Imperial College London

MRC Centre for Molecular Bacteriology and Infection

Investigating the protein subcomplexes from a conjugative Type IV Secretion System

Himani Amin

Centre for Molecular Bacteriology and Infection

Department of Life Sciences

Imperial College London

This thesis is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

February 2023

Statement of Originality

I declare that the work presented in this thesis is my own and that the contributions of others are acknowledged and appropriately referenced in text. The results described in this PhD thesis have not been submitted for any other degree, diploma or qualification.

Copyright declaration

List of publications arising from this work

AMIN, H., ILANGOVAN, A. & COSTA, T. R. D. 2021. Architecture of the outer-membrane core complex from a conjugative type IV secretion system. *Nat Commun,* 12**,** 6834.

LOCKWOOD, D. C., AMIN, H., COSTA, T. R. