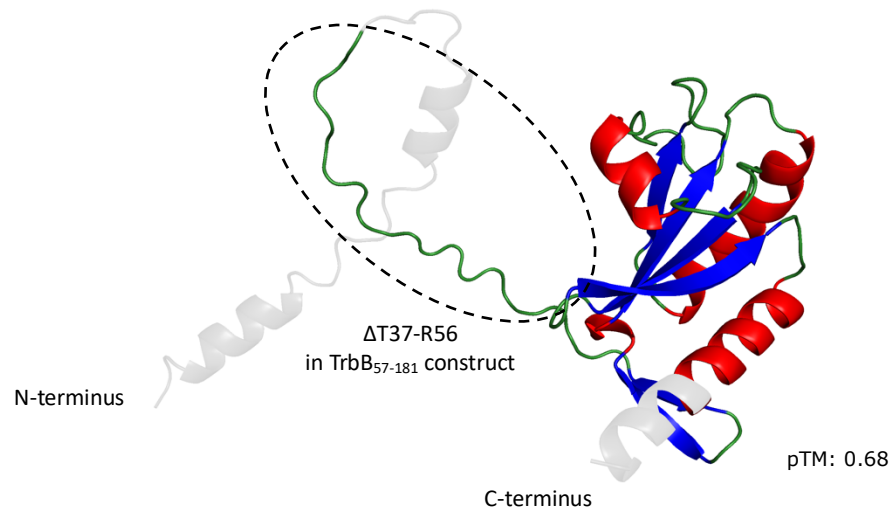
Data collection parameters		
Instrument	BioCAT (Sector 18, APS)	
Detector	Eiger2 XE 9M	
Wavelength (Å)	1.033	
q-measurement range (1/Å)	0.0028 to 0.42	
Exposure time (s)	0.5	
Size exclusion column	Superdex 200 10/300 Increase	
Flow rate (mL/min)	0.6	
Temperature (°C)	20	
Protein concentration (mg/mL)	5.5	
Loaded volume (μL)	300	
Buffer	20 mM HEPES pH 7.0, 100 mM NaCl, 5% glycerol, 0.05% NP40	
Structural parameters		
	from P(r) or GNOM	from Guinier
I(0)	0.0348 ± 1.15 x 10 ⁻⁴	0.0342 ± 7.41 x 10 ⁻⁵
R _g (Å)	40.41 ± 0.14	43.14 ± 0.27
D _{max} (Å)	180	

3.4. Design of a more stable truncation construct, TrbB₅₇₋₁₈₁

3.4.1. Truncation mutant protein, TrbB₅₇₋₁₈₁, design

Intrinsically disordered and/or highly dynamic regions are increasingly becoming appreciated as functional moieties in proteins, highlighted by the discoveries of their function in interactomes¹³⁷. These disorder-based interactions are gaining more research attention and are being found to transiently bind a diverse set of partner proteins at high specificity^{138–140}. Furthermore, these proteins tend to self-associate and form stable aggregates¹⁴¹, and aberrations to their typical physicochemical characteristics *in vivo* are implicated with cancer and neurodegenerative diseases, such as Tau aggregation, one of the hallmarks of Alzheimer's disease^{142–144}.

T4SS_F proteins are dynamic, and these dynamic disordered regions may play critical roles in their functions in the context of the T4SS_F. However, disordered regions pose significant challenges to structural analyses¹⁴⁵, and some have been shown to be recalcitrant toward crystallization^{5,146}. There are a multitude of factors that dissuade proteins from adopting a crystalline structure and one of them is the presence of intrinsically disordered regions¹⁴⁵. TrbB₅₇₋₁₈₁ was designed based on the computational CF-AF2 model of GST-TrbB_{WT} (**Figs. 20; 16A, 19E-F**) to remove the dynamic N-terminal region comprised of residues M1-R56 (**Fig. 20**), decrease the protein's dynamicity that is observed from the SAXS bead model (**Fig. 19E-F**), and favor crystallization for subsequent crystallographic analysis.



TrbB full sequence

MSLTKSLLFTLLLSAAAVQASTRDEIERLWNPQGMATQPAQPAAGTSART
AKPAPRWFRLSNGRQVNLADWKVVLFMQGHCPYCHQFDPVLKQLAQQYGF
SVFSYTLDGQGDATFPEALPVPDVMQTFFPNIPVATPTTFLVNVNTLEA
LPLLQGATDAAGFMARVDTVLMYGGKKGAK

Signal sequence ¹⁴⁷

M₁SLTKSLLFTLLLSAAAVQA₂₀

Thioredoxin domain

T₃₇QPAQPAAGTSARTAKPAPRWFRLSNGRQVNLADWKVVLFMQGHCPYC
HQFDPVLKQLAQQYGFSVFSYTLDGQGDATFPEALPVPDVMQTFFPNIP
VATPTTFLVNVNTLEALPLLQGATDAAGFMARVDTVLMY₁₇₂

Residues deleted in TrbB₅₇₋₁₈₁

T₃₇QPAQPAAGTSARTAKPAPR₅₆

Figure 20. ColabFold-AlphaFold2 model for TrbB_{WT} emphasizing the thioredoxin domain and residues T37-R56 deleted in TrbB₅₇₋₁₈₁ construct. The full amino acid sequence, the signal sequence (predicted by SinalP¹⁴⁷) that gets cleaved to form the mature protein *in vivo*, the putative thioredoxin domain (Uniprot: P18035; Prosite: PRU00691), and the residues deleted in TrbB₅₇₋₁₈₁ (dotted circle) are shown. Residues are coloured based on CF-AF2's predicted secondary structures; red for α -helices, blue for β -sheets, and green for loops. Only the thioredoxin domain is shown in colour. CF-AF2 is described by Mirdita and colleagues (2022)¹⁰⁹.

3.4.2. CD spectroscopy thermal denaturation studies support TrbB₅₇₋₁₈₁'s thermal stability

Following successful design, cloning, expression, and purification of TrbB₅₇₋₁₈₁ (**Fig. 14**), the next step was to test whether the deletion of the CF-AF2-predicted disordered N-terminus would result in meaningful physicochemical changes and stability. CD spectroscopy analyses were performed to probe changes in the secondary structure between the full-length and truncated forms of TrbB. The percentage of α -helices was observed to decrease from 52.4% in TrbB to 6.1% in TrbB₅₇₋₁₈₁ (**Fig. 17B**). β -sheet content, on the other hand, was observed to increase in composition from 31.8% to 33.7%, respectively (**Fig. 17B**).

In addition, we examined if the deletion mutation altered the thermostability of the protein in solution. Thermal denaturation profiles were observed *via* CD at 222 nm (monitoring α -helices) as a function of temperature (**Fig. 21**). This wavelength was chosen because it is characteristic of α -helices, effectively allowing the observation of the thermal denaturation of α -helices. The observed midpoint of unfolding (T_M) of the population of TrbB₅₇₋₁₈₁ in solution is 78°C in all three replicates (**Fig. 21A**). Conversely, the thermal denaturation profile of full-length TrbB_{WT} is significantly inconsistent across replicates (**Fig. 21B**), making it difficult to reliably ascertain its T_M .

CD thermal denaturation profiles of TrbB₅₇₋₁₈₁ demonstrated that, while it is comprised of only 6.1% α -helices, its T_M is 78°C (**Fig. 21A**). This provides insight into the organization of α -helices with respect to the protein. Of the residues in the CF-AF2 model forming α -helices, 43.8% are polar and charged; 25.0% polar, 18.7% charged (**Fig. 22**). The presence of a considerable amount of polar and charged residues may be one of the factors contributing to the 78°C T_M of TrbB₅₇₋₁₈₁. Without a crystal structure, it is difficult to make a sound inference, but it cannot be ruled out that the α -helices may be surrounded by stabilizing intermolecular forces (e.g., a network of hydrogen bonds).

The T_M of TrbB₅₇₋₁₈₁ is strikingly different to that of TrbB_{WT} whose thermal denaturation profile is inconsistent across replicates (**Fig. 21B**), making it impossible to ascertain its T_M precisely. We infer that the reason why the full-length protein fails to have a consistent thermal denaturation profile, and midpoint of unfolding, is because it is unstable in solution. Nevertheless, the high T_M of TrbB₅₇₋₁₈₁ suggests that the truncation was effective in improving the protein's

thermal stability compared to the full-length form. As well, the finding provides optimism for ongoing crystallization efforts, primarily because thermally stable proteins are typically easier to crystallize^{116,148}.

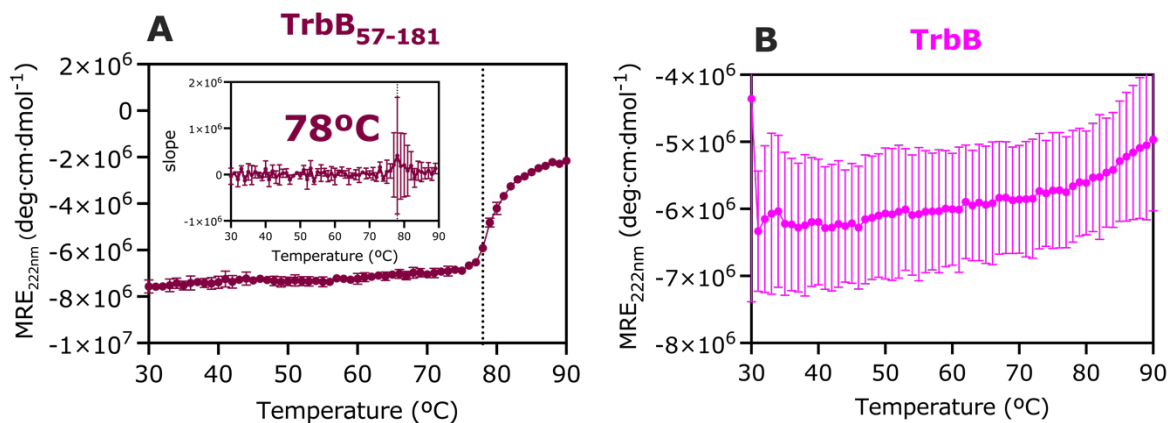


Figure 21. Thermal denaturation profiles of α-helices in (A) TrbB₅₇₋₁₈₁ and (B) full-length TrbB_{WT}. Mean residue ellipticity (MRE) at 222 nm of 5 μM protein samples was monitored as a function of temperature. The midpoint of unfolding (T_M) is given by the temperature at which the first derivative of the CD vs T curve is at its highest; T_M of TrbB₅₇₋₁₈₁ is 78°C (in all replicates), that of TrbB_{WT} cannot be reliably ascertained due to significant data deviations. Data points are shown as mean ± SD, n = 3 replicates.

A TrbB₅₇₋₁₈₁ amino acid sequence

WFRLSNGRQVNLADWKVVLFMQGHCPYCHQFDPVLKQLAQQYGFVSFYSYTLDGQGDATFPEALPVPDVMQTFFPNI
PVATPTTFLVNVNTLEALPLLQGATDAAGFMARVDTVLMYGGKKGAK

B TrbB₅₇₋₁₈₁ amino acids forming α -helices

	residues	No. residues	%
α -helical residues	LADPYCHQFDPVLKQLAQQYPDVMQTF AAGFMARVDTVLM YGGKKGA	48	38.4
polar uncharged	YCQQQQYQTTQY	12	25.0
negatively charged*	DDDD	4	8.3
positively charged*	HKRKK	5	10.4
polar & charged		21	43.8

*at physiological pH

C CF-AF2 model for TrbB₅₇₋₁₈₁ in 2 orientations

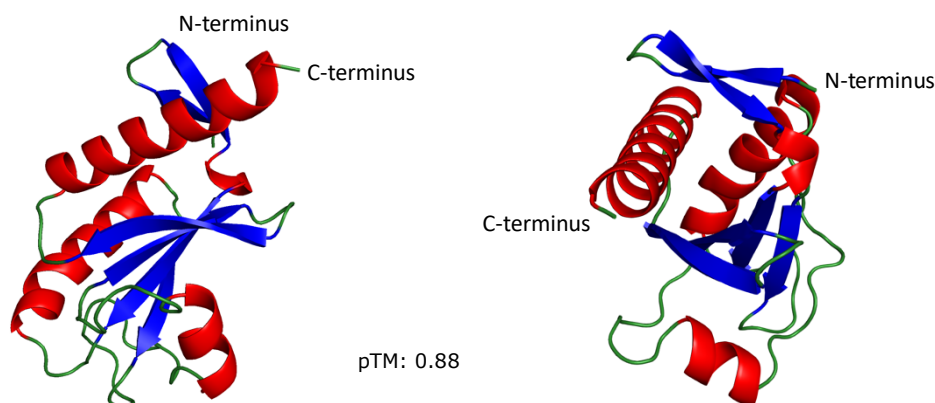


Figure 22. TrbB₅₇₋₁₈₁ residues predicted to form α -helices. (A) The full amino acid sequence, (B) amino acids predicted by ColabFold-AlphaFold2 (CF-AF2) to form α -helices (in red; in table), and the amino acids classified according to their polarity and charge (table) are shown. (C) CF-AF2 3D structure is shown to provide some insights into the organization of the secondary structures in the protein in 3D space; left model distinguishably shows the N- and C-terminus of the protein. Residues in blue are β -sheets, and those in green are loops. CF-AF2 is described by and colleagues (2022)¹⁰⁹. Structural analysis was performed using PyMOL v2.5.2.

3.4.3. Heteronuclear Single Quantum Coherence supports the stability of TrbB₅₇₋₁₈₁

The two forms of TrbB were also compared using ^1H - ^{15}N Heteronuclear Single Quantum Correlation (HSQC) Nuclear Magnetic Resonance (NMR) spectroscopy to determine if the truncation mutant, TrbB₅₇₋₁₈₁ (black), would improve the protein's stability compared to TrbB_{WT} (magenta). Discernable differences between the two TrbB constructs were observed (**Fig. 23**). Each concentric signal indicates the presence of a correlated ^{15}N - ^1H nuclei that are directly bonded to each other, providing a good estimate of the number of N-H bonds in the protein^{149,150}. Peak assignment and structure solution is more feasible for TrbB₅₇₋₁₈₁, whose spectrum is composed of more resolved signals compared to the overlapping peaks of TrbB_{WT} (**Fig. 23**). Moreover, the increased dispersion of TrbB₅₇₋₁₈₁ signals in the ^1H dimension (x-axis), compared to that of TrbB_{WT}, indicates a more well-folded protein^{149,150}. Furthermore, the TrbB_{WT} sample has pronounced clustering of signals around 8 ppm at the ^1H dimension compared to the concentric and dispersed signals from TrbB₅₇₋₁₈₁ that is especially observable in 8-8.5 ppm of the ^1H dimension (**Fig. 23**). This data, along with that of the CD spectroscopy and thermal denaturation studies, indicate that the truncation mutant is more stable in solution.

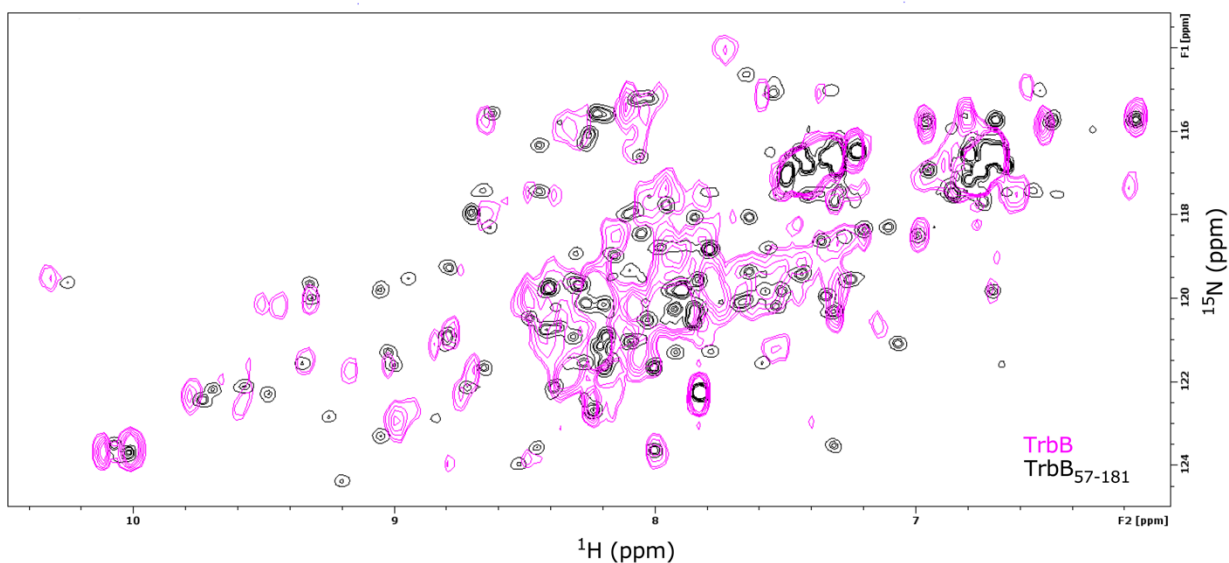


Figure 23. ^1H - ^{15}N Heteronuclear Single Quantum Correlation (HSQC) Spectra showing improved signals from TrbB₅₇₋₁₈₁ compared to TrbB_{WT}. Protein samples were at a concentration of 0.1 mM in 10 mM MES pH 6.0, 50 mM NaCl, 10% (v/v) D₂O, recorded at 600 MHz. TrbB_{WT} spectrum is shown in magenta, and TrbB₅₇₋₁₈₁ in black.

3.5. Comparative structural studies on cognate T4SS_F proteins

3.5.1. GST-TrbB is more similar to GST-TrbI at the secondary structure level compared to GST-TraF

TrbB is reported to have a thioredoxin domain, while TraF has a thioredoxin-like fold without the active CXXC motif^{5,89,100}. These proteins were compared at the secondary structure level using CD spectroscopy, and a stark difference in α -helical content is observed (GST-TrbB 24.6% and GST-TraF 51.7%; **Fig. 17B**). In terms of α -helical content, GST-TrbB (24.6%) is more similar to GST-TrbI (38.5%), another T4SS_F protein (**Fig. 17B**). β -sheets content (**Fig. 17B**) of GST-TrbB (28.3%) and GST-TraF (32.2%) are more similar compared to that of GST-TrbI (12.3%). Interestingly, while all three bear GST-tags, only GST-TrbB and GST-TrbI are comprised of “other” secondary structures (3,10 helix, π -helix, β -bridge, bend, loops, irregular and/or disordered regions¹⁰⁸), suggesting that these “other” structures are not owing to the GST-tag. GST-TrbB is indicated to have 36.8% “other” secondary structures, while that of GST-TrbI is 41.1% (**Fig. 17B**). Observation of these “other” structures, and the possibility that they are disordered regions, might rationalize the more recalcitrant nature of GST-TrbB and GST-TrbI to crystallization compared to GST-TraF, which has been crystallized previously¹⁰⁰. While these comparisons provide comparative insights to the structure of T4SS_F proteins, it is notable that they all bear GST-tags, which is also a protein that affects the CD signals and estimated secondary structure compositions¹⁰⁸.

3.5.2. GST-TrbB_{WT} vs. GST-TrbI vs. GST-TraF ColabFold-AlphaFold2 models

In predicting the 3D structure of proteins, CF-AF2 necessarily provides a prediction for secondary structures. The α -helical content of GST-TrbB is predicted to be 43.7%, 57.1% for GST-TrbI, and 50.0% for GST-TraF (**Fig. 17C**), which are all similar. β -sheets content is predicted as follows: GST-TrbB_{WT} 13.2%, GST-TrbI 7.8%, GST-TraF 8.8% (**Fig. 17C**). CF-AF2 does not predict β -turns, unlike BeStSel, but instead groups secondary structures that are not helices and sheets as loops, predicting GST-TrbB_{WT} to have 43.2%, GST-TrbI 35.0%, and GST-TraF 41.2% (**Fig. 17C**). While the 3D CF-AF2 models for GST-TrbI, GST-TrbB, and GST-TraF are generally in agreement with previous research (see Section 3.3.1, second par.), these secondary structure estimations must be taken with a grain of salt, especially since the pTM values of the models are low.

3.5.3. Comparing CF-AF2 and CD spectroscopy 2° structure estimations

CD estimation for α -helical content indicates that GST-TrbB (24.6%) and GST-TrbI (38.5%) are more similar to each other compared to GST-TraF (51.7%) (**Fig. 17B**), but this pattern is not consistent with CF-AF2, which indicates that α -helical content is comparable among GST-TrbB_{WT} (43.7%), GST-TraF (50.0%), and GST-TrbI (57.1%) (**Fig. 17C**). CD estimation of β -sheets content (**Fig. 17B**) suggests that GST-TrbB (28.3%) and GST-TraF (32.2%) are more similar compared to that of GST-TrbI (12.3%). CD estimation among the three proteins is again contrasted by CF-AF2 β -sheets content estimation (GST-TrbB 13.2%, GST-TraF 8.8%, GST-TrbI 7.8%; **Fig. 17C**), which predicts a close similarity among the three proteins. These differences support the idea that a more nuanced viewpoint must be adopted when analyzing computational and experimental structures, and in some ways also support the use of multi-methods approach (such as our work described in section 3.3.3 of this thesis) in a post-AlphaFold-breakthrough era.

3.6. TrbB functions as a disulfide isomerase but it does not bind its client protein *in vitro*

3.6.1. TrbB functions as a disulfide isomerase *in vitro*

The effects of N-terminal deletion to the enzymatic activity of TrbB *in vitro* was investigated. Firstly, we re-established¹⁰² that TrbB_{WT} does function as a disulfide isomerase (DI) *in vitro* compared to Bovine Serum Albumin (BSA), known not to function as a disulfide isomerase, to a statistically significant difference ($p < 0.0001$; **Fig. 24B, Table 4**). An *in vitro* kit (Abcam) employing a quenched-fluorophore substrate was utilized. In the presence of a DI, the substrate is cleaved, liberating the Fluorophore from the Quencher, leading to the observed fluorescence signals (excitation at 490 nm and emission detection at 580 nm). Our finding strengthens a previous experimental report on the DI activity of TrbB by providing a statistical support that the observed enzymatic activity is not due to chance. The rising trend observed for that of BSA can be inferred as owing to background signals; for example, the degradation of the substrate and the liberation of the fluorophore from the quencher.

3.6.2. The N-terminus of TrbB_{WT} (M1-R56) is not required for enzymatic activity *in vitro*

The functional importance of the N-terminal region of TrbB (M1-R56) *in vitro* was investigated by assaying 50 μ M TrbB₅₇₋₁₈₁ as compared to 50 μ M BSA (**Fig. 24A, Table 4**), still observing a statistically significant difference ($p < 0.0001$). Comparison between TrbB_{WT} and TrbB₅₇₋₁₈₁ indicated no significant difference ($p = 0.76$) between the two protein forms (**Fig. 24C, Table 4**), providing evidence that the N-terminus is not required for disulfide isomerase activity *in vitro*. Consistent with the CF-AF2 model, the N-terminus has been previously reported to be unstructured¹⁰². SignalP predicts (with a high probability of 0.98) that the first 20 N-terminal residues are cleaved by signal peptidase I, as the protein is translocated to the periplasm¹⁴⁷, and the cleavage site is located between A20 and S21 (**Fig. 20**). This proposed cleavage event and the *in vitro* DI activity assays presented suggest that residues M1-A20 do not play a role in TrbB's function in the context of the T4SS_F because *in vivo*, and the mature protein would only constitute residues S21-K181 following its translocation into the periplasm.

TrbB₅₇₋₁₈₁ was therefore tested to ascertain if it would retain its disulfide isomerase function *in vitro* (**Fig. 24**). Interestingly, the set of fluorometric signals (obtained from two independent protein samples, each measured in triplicates) for TrbB₅₇₋₁₈₁ is statistically distinct compared to the negative control BSA to a p -value < 0.0001 (**Fig. 24A**). Moreover, statistical t-test comparison of TrbB_{WT} and TrbB₅₇₋₁₈₁ provides evidence that deletion of the 56 N-terminal residues confers no meaningful difference ($p = 0.76$) with respect to the disulfide isomerase activity *in vitro* (**Fig. 24C**). These findings provide evidence that, at least *in vitro*, the N-terminal region of the protein is not required for enzymatic activity.

Since a good portion (34 amino acids) of TrbB's N-terminus following the putative signal peptide was deleted in the TrbB₅₇₋₁₈₁ construct, it is still a valid concern whether the truncation would affect the function and structure of TrbB. After all, the initial and primary intention in designing the truncation mutant was to improve protein crystallization outcomes, and if the truncation affects structure and function compared to the wild-type protein, one could argue the veracity of a determined crystal structure for further studies such as structure-based drug design. However, following the deletion of the N-terminal residues, TrbB₅₇₋₁₈₁ retains its disulfide isomerase function to a statistically comparable extent to that of TrbB_{WT} (**Fig. 24C, $p = 0.76$**).

Three implications could be drawn from these data. Firstly, preliminary empirical evidence that residues M1-A20 do not play a role in the catalytic activity of TrbB, as they are likely cleaved to form the mature protein, is provided, supporting the prediction by SignalP¹⁴⁷. Secondly, the catalytic domain of TrbB is not dependent on residues S21-R56. This is an interesting finding because UniProt (P18035) and Prosite (PRU00691) predict that residues T37-Q172 constitute the thioredoxin domain of TrbB, which means part of the putative thioredoxin domain of TrbB_{WT} was deleted in TrbB₅₇₋₁₈₁ (**Fig. 20**). Thirdly, we provide support for the assertion that the C81-XX-C84 moiety is essential for function as it is retained in our TrbB₅₇₋₁₈₁ construct^{89,102}. Stated generally, the truncated TrbB₅₇₋₁₈₁ protein remains functional and therefore structure solution of this protein will serve to forward future structure-based studies aiming toward disrupting the action of T4SS_F and hence mitigating T4SS-mediated pathologies.

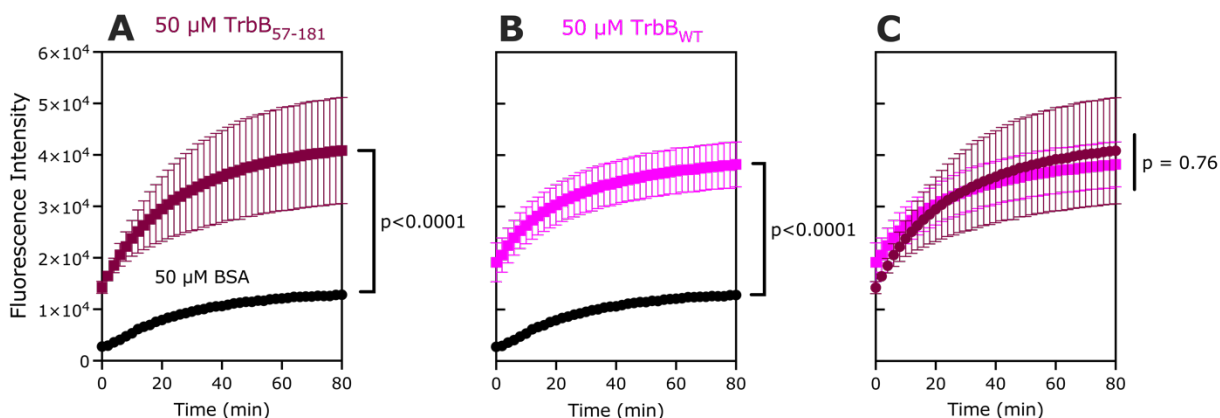


Figure 24. TrbB functions as a disulfide isomerase *in vitro*. Fluorometric assay at Excitation/Emission = 490nm/580nm on (A) N-terminal-truncated TrbB₅₇₋₁₈₁, (B) TrbB_{WT}, and (C) comparison between the two TrbB forms. Data points are mean \pm SD, n = 6 (two independent samples, each with 3 replicates). Bovine Serum Albumin (BSA), having no disulfide isomerase activity, serves as a negative control. The rising trend observed for that of BSA can be inferred as owing to background signals; for example, the degradation of the substrate and the liberation of the fluorophore from the quencher.

Table 4. Welch's T-test for Figure 24. Analysis was performed using GraphPad Prism v.9.5.1 for macOS. Bovine Serum Albumin (BSA) is a negative control.			
	TrbB _{WT} vs. BSA	TrbB ₅₇₋₁₈₁ vs. BSA	TrbB _{WT} vs. TrbB ₅₇₋₁₈₁
Column B	50 μ M TrbB _{WT}	50 μ M TrbB ₅₇₋₁₈₁	50 μ M TrbB _{WT}
vs.	vs.	vs.	vs.
Column A	50 μ M BSA	50 μ M BSA	50 μ M TrbB ₅₇₋₁₈₁
Unpaired t test with Welch's correction			
P value	<0.0001	<0.0001	0.7633
P value summary	****	****	ns
Significantly different (P < 0.05)?	Yes	Yes	No
One- or two-tailed P value?	Two-tailed	Two-tailed	Two-tailed
Welch-corrected t, df	t=24.97, df=64.83	t=19.13, df=53.29	t=0.3023, df=71.67
How big is the difference?			
Mean of column A	9662	9662	33329
Mean of column B	32907	33329	32907
Difference between means (B - A) \pm SEM	23245 \pm 930.8	23667 \pm 1237	-422.2 \pm 1397
95% confidence interval	21386 to 25104	21185 to 26149	-3206 to 2362
R squared (eta squared)	0.9058	0.8728	0.001274
F test to compare variances			
F, DFn, Dfd	2.874, 40, 40	5.847, 40, 40	2.035, 40, 40
P value	0.0012	<0.0001	0.0272
P value summary	**	****	*
Significantly different (P < 0.05)?	Yes	Yes	Yes
Data analyzed			
Sample size, column A	41	41	41
Sample size, column B	41	41	41

3.6.3. TrbB does not bind GST-TrbI *in vitro* using BLI

Establishing that TrbB functions as a DI provides insights into its enzymatic function in and of itself but also its role in the context of the T4SS_F. However, it does not necessarily mean that it functions as a T4SS_F chaperone as previously hypothesized^{5,89,99,100,103}. To establish that TrbB is a protein chaperone, a direct question to investigate is: does it bind another T4SS_F protein? Therefore, a binding study between TrbB and GST-TrbI was performed using Bio-Layer Interferometry (**Fig. 25**), and no difference between the buffer signal and that from the GST-TrbI/TrbB was observed, suggesting that GST-TrbI and TrbB do not form a stable complex *in vitro*.

TrbB is conclusively a disulfide isomerase (**Fig. 24** of this paper and ref¹⁰²), which necessarily means TrbB catalyzes the proper formation of disulfide bonds. However, TrbI does not have any potential to form a disulfide bond because it only consists of one cysteine residue (**Fig. 16C** primary structure), which in principle, rules out the possibility that it is a client protein to a disulfide isomerase such as TrbB.

3.6.4. Changes in secondary structure level in the presence of TrbB

Attention was turned to another potential T4SS_F client protein for TrbB. Harris and Silverman (2004)⁹⁵ previously reported an interaction between TrbB and TraW, prompting us to focus on it. Before a binding study was performed, we tested whether the presence of TrbB leads to some observable structural changes using CD spectroscopy, a readily available biophysical method to investigate protein-protein interactions^{151–153}.

The presence of GST-TrbB in solution with His₆TraW causes some 2° structure changes compared to the CD spectra of TrbB or His₆TraW alone (**Fig. 26**). The observed increase in antiparallel β -sheets, β -turns, and “other” 2° structures in the GST-TrbB/His₆TraW mixture (**Fig. 26B**) can be interpreted as superimpositions of the signals from the two proteins on their own. However, the apparent significant decrease in α -helical content and complete disappearance of parallel β -sheets in the mixture compared to the proteins on their own provide some evidence that structural changes occur in solution.

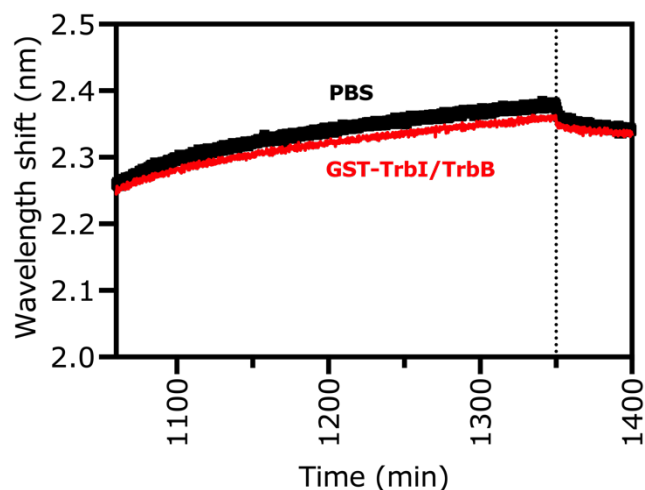


Figure 25. TrbB does not bind TrbI *in vitro*. There is no significant difference in wavelength shift between the protein (10 μ M TrbB and 10 μ M GST-TrbI) and buffer only samples. Bio-Layer Interferometry (BLI) study was performed using a Streptavidin biosensor tip against a biotinylated GST-TrbI prior to analysis. Shown above is the association (pre-dotted line) and dissociation (post-dotted line) steps.

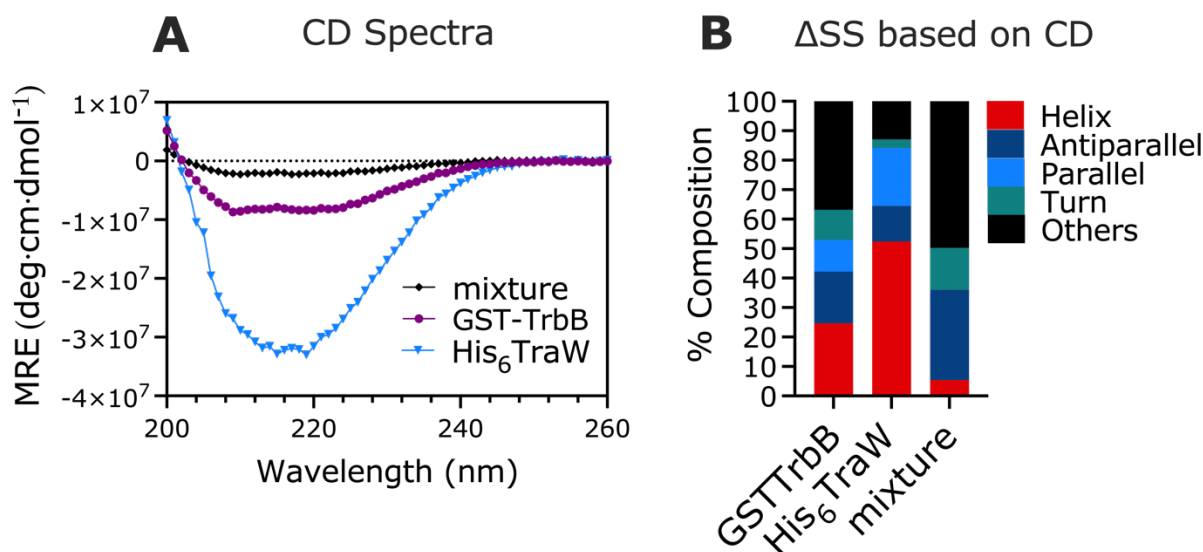


Figure 26. Observed changes in the secondary structure composition of GST-TrbB/His₆TraW compared to the proteins alone. (A) CD spectra of GST-TrbB/His₆TraW mixture, GST-TrbB alone, or His₆TraW alone. Protein samples were analyzed at a concentration of 5 μ M at 22°C. Data points are mean values from triplicate measurements, expressed in mean residue ellipticity (MRE). (B) Secondary structure composition prediction by BeStSel^{108,154}, with 3,10 helix, π -helix, β -bridge, bend, loops, irregular and/or disordered regions, grouped as “others”.

3.6.5. TrbB does not bind TraW *in vitro* using SEC, even in the presence of DTT

Following the slightly optimistic results from the CD experiment (Section 3.6.4), TrbB/TraW binding was investigated directly building on the work of Harris and Silverman (2004)⁹⁵ which utilized yeast two-hybrid analysis. To test whether TrbB binds TraW as previously reported, 19 μ M of TrbB and TraW as well as their truncation mutants were mixed in solution and allowed to equilibrate on ice for at least 30 min and analyzed using SEC (**Fig. 27**). There is no apparent difference in the elution profiles between the individual proteins (TrbB, His₆TraW, His₆ Δ TraW) and their mixture (TrbB/His₆TraW, TrbB/His₆ Δ TraW), which all elute within 8-9 min, though there is an increase in the magnitude of the peak absorbances when comparing the protein mixture and individual proteins attributable to the increased protein in solution. This result is consistent even when TrbB and His₆TraW are reduced with 2 mM dithiothreitol (DTT) (**Fig. 27B**).

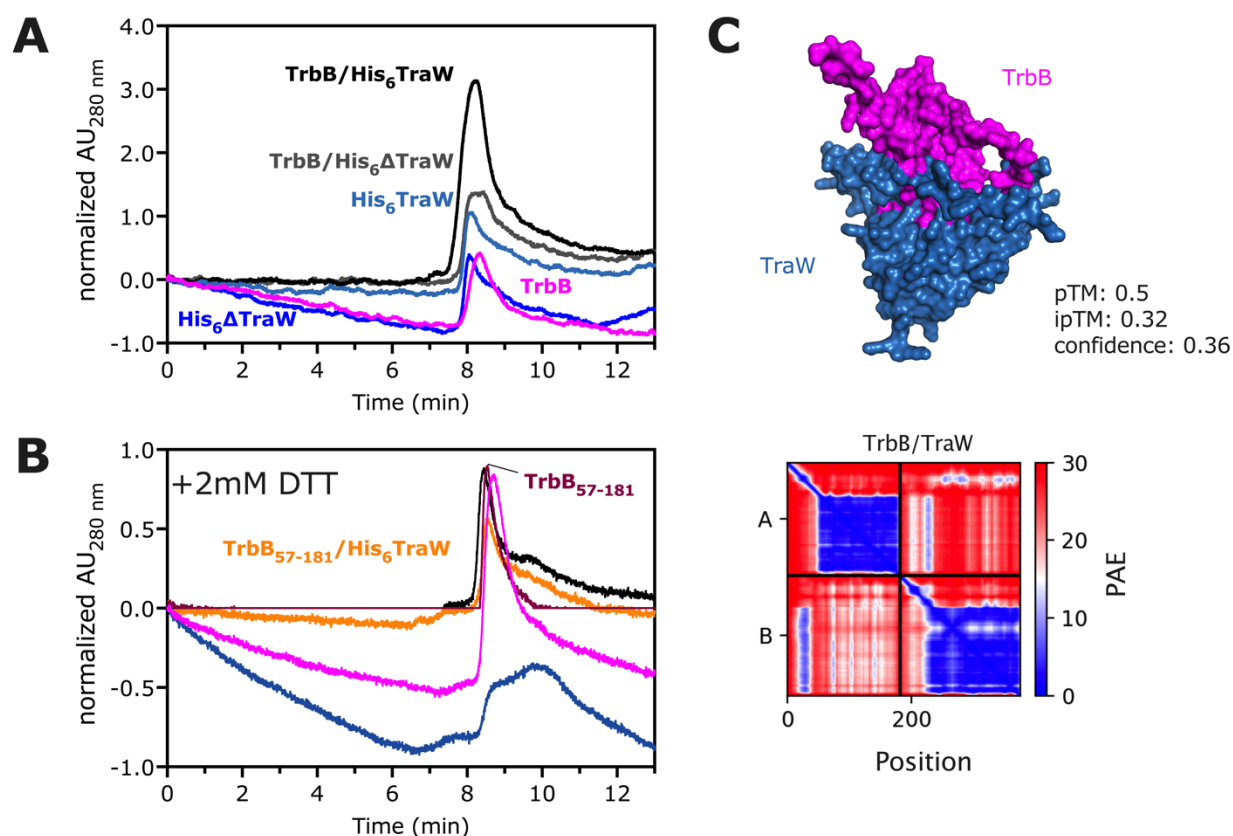


Figure 27. Preliminary investigation of the TrbB/His₆TraW putative interaction. (A) Size Exclusion Chromatography (SEC) elution profiles of TrbB, His₆TraW, His₆ Δ TraW, TrbB/His₆TraW, and TrbB/His₆ Δ TraW mixtures. (B) SEC elution profiles of TrbB, TrbB₅₇₋₁₈₁, His₆TraW, and TrbB₅₇₋₁₈₁/His₆TraW mixture in the presence of 2 mM DTT. Protein samples were incubated for 30 min on ice prior to SEC analysis on a pre-equilibrated column at a flow rate of 1.0 mL/min. (C)

Structural prediction of the putative TrbB/TraW complex (top) with model accuracy metrics (predicted Template Modelling scores, pTM; interface pTM, ipTM), and predicted alignment error (PAE) plot (bottom). The confidence metric, in the range [0,1], is calculated by $0.8 * ipTM + 0.2 * pTM$, weighing the metric for the reliability of binding interfaces (ipTM) more¹¹⁰.

Computational modelling using CF-AF2 multimer v3 was engaged to further understand the putative binding interaction between TrbB and TraW. The computational model support that TrbB and TraW could interact (**Fig. 27C** top); however, with high predicted alignment errors (PAE) (**Fig. 27C** bottom), a low predicted Template Modelling score (pTM, 0.5), interface pTM (ipTM, 0.32), and confidence¹¹⁰ (a weighted combination of ipTM and pTM, 0.36). The model predicts that the N-terminal residues of TrbB are involved in binding N-terminus of TraW (**Fig. 28A**).

TrbB is hypothesized to function as a protein chaperone for other T4SS_F proteins^{5,89,91}, yet it does not bind its previously reported binding partner, TraW, *in vitro*. The difference in our current finding (**Fig. 27A-B**) and that of Harris and Silverman (2004)⁹⁵ requires further empirical investigations to ascertain whether TrbB binds TraW. However, we wish to emphasize the following to support our preliminary evidence in this paper. Firstly, while yeast two-hybrid analysis is a powerful method to identify protein-protein interactions from the plethora of possibilities, the high intracellular traffic can lead to non-specific binding and the detection of a confounding protein-protein interaction, among other limitations^{155–157}. Secondly, we utilized purified proteins which may raise the concern that perhaps the proteins require other conditions present *in vivo*, such as the presence of lipid membranes in the microenvironment, to bind one another. Other concerns related to an *in vitro* binding study is that perhaps the concentrations of the proteins in our binding study were too low to observe an apparent difference in elution profiles, or that the time at which the proteins were equilibrated was not sufficient for stable complex formation. As such, we emphasize that the *in vitro* binding study we report here is preliminary because too many factors are yet to be established, *e.g.* assuming they bind, the time at which TrbB binds TraW can only be determined empirically and this is given by the k_{on} ¹⁵⁸. The use of more sensitive methods such as Bio-Layer Interferometry (which would also provide the k_{on} , k_{off} , and K_D of the interaction)¹⁵⁹ to test the TrbB/TraW binding is the subject of on-going efforts.

Thirdly, the His₆ tag may interfere with the protein-protein binding. Still, the CF-AF2 multimer v3 model for TrbB in complex with His₆TraW shows that the hexa-Histidine tag protrudes

from the binding interface and the confidence score of TrbB/TraW (0.36) and TrbB/His₆TraW (0.37) only differ by 0.01 (**Fig. 28B**), providing some evidence that the affinity tag does not interfere with the protein-protein binding interface. Interestingly, CF-AF2 multimer v3 predicts that the presence of a Hig₆-tag changes the predicted TrbB residues that bind TraW or His₆TraW. Unlike in the TrbB/TraW model, where N-terminal residues of TrbB bind N-terminal residues of TraW (**Fig. 28A, C**), C-terminal residues of TrbB bind N- and C-terminal residues of His₆TraW in the heterodimeric TrbB/His₆TraW model (**Fig. 28B, C**).

It is also important to note that TrbB is conclusively a disulfide isomerase (**Fig. 24** of this paper and ref¹⁰²), which necessarily means TrbB catalyzes the proper formation of disulfide bonds. Yet, TraW, like TrbI, does not have any potential to form a disulfide bond because it only consists of one cysteine residue (aa sequence shown in **Fig. 28C**), which in principle, rules out the possibility that it is a client protein to a disulfide isomerase such as TrbB. Nevertheless, we investigated whether it was important for TrbB to be reduced to bind TraW as TrbB is reported to be active only in its reduced form¹⁰². In the cell, DsbD is important for the catalytic activity of TrbB and DsbD's role is to reduce TrbB. *In lieu* of DsbD, TrbB was reduced using 2 mM DTT as previously performed to assay the activity of DsbC¹⁶⁰, now a well-studied prokaryotic disulfide isomerase. Even in its reduced form, TrbB does not bind TraW (**Fig. 27B**).

Contrary to our empirical evidence, computational modelling using CF-AF2 multimer v3 provides evidence that the N-terminus of TrbB binds the N-terminus of TraW (**Fig. 27C; Fig. 28A**) albeit with low confidence. Moreover, CF-AF2 multimer v3 predicts that there are differences between TrbB/TraW (N-terminus/N-terminus interaction) and TrbB/His₆TraW (C-terminus/N- and C-terminus interaction) (**Fig. 28**). Still, Mirdita and colleagues (2022)¹⁰⁹ report that a high confidence prediction by CF-AF2 Multimer v3 is backed by low Predicted Alignment Errors (PAEs) (blue) across the plot, and the presence of high PAEs in the plots for TrbB/TraW (**Fig. 27C** bottom) and TrbB/His₆TraW (**Fig. 28B**) are characteristic of a low confidence prediction. Additionally, the CF-AF2 heterodimeric models have pTM and ipTM scores (TrbB/TraW, pTM: 0.5, ipTM: 0.32; TrbB/His₆TraW, pTM: 0.48, ipTM: 0.34; **Fig. 28**) that are lower than the proposed threshold established by Yin and colleagues (2022) to be characteristic of an accurate model (pTM: 0.8).

Furthermore, the model confidence¹¹⁰ for each heterodimeric model are low; 0.36 for TrbB/TraW (**Fig. 27C**) and 0.37 for TrbB/His₆TraW (**Fig. 28**).

Therefore, the low confidence of the CF-AF2 models may weaken, but cannot invalidate, our empirical observation of no apparent difference in elution profiles and inference of no stable interaction between TrbB/His₆TraW and TrbB/His₆ΔTraW (a mutant lacking residues M1-S50) (**Fig. 28A-B**). Furthermore, while it is claimed that AF2 predicts bacterial protein-protein complexes more accurately, analysis by Bryant and colleagues show that only 60% of bacterial complexes among those complexes they analyzed are modelled correctly¹⁶¹. This percentage suggests that while complex modelling by AF2 is currently viewed among structural biologists with much confidence, it currently does not predict all protein-protein interactions definitively. Empirical investigations of protein-protein interactions remain necessary.

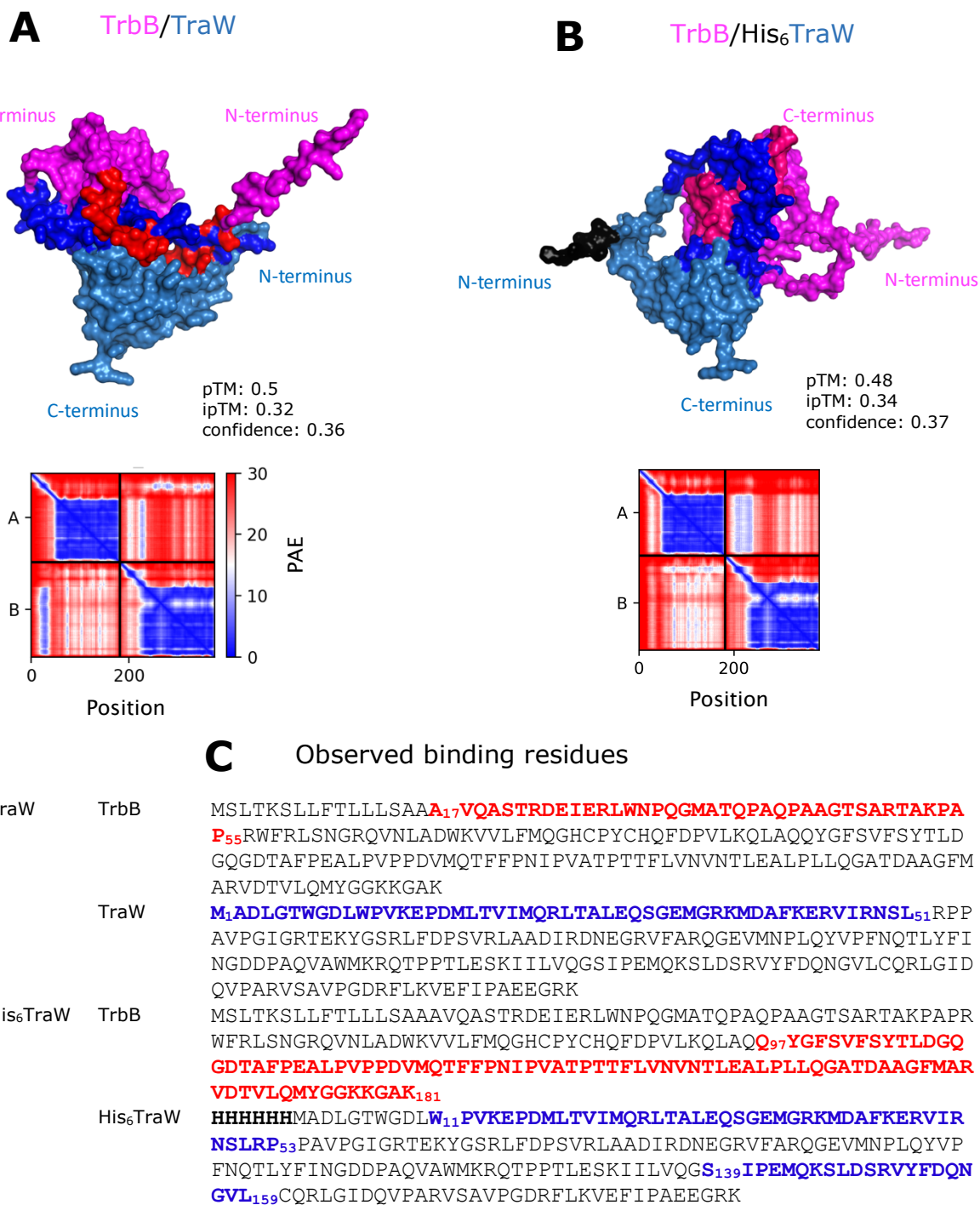


Figure 28. ColabFold-AlphaFold2 (CF-AF2) heterodimeric models and their observed binding residues. (A) TrbB/TraW model. N-terminal residues of TrbB (red) bind N-terminal residues of TraW (dark blue). (B) TrbB/His₆TraW model. C-terminal residues of TrbB (red) bind N- and C-terminal residues of TraW (dark blue). His₆ tag protrudes in TrbB/His₆TraW. Predicted Alignment Errors (PAE) plots and predicted TM scores (pTM) and interface pTM (ipTM) generated by CF-AF2 are shown below each model. CF-AF2 is described by Mirdita and colleagues (2022)¹⁰⁹.

3.6.6. Preliminary investigations on the chaperone activity of TrbB

In response to the inconclusive results, it was important to take a step back and understand whether TrbB functions as a chaperone at all before embarking on investigating its putative chaperone role of the T4SS_F once again. For its accessibility and its remarkable potential in forming disulfide bonds, lysozyme (which can form 4 disulfide bonds^{162,163}) was investigated as a TrbB client protein. To ensure that it is in its active form, TrbB was also reduced in 2 mM DTT overnight at 4°C prior to subsequent analyses, based on what was previously performed by Chen and colleagues¹⁶⁰ and in our SEC binding study (**Fig. 27B**).

To establish that a protein functions as a chaperone, it must observably increase a client protein's thermostability or decrease the client protein's propensity to aggregation^{164–166}. We monitored CD thermal denaturation profiles of a solution of GST-TrbB/Lysozyme reduced by DTT and the proteins on their own to determine whether the presence of GST-TrbB increases the observed T_M of lysozyme (**Fig. 29**). Notably, Lysozyme on its own has a T_M of 81°C, but when it is reduced by DTT, its T_M decreases to $45 \pm 5^\circ\text{C}$. When GST-TrbB is present in solution with Lysozyme in a reducing environment, the observed T_M of the mixture ($57 \pm 3^\circ\text{C}$) increases by 12°C compared to Lysozyme + DTT ($45 \pm 5^\circ\text{C}$), but only 1°C higher than GST-TrbB + DTT alone (T_M $56 \pm 2^\circ\text{C}$) (**Fig. 29**). Therefore, the results are inconclusive; it is difficult to ascertain whether the apparent increase in T_M in the mixture is a result of TrbB conferring thermostability to Lysozyme or the observed T_M is simply from the GST-TrbB in solution. Nevertheless, large secondary structure difference can be observed between the reduced Lysozyme (Lysozyme + DTT) and Lysozyme only, providing some evidence that the reducing environment changes the structure of Lysozyme, likely due to the unfolding of its tertiary and quaternary structures.

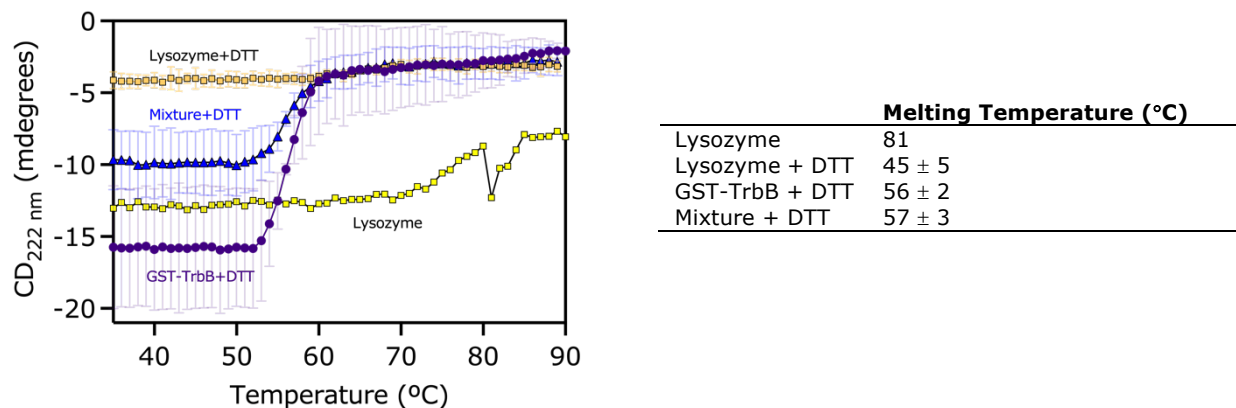


Figure 29. TrbB can increase the thermostability of Lysozyme. Protein sample melting temperatures, apart from Lysozyme without 2 mM DTT, are reported as mean \pm standard deviation from triplicate measurements. All protein samples were analyzed at a concentration of 5 μ M.

3.6.7. Other putative T4SS_F TrbB client proteins and CF-AF2 modelling

Attention was turned to another T4SS_F protein for two reasons. Firstly, understanding the interactome of TrbB provides rich information on how it behaves in the context of the T4SS_F apparatus. Secondly, having a known protein-protein binder can improve crystallization outcomes. It is widely practiced among crystallographers to use co-binding of interacting proteins to increase the stability of the recalcitrant protein and thus favor crystallization. For example, the complexing of a protein to antibodies, with its natural ligands, with engineered scaffolds, or its chaperones have been shown to favor protein crystallization^{167,168}.

Building on the hypothesis by Hemmis et al. (2011)¹⁰² and the premise that disulfide isomerases catalyze the proper formation of disulfide bonds of client proteins, we explored the potential of periplasmic TraU (which contains 11 cysteine residues) as a client protein. Presently, we analyzed the putative interaction using CF-AF2. Compared to TrbB/TraW, the PAE plot of TrbB/TraU (**Fig. 30**) appears to have lower PAE values for residues 200 and beyond (see more blue signals) and a higher pTM (0.61 vs. TrbB/TraW pTM: 0.5), providing optimism for future empirical investigations. However, the ipTM (TrbB/TraU ipTM: 0.25; TrbB/TraW ipTM: 0.32) and confidence (TrbB/TraU, 0.32; TrbB/TraW, 0.36) are lower. It is interesting to note, however, that the C81-XX-C84 moiety of TrbB is not in the interface that binds TraU, TraH, or TraN (**Fig. 30**), suggesting that TrbB may function as a disulfide isomerase chaperone in an action-at-a-distance mechanism or

that other parts of the protein function to bring the CXXC moiety to and from the substrate interface at a dynamic fashion. Overall, it is also important to consider that chaperones are known weak and transient binders¹⁶⁴. For example, some chaperones have dissociation constant (K_D) values in the order of 200 μM ¹⁶⁹.

3.6.8. Inconclusive findings as indications for the use of less stringent methods

Other state-of-the art methods and instrumentation to study protein-protein binding are now viewed with more confidence^{158,159,170}, but the insights provided by CD remain. The principle of CD must be re-iterated. Signals are based on how secondary structures absorb circularly polarized light¹⁷¹. That means it can be sensitive to slight structural changes. Therefore, the findings we report here, taken together by the lack of observation of a stable complex (see 3.6.5) and the typical reported behaviour of chaperones to be transient and weak binders, calls for the need for further binding interaction studies utilizing less stringent methods. A precedent of this in history is the work of the group of Richard H. Ebright with RNA polymerase and transcription in *E. coli*. They utilized Fluorescence Resonance Energy Transfer (FRET) to show the transient association and disassociation of the RNA polymerase σ -domain with the holoenzyme¹⁷². The use of FRET can probe the weak and transient binding of TrbB, as a chaperone protein¹⁶⁴, with its client proteins.

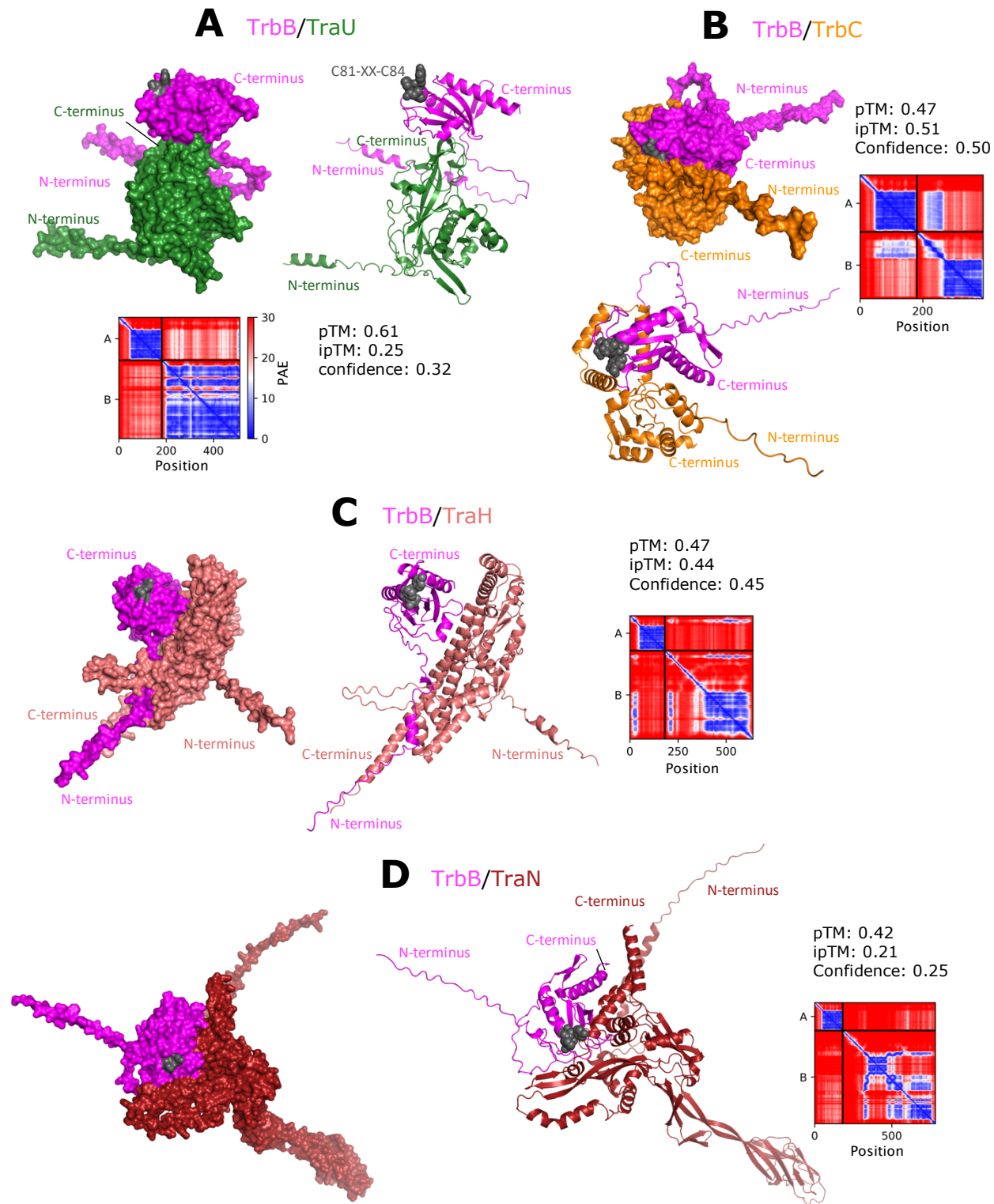


Figure 30. The active CXXC motif is not located at the binding interface of CF-AF2 heterodimeric models aside from TrbC. Surface and cartoon representations for (A) TrbB/TraU, (B) TrbB/TrbC, (C) TrbB/TraH, (D) TrbB/TraN are shown; C81-XX-C84 motif shown in dark grey. Indicated confidence metric, in the range [0,1], is given by $0.8 * ipTM + 0.2 * pTM$ ¹¹⁰, weighing the metric

for the reliability of binding interfaces (ipTM) more. CF-AF2 is described by Mirdita and colleagues (2022)¹⁰⁹.

4. CHAPTER FOUR: CONCLUSIONS & FUTURE WORK

Initial crystallization trials for GST-TrbI and GST-TrbB_{WT} have provided leads for conditions that facilitate protein crystallization. A low-resolution SEC-MALS-SAXS model of GST-TrbB_{WT} is reported, illustrating the dynamics of TrbB_{WT} when its disordered N-terminus is intact. Computational predictions indicate that residues M1-A20 comprise the signal sequence of TrbB_{WT} and residues S21-R56 are disordered. Accordingly, N-terminal residues of TrbB_{WT} were deleted in the truncation construct, TrbB₅₇₋₁₈₁, and spectroscopic methods support its relative physicochemical stability compared to TrbB_{WT}, providing optimism for future crystallization campaigns, and even structure solution by Nuclear Magnetic Resonance Spectroscopy. Lastly, Circular Dichroism spectroscopy provided empirical estimations of the secondary structures of GST-TrbI, GST-TrbB, TrbB_{WT}, and TrbB₅₇₋₁₈₁. GST-TraF secondary structures were also estimated for comparison with GST-TrbB_{WT} because previous research implicate the presence of thioredoxin domain in TrbB but only a thioredoxin-like domain in TraF due to the absence of the active CXXC motif in TraF.

The enzymatic disulfide isomerase activity of TrbB_{WT} was shown to be statistically different ($p < 0.0001$) compared to BSA (negative control) using a quenched-fluorophore *in vitro* assay, corroborating previous research. However, no significant change in disulfide isomerase activity ($p = 0.76$) was observed in the TrbB₅₇₋₁₈₁ construct compared to TrbB_{WT}, providing some evidence that the N-terminal M1-R56 residues are not required for *in vitro* function. The hypothesized chaperone function of TrbB led to the investigations of its binding activity and whether its presence in solution would confer increased thermostability to a potential client protein. Our investigations showed that TrbB does not form a stable complex with TrbI or TraW *in vitro*, despite previous reports that it binds TraW. Still, the presence of TrbB in solution in a TrbB/TraW mixture led to the observation of secondary structure changes compared to TrbB or TraW on their own, suggesting that TrbB may bind its client proteins weakly and transiently. Hence, the use of FRET to demonstrate TrbB's chaperone activity may be a promising avenue for future research.

The importance of the CXXC motif of TrbB, its chaperone activity, and the kinetics of its enzymatic activity need to be investigated in future work. Firstly, I hypothesize that the C81-XX-C84 moiety is necessary but not sufficient for DI catalytic activity. Thus, we need to understand its role and what extent of residues surrounding this moiety is necessary and sufficient for proper DI function. For future work, I propose a systematic mutational analysis on TrbB, and the design of a collection of: (a) truncation mutants, and (b) specific amino acid substitution mutants. Secondly, investigations into the chaperone activity of TrbB with other putative client proteins such as TraU and the kinetics of that activity will advance our understanding of T4SS_F and the protein that is central in its stability.

5. REFERENCES

- (1) Newsom, S.; Parameshwaran, H. P.; Martin, L.; Rajan, R. The CRISPR-Cas Mechanism for Adaptive Immunity and Alternate Bacterial Functions Fuels Diverse Biotechnologies. *Front Cell Infect Microbiol* **2021**, *10*.
- (2) Sgro, G. G.; Oka, G. U.; Souza, D. P.; Cenens, W.; Bayer-Santos, E.; Matsuyama, B. Y.; Bueno, N. F.; Dos Santos, T. R.; Alvarez-Martinez, C. E.; Salinas, R. K.; Farah, C. S. Bacteria-Killing Type IV Secretion Systems. *Front Microbiol* **2019**, *10* (1078), 1–20. <https://doi.org/10.3389/fmicb.2019.01078>.
- (3) Cavalli-Sforza, L. L. Forty Years Ago in “Genetics”: The Unorthodox Mating Behavior of Bacteria. *Genetics* **1992**, *132* (3), 635–637. <https://doi.org/10.1093/genetics/132.3.635>.
- (4) Sapp, J. Lucky. In *Genes, Germs and Medicine*; World Scientific, 2020; pp 31–40. https://doi.org/doi:10.1142/9789811225482_0004.
- (5) Bragagnolo, N.; Rodriguez, C.; Samari-Kermani, N.; Fours, A.; Korouzhdehi, M.; Lysenko, R.; Audette, G. F. Protein Dynamics in F-like Bacterial Conjugation. *Biomedicines* **2020**, *8* (9). <https://doi.org/10.3390/BIOMEDICINES8090362>.
- (6) Huang, L.; Wu, C.; Gao, H.; Xu, C.; Dai, M.; Huang, L.; Hao, H.; Wang, X.; Cheng, G. Bacterial Multidrug Efflux Pumps at the Frontline of Antimicrobial Resistance: An Overview. *Antibiotics* **2022**, *11* (4). <https://doi.org/10.3390/antibiotics11040520>.
- (7) Keeling, P. J.; Palmer, J. D. Horizontal Gene Transfer in Eukaryotic Evolution. *Nat Rev Genet* **2008**, *9* (8), 605–618. <https://doi.org/10.1038/nrg2386>.
- (8) Lurias, S. E.; Delbrock, M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* **1943**, *28*, 491–511.
- (9) Murray, A. Salvador Luria and Max Delbrück on Random Mutation and Fluctuation Tests. *Genetics* **2016**, *202* (2), 367–368. <https://doi.org/10.1534/genetics.115.186163>.
- (10) Avery, O. T.; Macleod, C. M.; McCarty, M. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types. *Journal of Experimental Medicine* **1943**, *79* (2), 137–158.
- (11) Gyles, C.; Boerlin, P. Horizontally Transferred Genetic Elements and Their Role in Pathogenesis of Bacterial Disease. *Vet Pathol* **2014**, *51* (2), 328–340. <https://doi.org/10.1177/0300985813511131>.
- (12) Blokesch, M. Protocols for Visualizing Horizontal Gene Transfer in Gram-Negative Bacteria Through Natural Competence. In *Hydrocarbon and Lipid Microbiology Protocols: Single-Cell and Single-Molecule Methods*; McGenity, T. J., Timmis, K. N., Nogales, B., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2016; pp 189–204. https://doi.org/10.1007/8623_2015_46.
- (13) Lederberg, J.; Tatum, E. Gene Recombination in Escherichia Coli. *Nature* **1946**, *158* (4016), 558. <https://doi.org/10.1038/158558a0>.
- (14) Sardesai, A. A.; Gowrishankar, J. Joshua Lederberg-a Remembrance. *J Genet* **2008**, *87* (3), 311–313.
- (15) Sapp, J. A Field of Their Own. In *Genes, Germs and Medicine*; World Scientific, 2020; pp 57–71. https://doi.org/doi:10.1142/9789811225482_0007.

- (16) Lederberg, E. M.; Lederberg, J. Genetic Studies Of Lysogenicity In Escherichia Coli. *Genetics* **1953**, *38* (1), 51–64. <https://doi.org/10.1093/genetics/38.1.51>.
- (17) Du Toit, A. CRISPR–Cas Enhances HGT by Transduction. *Nat Rev Microbiol* **2018**, *16* (4), 186. <https://doi.org/10.1038/nrmicro.2018.28>.
- (18) Zinder, N. D.; Lederberg, J. Genetic Exchange In Salmonella. *J Bacteriol* **1952**, *64* (5), 679–699. <https://doi.org/10.1128/jb.64.5.679-699.1952>.
- (19) McClintock, B. The Origin and Behavior of Mutable Loci in Maize. *Proceedings of the National Academy of Sciences* **1950**, *36* (6), 344–355. <https://doi.org/10.1073/pnas.36.6.344>.
- (20) Ravindran, S. Barbara McClintock and the Discovery of Jumping Genes. *Proc Natl Acad Sci U S A* **2012**, *109* (50), 20198–20199. <https://doi.org/10.1073/pnas.1219372109>.
- (21) Siguier, P.; Gourbeyre, E.; Chandler, M. Bacterial Insertion Sequences: Their Genomic Impact and Diversity. *FEMS Microbiol Rev* **2014**, *38* (5), 865–891. <https://doi.org/10.1111/1574-6976.12067>.
- (22) Ross, K.; Varani, A. M.; Snesrud, E.; Huang, H.; Alvarenga, D. O.; Zhang, J.; Wu, C.; McGann, P.; Chandlere, M. TnCentral: A Prokaryotic Transposable Element Database and Web Portal for Transposon Analysis. *mBio* **2021**, *12* (5). <https://doi.org/10.1128/mBio.02060-21>.
- (23) Green, E. R.; Meccas, J. Bacterial Secretion Systems: An Overview. *Microbiol Spectr* **2016**, *4* (1). <https://doi.org/10.1128/microbiolspec.vmbf-0012-2015>.
- (24) Trivedi, A.; Gosai, J.; Nakane, D.; Shrivastava, A. Design Principles of the Rotary Type 9 Secretion System. *Front Microbiol* **2022**, *13*.
- (25) Costa, T. R. D.; Felisberto-Rodrigues, C.; Meir, A.; Prevost, M. S.; Redzej, A.; Trokter, M.; Waksman, G. Secretion Systems in Gram-Negative Bacteria: Structural and Mechanistic Insights. *Nat Rev Microbiol* **2015**, *13* (6), 343–359. <https://doi.org/10.1038/nrmicro3456>.
- (26) Nicholson, K. R.; Champion, P. A. Bacterial Secretion Systems: Networks of Pathogenic Regulation and Adaptation in Mycobacteria and Beyond. *PLoS Pathog* **2022**, *18* (7), e1010610-.
- (27) Abrusci, P.; Vergara-Irigaray, M.; Johnson, S.; Beeby, M. D.; Hendrixson, D. R.; Roversi, P.; Friede, M. E.; Deane, J. E.; Jensen, G. J.; Tang, C. M.; Lea, S. M. Architecture of the Major Component of the Type III Secretion System Export Apparatus. *Nat Struct Mol Biol* **2013**, *20* (1), 99–104. <https://doi.org/10.1038/nsmb.2452>.
- (28) Sato, K.; Naito, M.; Yukitake, H.; Hirakawa, H.; Shoji, M.; McBride, M. J.; Rhodes, R. G.; Nakayama, K. A Protein Secretion System Linked to Bacteroidete Gliding Motility and Pathogenesis. *Proceedings of the National Academy of Sciences* **2010**, *107* (1), 276–281. <https://doi.org/10.1073/pnas.0912010107>.
- (29) Shrivastava, A.; Berg, H. C. A Molecular Rack and Pinion Actuates a Cell-Surface Adhesin and Enables Bacterial Gliding Motility. *Sci Adv* **2023**, *6* (10), eaay6616. <https://doi.org/10.1126/sciadv.aay6616>.
- (30) Olivia, S.; N, E. I.; Tobias, B.; Kerstin, K.; Barry, H. I.; Lutz, S. Type I Secretion Systems—One Mechanism for All? *Microbiol Spectr* **2019**, *7* (2), 10.1128/microbiolspec.psib-0003–2018. <https://doi.org/10.1128/microbiolspec.psib-0003-2018>.
- (31) Korotkov, K. V.; Maria, S. Architecture, Function, and Substrates of the Type II Secretion System. *EcoSal Plus* **2019**, *8* (2), 10.1128/ecosalplus.ESP-0034–2018. <https://doi.org/10.1128/ecosalplus.esp-0034-2018>.

- (32) Portaliou, A. G.; Tsois, K. C.; Loos, M. S.; Zorzini, V.; Economou, A. Type III Secretion: Building and Operating a Remarkable Nanomachine. *Trends Biochem Sci* **2016**, *41* (2), 175–189. <https://doi.org/https://doi.org/10.1016/j.tibs.2015.09.005>.
- (33) Grace, L. Y.; Bo, H.; J, C. P. Biological and Structural Diversity of Type IV Secretion Systems. *Microbiol Spectr* **2019**, *7* (2), 10.1128/microbiolspec.psib-0012–2018. <https://doi.org/10.1128/microbiolspec.psib-0012-2018>.
- (34) Enguo, F.; Nandini, C.; Gupta, U. D. B. R. K.; C, L. J.; Dirk, L. Type V Secretion Systems in Bacteria. *Microbiol Spectr* **2016**, *4* (1), 10.1128/microbiolspec.vmbf-0009–2015. <https://doi.org/10.1128/microbiolspec.vmbf-0009-2015>.
- (35) Meuskens, I.; Saragliadis, A.; Leo, J. C.; Linke, D. Type V Secretion Systems: An Overview of Passenger Domain Functions. *Front Microbiol* **2019**, *10*.
- (36) Cianfanelli, F. R.; Monlezun, L.; Coulthurst, S. J. Aim, Load, Fire: The Type VI Secretion System, a Bacterial Nanoweapon. *Trends Microbiol* **2016**, *24* (1), 51–62. <https://doi.org/https://doi.org/10.1016/j.tim.2015.10.005>.
- (37) Allsopp, L. P.; Bernal, P.; Nolan, L. M.; Filloux, A. Causalities of War: The Connection between Type VI Secretion System and Microbiota. *Cell Microbiol* **2020**, *22* (3), e13153. <https://doi.org/https://doi.org/10.1111/cmi.13153>.
- (38) Beckham, K. S. H.; Ritter, C.; Chojnowski, G.; Ziemianowicz, D. S.; Mullapudi, E.; Rettel, M.; Savitski, M. M.; Mortensen, S. A.; Kosinski, J.; Wilmanns, M. Structure of the Mycobacterial ESX-5 Type VII Secretion System Pore Complex. *Sci Adv* **2023**, *7* (26), eabg9923. <https://doi.org/10.1126/sciadv.abg9923>.
- (39) Bunduc, C. M.; Bitter, W.; Houben, E. N. G. Structure and Function of the Mycobacterial Type VII Secretion Systems. *Annu Rev Microbiol* **2020**, *74* (1), 315–335. <https://doi.org/10.1146/annurev-micro-012420-081657>.
- (40) Rickman, H. M.; Kamchedzera, W.; Schwalb, A.; Phiri, M. D.; Ruhwald, M.; Shanaube, K.; Dodd, P. J.; Houben, R. M. G. J.; Corbett, E. L.; MacPherson, P. Know Your Tuberculosis Epidemic—Is It Time to Add Mycobacterium Tuberculosis Immunoreactivity Back into Global Surveillance? *PLOS Global Public Health* **2022**, *2* (10), e0001208-.
- (41) Sujeet, B.; Nani, van G.; R, C. M.; Han, R. Curli Biogenesis: Bacterial Amyloid Assembly by the Type VIII Secretion Pathway. *EcoSal Plus* **2019**, *8* (2), 10.1128/ecosalplus.ESP-0037–2018. <https://doi.org/10.1128/ecosalplus.esp-0037-2018>.
- (42) Palmer, T.; Finney, A. J.; Saha, C. K.; Atkinson, G. C.; Sargent, F. A Holin/Peptidoglycan Hydrolase-Dependent Protein Secretion System. *Mol Microbiol* **2021**, *115* (3), 345–355. <https://doi.org/https://doi.org/10.1111/mmi.14599>.
- (43) Grossman, A. S.; Mauer, T. J.; Forest, K. T.; Goodrich-Blair, H. A Widespread Bacterial Secretion System with Diverse Substrates. *mBio* **2021**, *12* (4), 10.1128/mbio.01956-21. <https://doi.org/10.1128/mbio.01956-21>.
- (44) Ikryannikova, L. N.; Kurbatov, L. K.; Gorokhovets, N. V.; Zamyatnin, A. A. Contact-Dependent Growth Inhibition in Bacteria: Do Not Get Too Close! *Int J Mol Sci* **2020**, *21* (21). <https://doi.org/10.3390/ijms21217990>.
- (45) Aoki, S. K.; Pamma, R.; Hernday, A. D.; Bickham, J. E.; Braaten, B. A.; Low, D. A. Contact-Dependent Inhibition of Growth in Escherichia Coli. *Science (1979)* **2005**, *309* (5738), 1245–1248. <https://doi.org/10.1126/science.1115109>.

- (46) Beck, C. M.; Diner, E. J.; Kim, J. J.; Low, D. A.; Hayes, C. S. The F Pilus Mediates a Novel Pathway of CDI Toxin Import. *Mol Microbiol* **2014**, *93* (2), 276–290. <https://doi.org/10.1111/mmi.12658>.
- (47) Allen, J. P.; Hauser, A. R. Diversity of Contact-Dependent Growth Inhibition Systems of *Pseudomonas Aeruginosa*. *J Bacteriol* **2019**, *201* (14), e00776-18. <https://doi.org/10.1128/JB>.
- (48) Bartelli, N. L.; Sun, S.; Gucinski, G. C.; Zhou, H.; Song, K.; Hayes, C. S.; Dahlquist, F. W. The Cytoplasm-Entry Domain of Antibacterial CdiA Is a Dynamic α -Helical Bundle with Disulfide-Dependent Structural Features. *J Mol Biol* **2019**, *431* (17), 3203–3216. <https://doi.org/10.1016/j.jmb.2019.05.049>.
- (49) Jones, A. M.; Low, D. A.; Hayes, C. S. Can't You Hear Me Knocking: Contact-Dependent Competition and Cooperation in Bacteria. *Emerg Top Life Sci* **2017**, *1* (1), 75–83. <https://doi.org/10.1042/ETLS20160019>.
- (50) Jones, A. M.; Garza-Sánchez, F.; So, J.; Hayes, C. S.; Low, D. A. Activation of Contact-Dependent Antibacterial TRNase Toxins by Translation Elongation Factors. *Proc Natl Acad Sci U S A* **2017**, *114* (10), E1951–E1957. <https://doi.org/10.1073/pnas.1619273114>.
- (51) Beck, C. M.; Willett, J. L. E.; Cunningham, D. A.; Kim, J. J.; Low, D. A.; Hayes, C. S. CdiA Effectors from Uropathogenic *Escherichia Coli* Use Heterotrimeric Osmoporins as Receptors to Recognize Target Bacteria. *PLoS Pathog* **2016**, *12* (10). <https://doi.org/10.1371/journal.ppat.1005925>.
- (52) Maneewannakul, S.; Maneewannakul, K. Characterization, Localization, and Sequence of F Transfer Region Products: The Pilus Assembly Gene Product TraW and a New Product, Trbl. *J Bacteriol* **1992**, *174* (17), 5567–5574.
- (53) Norkunas, K.; Harding, R.; Dale, J.; Dugdale, B. Improving Agroinfiltration-Based Transient Gene Expression in *Nicotiana Benthamiana*. *Plant Methods* **2018**, *14* (1), 71. <https://doi.org/10.1186/s13007-018-0343-2>.
- (54) Wallden, K.; Rivera-Calzada, A.; Waksman, G. Type IV Secretion Systems: Versatility and Diversity in Function. *Cell Microbiol* **2010**, *12* (9), 1203–1212. <https://doi.org/10.1111/j.1462-5822.2010.01499.x>.
- (55) Waksman, G. From Conjugation to T4S Systems in Gram-Negative Bacteria: A Mechanistic Biology Perspective. *EMBO Rep* **2019**, *20* (2), e47012. <https://doi.org/https://doi.org/10.15252/embr.201847012>.
- (56) Lee Ventola, C. The Antibiotic Resistance Crisis Part 1: Causes and Threats. *P&T* **2015**, *40* (4), 277–283.
- (57) Spellberg, B.; Guidos, R.; Gilbert, D.; Bradley, J.; Boucher, H. W.; Scheld, W. M.; Bartlett, J. G.; Edwards, J. The Epidemic of Antibiotic-Resistant Infections: A Call to Action for the Medical Community from the Infectious Diseases Society of America. *Clinical Infectious Diseases* **2008**, *46* (2), 155–164. <https://doi.org/10.1086/524891>.
- (58) Klein, E. Y.; Van Boeckel, T. P.; Martinez, E. M.; Pant, S.; Gandra, S.; Levin, S. A.; Goossens, H.; Laxminarayan, R. Global Increase and Geographic Convergence in Antibiotic Consumption between 2000 and 2015. *Proceedings of the National Academy of Sciences* **2018**, *115* (15), E3463–E3470. <https://doi.org/10.1073/pnas.1717295115>.

- (59) Zavodszky, M.; Chen, C.-W.; Huang, J.-K.; Zolkiewski, M.; Wen, L.; Krishnamoorthi, R. Disulfide Bond Effects on Protein Stability: Designed Variants of Cucurbita Maxima Trypsin Inhibitor-V. *Protein Science* **2001**, *10* (1), 149–160. <https://doi.org/10.1110/ps.26801>.
- (60) Karimi, M.; Ignasiak, M. T.; Chan, B.; Croft, A. K.; Radom, L.; Schiesser, C. H.; Pattison, D. I.; Davies, M. J. Reactivity of Disulfide Bonds Is Markedly Affected by Structure and Environment: Implications for Protein Modification and Stability. *Sci Rep* **2016**, *6*. <https://doi.org/10.1038/srep38572>.
- (61) Depuydt, M.; Messens, J.; Collet, J.-F. How Proteins Form Disulfide Bonds. *Antioxid Redox Signal* **2011**, *15* (1), 49–66.
- (62) Sevier, C. S.; Kaiser, C. A. Formation and Transfer of Disulphide Bonds in Living Cells. *Nat Rev Mol Cell Biol* **2002**, *3* (11), 836–847. <https://doi.org/10.1038/nrm954>.
- (63) Christis, C.; Lubsen, N. H.; Braakman, I. Protein Folding Includes Oligomerization - Examples from the Endoplasmic Reticulum and Cytosol. *FEBS Journal* **2008**, *275* (19), 4700–4727. <https://doi.org/10.1111/j.1742-4658.2008.06590.x>.
- (64) Hatahet, F.; Ruddock, L. W. Protein Disulfide Isomerase: A Critical Evaluation of Its Function in Disulfide Bond Formation. *Antioxid Redox Signal* **2009**, *11* (11), 2807–2850. <https://doi.org/10.1089/ars.2009.2466>.
- (65) Matsusaki, M.; Kanemura, S.; Kinoshita, M.; Lee, Y. H.; Inaba, K.; Okumura, M. The Protein Disulfide Isomerase Family: From Proteostasis to Pathogenesis. *Biochim Biophys Acta Gen Subj* **2020**, *1864* (2). <https://doi.org/10.1016/j.bbagen.2019.04.003>.
- (66) Chacinska, A.; Pfannschmidt, S.; Wiedemann, N.; Kozjak, V.; Sanjuán Szklarz, L. K.; Schulze-Specking, A.; Truscott, K. N.; Guiard, B.; Meisinger, C.; Pfanner, N. Essential Role of Mia40 in Import and Assembly of Mitochondrial Intermembrane Space Proteins. *EMBO Journal* **2004**, *23* (19), 3735–3746. <https://doi.org/10.1038/sj.emboj.7600389>.
- (67) Mesecke, N.; Terziyska, N.; Kozany, C.; Baumann, F.; Neupert, W.; Hell, K.; Herrmann, J. M. A Disulfide Relay System in the Intermembrane Space of Mitochondria That Mediates Protein Import. *Cell* **2005**, *121* (7), 1059–1069. <https://doi.org/10.1016/j.cell.2005.04.011>.
- (68) Naoé, M.; Ohwa, Y.; Ishikawa, D.; Ohshima, C.; Nishikawa, S. I.; Yamamoto, H.; Endo, T. Identification of Tim40 That Mediates Protein Sorting to the Mitochondrial Intermembrane Space. *Journal of Biological Chemistry* **2004**, *279* (46), 47815–47821. <https://doi.org/10.1074/jbc.M410272200>.
- (69) Arunachalam, B.; Phan, U. T.; Geuze, H. J.; Cresswell, P.; Amos, D. B. Enzymatic Reduction of Disulfide Bonds in Lysosomes: Characterization of a Gamma-Interferon-Inducible Lysosomal Thiol Reductase (GILT). *PNAS* **2000**, *97* (2), 745–750.
- (70) de Duve, C.; Wattiaux, R. Functions of Lysosomes. *Annu Rev Physiol* **1966**, *28* (1), 435–492. <https://doi.org/10.1146/annurev.ph.28.030166.002251>.
- (71) Messens, J.; Collet, J. F.; Van Belle, K.; Brosens, E.; Loris, R.; Wyns, L. The Oxidase DsbA Folds a Protein with a Nonconsecutive Disulfide. *Journal of Biological Chemistry* **2007**, *282* (43), 31302–31307. <https://doi.org/10.1074/jbc.M705236200>.
- (72) Hooper, K. L.; Sheasley, S. L.; Gilbert, H. F.; Thorpe, C. Sulfhydryl Oxidase from Egg White. A Facile Catalyst for Disulfide Bond Formation in Proteins and Peptides. *Journal of Biological Chemistry* **1999**, *274* (32), 22147–22150. <https://doi.org/10.1074/jbc.274.32.22147>.

- (73) Hwang, C.; Sinskey, A. J.; Lodish, H. F. Oxidized Redox State of Glutathione in the Endoplasmic Reticulum. *Science* (1979) **1992**, 257 (5076), 1496–1502. <https://doi.org/10.1126/science.1523409>.
- (74) Silhavy, T. J.; Kahne, D.; Walker, S. The Bacterial Cell Envelope. *Cold Spring Harb Perspect Biol* **2010**, 2 (5). <https://doi.org/10.1101/cshperspect.a000414>.
- (75) Reardon-Robinson, M. E.; Ton-That, H. Disulfide-Bond-Forming Pathways in Gram-Positive Bacteria. *J Bacteriol* **2016**, 198 (5), 746–754. <https://doi.org/10.1128/JB.00769-15>.
- (76) Daniels, R.; Mellroth, P.; Bernsel, A.; Neiers, F.; Normark, S.; Von Heijne, G.; Henriques-Normark, B. Disulfide Bond Formation and Cysteine Exclusion in Gram-Positive Bacteria. *Journal of Biological Chemistry* **2010**, 285 (5), 3300–3309. <https://doi.org/10.1074/jbc.M109.081398>.
- (77) Dutton, R. J.; Boyd, D.; Berkmen, M.; Beckwith, J. Bacterial Species Exhibit Diversity in Their Mechanisms and Capacity for Protein Disulfide Bond Formation. *PNAS* **2008**, 105 (33), 11933–11938.
- (78) Bardwell, J. C. A.; McGovern, K.; Beckwith, J. Identification of a Protein Required for Disulfide Bond Formation In Vivo. *Cell* **1991**, 67, 581–569.
- (79) Heras, B.; Shouldice, S. R.; Totsika, M.; Scanlon, M. J.; Schembri, M. A.; Martin, J. L. DSB Proteins and Bacterial Pathogenicity. *Nat Rev Microbiol* **2009**, 7 (3), 215–225. <https://doi.org/10.1038/nrmicro2087>.
- (80) Anfinsen, C. B.; Haber, E. Studies on the Reduction and Re-Formation of Protein Disulfide Bonds. *J Biol Chem* **1961**, 236, 1361–1363.
- (81) Goldberger, R. F.; Epstein, C. J.; Anfinsen, C. B. Acceleration of Reactivation of Reduced Bovine Pancreatic Ribonuclease by a Microsomal System from Rat Liver. *J Biol Chem* **1963**, 238, 628–635. [https://doi.org/10.1016/s0021-9258\(18\)81309-6](https://doi.org/10.1016/s0021-9258(18)81309-6).
- (82) Freedman, R. B. Native Disulphide Bond Formation in Protein Biosynthesis: Evidence for the Role of Protein Disulphide Isomerase. *Trends Biochem Sci* **1984**, 9 (10), 438–441. [https://doi.org/https://doi.org/10.1016/0968-0004\(84\)90152-X](https://doi.org/https://doi.org/10.1016/0968-0004(84)90152-X).
- (83) Sato, Y.; Inaba, K. Disulfide Bond Formation Network in the Three Biological Kingdoms, Bacteria, Fungi and Mammals. *FEBS Journal* **2012**, 279 (13), 2262–2271. <https://doi.org/10.1111/j.1742-4658.2012.08593.x>.
- (84) McArthur, A. G.; Knodler, L. A.; Silberman, J. D.; Davids, B. J.; Gillin, F. D.; Sogin, M. L. The Evolutionary Origins of Eukaryotic Protein Disulfide Isomerase Domains: New Evidence from the Amitochondriate Protist Giardia Lamblia. *Mol. Biol. Evol* **2001**, 18 (8), 1455–1463.
- (85) Arutyunov, D.; Arenson, B.; Manchak, J.; Frost, L. S. F Plasmid TraF and TraH Are Components of an Outer Membrane Complex Involved in Conjugation. *J Bacteriol* **2010**, 192 (6), 1730–1734. <https://doi.org/10.1128/JB.00726-09>.
- (86) Lawley, T. D.; Klimke, W. A.; Gubbins, M. J.; Frost, L. S. F Factor Conjugation Is a True Type IV Secretion System. *FEMS Microbiol Lett* **2003**, 224 (1), 1–15. [https://doi.org/10.1016/S0378-1097\(03\)00430-0](https://doi.org/10.1016/S0378-1097(03)00430-0).
- (87) Bradley, D. E.; Taylor, D. E.; Cohen, D. R. Specification of Surface Mating Systems among Conjugative Drug Resistance Plasmids in Escherichia Coli K-12. *J Bacteriol* **1980**, 143 (3), 1466–1470. <https://doi.org/10.1128/jb.143.3.1466-1470.1980>.
- (88) Christie, P. J. The Mosaic Type IV Secretion Systems. *EcoSal Plus* **2016**, 7 (1), 10.1128/ecosalplus.ESP-0020–2015. <https://doi.org/10.1128/ecosalplus.esp-0020-2015>.

- (89) Elton, T. C.; Holland, S. J.; Frost, L. S.; Hazes, B. F-like Type IV Secretion Systems Encode Proteins with Thioredoxin Folds That Are Putative DsbC Homologues. *J Bacteriol* **2005**, *187* (24), 8267–8277. <https://doi.org/10.1128/JB.187.24.8267-8277.2005>.
- (90) Virolle, C.; Goldlust, K.; Djermoun, S.; Bigot, S.; Lesterlin, C. Plasmid Transfer by Conjugation in Gram-Negative Bacteria: From the Cellular to the Community Level. *Genes (Basel)* **2020**, *11* (11). <https://doi.org/10.3390/genes11111239>.
- (91) Frost, L. S.; Ippen-Ihler, K.; Skurray, R. A. Analysis of the Sequence and Gene Products of the Transfer Region of the F Sex Factor. *Microbiol Rev* **1994**, *58* (2), 162–210.
- (92) Curtiss, R.; Caro, L. G.; Allison, D. P. Early Stages of Conjugation in Escherichia Coli. *J Bacteriol* **1969**, *100* (2), 1091–1104.
- (93) Arutyunov, D.; Frost, L. S. F Conjugation: Back to the Beginning. *Plasmid* **2013**, *70* (1), 18–32. <https://doi.org/https://doi.org/10.1016/j.plasmid.2013.03.010>.
- (94) Pelat, C.; Kardaš-Stoma, L.; Birgand, G.; Ruppé, E.; Schwarzing, M.; Andremon, A.; Lucet, J.-C.; Yazdanpanah, Y. Hand Hygiene, Cohorting, or Antibiotic Restriction to Control Outbreaks of Multidrug-Resistant Enterobacteriaceae. *Infect Control Hosp Epidemiol* **2016**, *37* (3), 272–280. <https://doi.org/DOI: 10.1017/ice.2015.284>.
- (95) Harris, R. L.; Silverman, P. M. Tra Proteins Characteristic of F-like Type IV Secretion Systems Constitute an Interaction Group by Yeast Two-Hybrid Analysis. *J Bacteriol* **2004**, *186* (16), 5480–5485. <https://doi.org/10.1128/JB.186.16.5480-5485.2004>.
- (96) Lawley, T. D.; Klimke, W. A.; Gubbins, M. J.; Frost, L. S. F Factor Conjugation Is a True Type IV Secretion System. *FEMS Microbiol Lett* **2003**, *224* (1), 1–15. [https://doi.org/10.1016/S0378-1097\(03\)00430-0](https://doi.org/10.1016/S0378-1097(03)00430-0).
- (97) Missiakas, D.; Raina, S. Protein Folding in the Bacterial Periplasm. *J Bacteriol* **1997**, *179* (8), 2465–2471. <https://doi.org/10.1128/jb.179.8.2465-2471.1997>.
- (98) Kinch, L. N.; Baker, D.; Grishin, N. V. Deciphering a Novel Thioredoxin-like Fold Family. *Proteins: Structure, Function, and Bioinformatics* **2003**, *52* (3), 323–331. <https://doi.org/https://doi.org/10.1002/prot.10425>.
- (99) Collet, J.-F.; Bardwell, J. C. A. Oxidative Protein Folding in Bacteria. *Mol Microbiol* **2002**, *44* (1), 1–8. <https://doi.org/https://doi.org/10.1046/j.1365-2958.2002.02851.x>.
- (100) Audette, G. F.; Holland, S. J.; Elton, T. C.; Manchak, J.; Hayakawa, K.; Frost, L. S.; Hazes, B. Crystallization and Preliminary Diffraction Studies of TraF, a Component of the Escherichia Coli Type IV Secretory System. *Acta Crystallographica Section D* **2004**, *60* (11), 2025–2027. <https://doi.org/10.1107/S0907444904020724>.
- (101) Collet, J. F.; Bardwell, J. C. A. Oxidative Protein Folding in Bacteria. *Mol Microbiol* **2002**, *44* (1), 1–8. <https://doi.org/10.1046/j.1365-2958.2002.02851.x>.
- (102) Hemmis, C. W.; Berkmen, M.; Eser, M.; Schildbach, J. F. TrbB from Conjugative Plasmid F Is a Structurally Distinct Disulfide Isomerase That Requires DsbD for Redox State Maintenance. *J Bacteriol* **2011**, *193* (18), 4588–4597. <https://doi.org/10.1128/JB.00351-11>.
- (103) Anthony, K. G.; Sherburne, C.; Sherburne, R.; Frost, L. S. The Role of the Pilus in Recipient Cell Recognition during Bacterial Conjugation Mediated by F-like Plasmids. *Mol Microbiol* **1994**, *13* (6), 939–953. <https://doi.org/https://doi.org/10.1111/j.1365-2958.1994.tb00486.x>.

- (104) Addgene. Plasmid Cloning by PCR. <https://blog.addgene.org/plasmid-cloning-by-pcr> (accessed 2023-11-14).
- (105) Sayers, E. W.; Bolton, E. E.; Brister, J. R.; Canese, K.; Chan, J.; Comeau, D. C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; Madej, T.; Marchler-Bauer, A.; Lanczycki, C.; Lathrop, S.; Lu, Z.; Thibaud-Nissen, F.; Murphy, T.; Phan, L.; Skripchenko, Y.; Tse, T.; Wang, J.; Williams, R.; Trawick, B. W.; Pruitt, K. D.; Sherry, S. T. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **2022**, *50* (D1), D20–D26. <https://doi.org/10.1093/nar/gkab1112>.
- (106) Bruno, A. E.; Charbonneau, P.; Newman, J.; Snell, E. H.; So, D. R.; Vanhoucke, V.; Watkins, C. J.; Williams, S.; Wilson, J. Classification of Crystallization Outcomes Using Deep Convolutional Neural Networks. *PLoS One* **2018**, *13* (6). <https://doi.org/10.1371/journal.pone.0198883>.
- (107) Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nat Protoc* **2007**, *1* (6), 2876–2890. <https://doi.org/10.1038/nprot.2006.202>.
- (108) Micsonai, A.; Moussong, É.; Wien, F.; Boros, E.; Vadász, H.; Murvai, N.; Lee, Y. H.; Molnár, T.; Réfrégiers, M.; Goto, Y.; Tantos, Á.; Kardos, J. BeStSel: Webserver for Secondary Structure and Fold Prediction for Protein CD Spectroscopy. *Nucleic Acids Res* **2022**, *50* (W1), W90–W98. <https://doi.org/10.1093/nar/gkac345>.
- (109) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making Protein Folding Accessible to All. *Nat Methods* **2022**, *19* (6), 679–682. <https://doi.org/10.1038/s41592-022-01488-1>.
- (110) Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Židek, A.; Bates, R.; Blackwell, S.; Yim, J.; Ronneberger, O.; Bodenstein, S.; Zielinski, M.; Bridgland, A.; Potapenko, A.; Cowie, A.; Tunyasuvunakool, K.; Jain, R.; Clancy, E.; Kohli, P.; Jumper, J.; Hassabis, D. Protein Complex Prediction with AlphaFold-Multimer. *bioRxiv* **2021**. <https://doi.org/10.1101/2021.10.04.463034>.
- (111) The PyMOL Molecular Graphics System, Version 2.5. Schrödinger, LLC.
- (112) Hopkins, J. B.; Gillilan, R. E.; Skou, S. BioXTAS RAW: Improvements to a Free Open-Source Program for Small-Angle X-Ray Scattering Data Reduction and Analysis. *J Appl Crystallogr* **2017**, *50* (5), 1545–1553. <https://doi.org/10.1107/S1600576717011438>.
- (113) Manalastas-Cantos, K.; Konarev, P. V.; Hajizadeh, N. R.; Kikhney, A. G.; Petoukhov, M. V.; Molodenskiy, D. S.; Panjkovich, A.; Mertens, H. D. T.; Gruzinov, A.; Borges, C.; Jeffries, C. M.; Svergun, D. I.; Franke, D. ATSAS 3.0: Expanded Functionality and New Tools for Small-Angle Scattering Data Analysis. *J Appl Crystallogr* **2021**, *54* (1), 343–355. <https://doi.org/10.1107/S1600576720013412>.
- (114) Luft, J. R.; Wolfley, J. R.; Snell, E. H. What's in a Drop? Correlating Observations and Outcomes to Guide Macromolecular Crystallization Experiments. *Cryst Growth Des* **2011**, *11* (3), 651–663. <https://doi.org/10.1021/cg1013945>.
- (115) Lynch, M. L.; Snell, M. E.; Potter, S. A.; Snell, E. H.; Bowman, S. E. J. 20 Years of Crystal Hits: Progress and Promise in Ultrahigh-Throughput Crystallization Screening. *Acta Crystallogr D Struct Biol* **2023**, *79*, 198–205. <https://doi.org/10.1107/S2059798323001274>.
- (116) McPherson, A.; Gavira, J. A. Introduction to Protein Crystallization. *Acta Crystallographica Section F: Structural Biology Communications* **2014**, *70* (1), 2–20. <https://doi.org/10.1107/S2053230X13033141>.

- (117) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
- (118) Yin, R.; Feng, B. Y.; Varshney, A.; Pierce, B. G. Benchmarking AlphaFold for Protein Complex Modeling Reveals Accuracy Determinants. *Protein Science* **2022**, *31* (8). <https://doi.org/10.1002/pro.4379>.
- (119) Gavshina, A. V.; Marynich, N. K.; Khrenova, M. G.; Solovyev, I. D.; Savitsky, A. P. The Role of Cysteine Residues in the Allosteric Modulation of the Chromophore Phototransformations of Biphotochromic Fluorescent Protein SAASoti. *Sci Rep* **2021**, *11* (1), 24314. <https://doi.org/10.1038/s41598-021-03634-9>.
- (120) Khanal, P.; Jia, Z.; Yang, X. Cysteine Residues Are Essential for Dimerization of Hippo Pathway Components YAP2L and TAZ. *Sci Rep* **2018**, *8* (1), 3485. <https://doi.org/10.1038/s41598-018-21828-6>.
- (121) Rotoli, S. M.; Jones, J. L.; Caradonna, S. J. Cysteine Residues Contribute to the Dimerization and Enzymatic Activity of Human Nuclear DUTP Nucleotidohydrolase (NDut). *Protein Science* **2018**, *27* (10), 1797–1809. <https://doi.org/https://doi.org/10.1002/pro.3481>.
- (122) Bell, M. R.; Engleka, M. J.; Malik, A.; Strickler, J. E. To Fuse or Not to Fuse: What Is Your Purpose? *Protein Sci* **2013**, *22* (11), 1466–1477. <https://doi.org/10.1002/pro.2356>.
- (123) Maru, Y.; Afar, D. E.; Witte, O. N.; Shibuya, M. The Dimerization Property of Glutathione S-Transferase Partially Reactivates Bcr-Abl Lacking the Oligomerization Domain*. *Journal of Biological Chemistry* **1996**, *271* (26), 15353–15357. <https://doi.org/https://doi.org/10.1074/jbc.271.26.15353>.
- (124) Fabrini, R.; De Luca, A.; Stella, L.; Mei, G.; Orioni, B.; Ciccone, S.; Federici, G.; Lo Bello, M.; Ricci, G. Monomer–Dimer Equilibrium in Glutathione Transferases: A Critical Re-Examination. *Biochemistry* **2009**, *48* (43), 10473–10482. <https://doi.org/10.1021/bi901238t>.
- (125) Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. X-Ray Solution Scattering (SAXS) Combined with Crystallography and Computation: Defining Accurate Macromolecular Structures, Conformations and Assemblies in Solution. *Q Rev Biophys* **2007**, *40* (3), 191–285. <https://doi.org/10.1017/S0033583507004635>.
- (126) Jacques, D. A.; Trewella, J. Small-Angle Scattering for Structural Biology - Expanding the Frontier While Avoiding the Pitfalls. *Protein Science* **2010**, *19* (4), 642–657. <https://doi.org/10.1002/pro.351>.
- (127) Da Vela, S.; Svergun, D. I. Methods, Development and Applications of Small-Angle X-Ray Scattering to Characterize Biological Macromolecules in Solution. *Curr Res Struct Biol* **2020**, *2*, 164–170. <https://doi.org/10.1016/j.crstbi.2020.08.004>.
- (128) Rambo, R. P.; Tainer, J. A. Accurate Assessment of Mass, Models and Resolution by Small-Angle Scattering. *Nature* **2013**, *496* (7446), 477–481. <https://doi.org/10.1038/nature12070>.

- (129) Powers, K. T.; Gildenberg, M. S.; Washington, M. T. Modeling Conformationally Flexible Proteins With X-Ray Scattering and Molecular Simulations. *Comput Struct Biotechnol J* **2019**, *17*, 570–578. <https://doi.org/https://doi.org/10.1016/j.csbj.2019.04.011>.
- (130) Kikhney, A. G.; Svergun, D. I. A Practical Guide to Small Angle X-Ray Scattering (SAXS) of Flexible and Intrinsically Disordered Proteins. *FEBS Lett* **2015**, *589* (19), 2570–2577. <https://doi.org/10.1016/j.febslet.2015.08.027>.
- (131) Smilgies, D.-M.; Foltá-Stogniew, E. Molecular Weight-Gyration Radius Relation of Globular Proteins: A Comparison of Light Scattering, Small-Angle X-Ray Scattering and Structure-Based Data. *J Appl Crystallogr* **2015**, *48* (Pt 5), 1604–1606. <https://doi.org/10.1107/s1600576715015551>.
- (132) Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. ExPASy: The Proteomics Server for in-Depth Protein Knowledge and Analysis. *Nucleic Acids Res* **2003**, *31* (13), 3784–3788. <https://doi.org/10.1093/nar/gkg563>.
- (133) Piiadov, V.; Ares de Araújo, E.; Oliveira Neto, M.; Craievich, A. F.; Polikarpov, I. SAXSMoW 2.0: Online Calculator of the Molecular Weight of Proteins in Dilute Solution from Experimental SAXS Data Measured on a Relative Scale. *Protein Science* **2019**, *28* (2), 454–463. <https://doi.org/https://doi.org/10.1002/pro.3528>.
- (134) Rambo, R. P.; Tainer, J. A. Accurate Assessment of Mass, Models and Resolution by Small-Angle Scattering. *Nature* **2013**, *496* (7446), 477–481. <https://doi.org/10.1038/nature12070>.
- (135) Franke, D.; Jeffries, C. M.; Svergun, D. I. Machine Learning Methods for X-Ray Scattering Data Analysis from Biomacromolecular Solutions. *Biophys J* **2018**, *114* (11), 2485–2492. <https://doi.org/https://doi.org/10.1016/j.bpj.2018.04.018>.
- (136) Hajizadeh, N. R.; Franke, D.; Jeffries, C. M.; Svergun, D. I. Consensus Bayesian Assessment of Protein Molecular Mass from Solution X-Ray Scattering Data. *Sci Rep* **2018**, *8* (1), 7204. <https://doi.org/10.1038/s41598-018-25355-2>.
- (137) Norris, V.; Oláh, J.; Krylov, S. N.; Uversky, V. N.; Ovádi, J. The Sherpa Hypothesis: Phenotype-Preserving Disordered Proteins Stabilize the Phenotypes of Neurons and Oligodendrocytes. *NPJ Syst Biol Appl* **2023**, *9* (1). <https://doi.org/10.1038/s41540-023-00291-8>.
- (138) Wright, P. E.; Dyson, H. J. Intrinsically Disordered Proteins in Cellular Signalling and Regulation. *Nat Rev Mol Cell Biol* **2015**, *16* (1), 18–29. <https://doi.org/10.1038/nrm3920>.
- (139) Sharma, R.; Raduly, Z.; Miskei, M.; Fuxreiter, M. Fuzzy Complexes: Specific Binding without Complete Folding. *FEBS Lett* **2015**, *589* (19PartA), 2533–2542. <https://doi.org/https://doi.org/10.1016/j.febslet.2015.07.022>.
- (140) Uversky, V. N. Multitude of Binding Modes Attainable by Intrinsically Disordered Proteins: A Portrait Gallery of Disorder-Based Complexes. *Chem. Soc. Rev.* **2011**, *40* (3), 1623–1634. <https://doi.org/10.1039/C0CS00057D>.
- (141) Uversky, V. N. Targeting Intrinsically Disordered Proteins in Neurodegenerative and Protein Dysfunction Diseases: Another Illustration of the D2 Concept. *Expert Rev Proteomics* **2010**, *7* (4), 543–564. <https://doi.org/10.1586/epr.10.36>.
- (142) Irvine, G. B.; El-Agnaf, O. M.; Shankar, G. M.; Walsh, D. M. Protein Aggregation in the Brain: The Molecular Basis for Alzheimer's and Parkinson's Diseases. *Molecular Medicine* **2008**, *14* (7), 451–464. <https://doi.org/10.2119/2007-00100.Irvine>.

- (143) Trejo-Lopez, J. A.; Yachnis, A. T.; Prokop, S. Neuropathology of Alzheimer's Disease. *Neurotherapeutics* **2022**, *19* (1), 173–185. <https://doi.org/10.1007/s13311-021-01146-y>.
- (144) Mochizuki, H.; Choong, C.-J.; Masliah, E. A Refined Concept: α -Synuclein Dysregulation Disease. *Neurochem Int* **2018**, *119*, 84–96. <https://doi.org/https://doi.org/10.1016/j.neuint.2017.12.011>.
- (145) Oldfield, C. J.; Xue, B.; Van, Y. Y.; Ulrich, E. L.; Markley, J. L.; Dunker, A. K.; Uversky, V. N. Utilization of Protein Intrinsic Disorder Knowledge in Structural Proteomics. *Biochim Biophys Acta Proteins Proteom* **2013**, *1834* (2), 487–498. <https://doi.org/10.1016/j.bbapap.2012.12.003>.
- (146) Bragagnolo, N.; Audette, G. F. Solution Characterization of the Dynamic Conjugative Entry Exclusion Protein TraG. *Structural Dynamics* **2022**, *9* (6), 064702. <https://doi.org/10.1063/4.0000171>.
- (147) Teufel, F.; Almagro Armenteros, J. J.; Johansen, A. R.; Gíslason, M. H.; Pihl, S. I.; Tsirigos, K. D.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 6.0 Predicts All Five Types of Signal Peptides Using Protein Language Models. *Nat Biotechnol* **2022**, *40* (7), 1023–1025. <https://doi.org/10.1038/s41587-021-01156-3>.
- (148) Derewenda, Z. S. The Use of Recombinant Methods and Molecular Engineering in Protein Crystallization. *Methods* **2004**, *34* (3), 354–363. <https://doi.org/10.1016/j.ymeth.2004.03.024>.
- (149) Marion, D. An Introduction to Biological NMR Spectroscopy. *Molecular and Cellular Proteomics* **2013**, *12* (11), 3006–3025. <https://doi.org/10.1074/mcp.O113.030239>.
- (150) Kwan, A. H.; Mobli, M.; Gooley, P. R.; King, G. F.; MacKay, J. P. Macromolecular NMR Spectroscopy for the Non-Spectroscopist. *FEBS Journal* **2011**, *278* (5), 687–703. <https://doi.org/10.1111/j.1742-4658.2011.08004.x>.
- (151) He, S.; Zhao, J.; Zhang, Y.; Zhu, Y.; Li, X.; Cao, X.; Ye, Y.; Li, J.; Sun, H. Effects of Low-PH Treatment on the Allergenicity Reduction of Black Turtle Bean (*Phaseolus Vulgaris* L.) Lectin and Its Mechanism. *J Agric Food Chem* **2021**, *69* (4), 1379–1390. <https://doi.org/10.1021/acs.jafc.0c06524>.
- (152) Zhou, M.; Li, Q.; Wang, R. Current Experimental Methods for Characterizing Protein–Protein Interactions. *ChemMedChem* **2016**, *11* (8), 738–756. <https://doi.org/https://doi.org/10.1002/cmdc.201500495>.
- (153) Abrie, J. A.; Molero, C.; Ariño, J.; Strauss, E. Complex Stability and Dynamic Subunit Interchange Modulates the Disparate Activities of the Yeast Moonlighting Proteins Hal3 and Vhs3. *Sci Rep* **2015**, *5* (1), 15774. <https://doi.org/10.1038/srep15774>.
- (154) Micsonai, A.; Wien, F.; Kernya, L.; Lee, Y. H.; Goto, Y.; Réfrégiers, M.; Kardos, J. Accurate Secondary Structure Prediction and Fold Recognition for Circular Dichroism Spectroscopy. *Proc Natl Acad Sci U S A* **2015**, *112* (24), E3095–E3103. <https://doi.org/10.1073/pnas.1500851112>.
- (155) Stellberger, T.; Häuser, R.; Baiker, A.; Pothineni, V. R.; Haas, J.; Uetz, P. Improving the Yeast Two-Hybrid System with Permutated Fusions Proteins: The Varicella Zoster Virus Interactome. *Proteome Sci* **2010**, *8* (8), 1–9.
- (156) Rajagopala, S. V.; Hughes, K. T.; Uetz, P. Benchmarking Yeast Two-Hybrid Systems Using the Interactions of Bacterial Motility Proteins. *Proteomics* **2009**, *9* (23), 5296–5302. <https://doi.org/10.1002/pmic.200900282>.

- (157) Huang, H.; Jedynak, B. M.; Bader, J. S. Where Have All the Interactions Gone? Estimating the Coverage of Two-Hybrid Protein Interaction Maps. *PLoS Comput Biol* **2007**, *3* (11), 2155–2174. <https://doi.org/10.1371/journal.pcbi.0030214>.
- (158) Jarmoskaite, I.; AlSadhan, I.; Vaidyanathan, P. P.; Herschlag, D. How to Measure and Evaluate Binding Affinities. *Elife* **2020**, *9*, e57264. <https://doi.org/10.7554/eLife.57264>.
- (159) Petersen, R. Strategies Using Bio-Layer Interferometry Biosensor Technology for Vaccine Research and Development. *Biosensors (Basel)* **2017**, *7* (4), 49. <https://doi.org/10.3390/bios7040049>.
- (160) Chen, J.; Song, J. L.; Zhang, S.; Wang, Y.; Cui, D. F.; Wangt, C. C. Chaperone Activity of DsbC. *Journal of Biological Chemistry* **1999**, *274* (28), 19601–19605. <https://doi.org/10.1074/jbc.274.28.19601>.
- (161) Bryant, P.; Pozzati, G.; Elofsson, A. Improved Prediction of Protein-Protein Interactions Using AlphaFold2. *Nat Commun* **2022**, *13* (1). <https://doi.org/10.1038/s41467-022-28865-w>.
- (162) Muttathukattil, A. N.; Singh, P. C.; Reddy, G. Role of Disulfide Bonds and Topological Frustration in the Kinetic Partitioning of Lysozyme Folding Pathways. *J Phys Chem B* **2019**, *123* (15), 3232–3241. <https://doi.org/10.1021/acs.jpcc.9b00739>.
- (163) Bocedi, A.; Fabrini, R.; Pedersen, J. Z.; Federici, G.; Iavarone, F.; Martelli, C.; Castagnola, M.; Ricci, G. The Extreme Hyper-Reactivity of Selected Cysteines Drives Hierarchical Disulfide Bond Formation in Serum Albumin. *FEBS J* **2016**, *283* (22), 4113–4127. <https://doi.org/https://doi.org/10.1111/febs.13909>.
- (164) Arhar, T.; Shkedi, A.; Nadel, C. M.; Gestwicki, J. E. The Interactions of Molecular Chaperones with Client Proteins: Why Are They so Weak? *Journal of Biological Chemistry* **2021**, *297* (5). <https://doi.org/10.1016/j.jbc.2021.101282>.
- (165) Garbuz, D. G.; Sverchinsky, D.; Davletshin, A.; Margulis, B. A.; Mitkevich, V.; Kulikov, A. M.; Evgen'ev, M. B. The Molecular Chaperone Hsp70 from the Thermotolerant Diptera Species Differs from the Drosophila Paralog in Its Thermostability and Higher Refolding Capacity at Extreme Temperatures. *Cell Stress Chaperones* **2019**, *24* (6), 1163–1173. <https://doi.org/10.1007/s12192-019-01038-3>.
- (166) Ronzetti, M.; Baljinnyam, B.; Jalal, I.; Pal, U.; Simeonov, A. Application of Biophysical Methods for Improved Protein Production and Characterization: A Case Study on an High-Temperature Requirement A-Family Bacterial Protease. *Protein Science* **2022**, *31* (12). <https://doi.org/10.1002/pro.4498>.
- (167) Anderson, J.; Ptashne, M.; Harrison, S. C. Cocystals of the DNA-Binding Domain of Phage 434 Repressor and a Synthetic Phage 434 Operator (Protein-DNA Cocystals/x-Ray Crystallography/B-DNA/Helical Diffraction). *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 1307–1311.
- (168) Griffin, L.; Lawson, A. Antibody Fragments as Tools in Crystallography. *Clin Exp Immunol* **2011**, *165* (3), 285–291. <https://doi.org/10.1111/j.1365-2249.2011.04427.x>.
- (169) Lee, J. H.; Zhang, D.; Hughes, C.; Okuno, Y.; Sekhar, A.; Cavagnero, S. Heterogeneous Binding of the SH3 Client Protein to the DnaK Molecular Chaperone. *Proc Natl Acad Sci U S A* **2015**, *112* (31), E4206–E4215. <https://doi.org/10.1073/pnas.1505173112>.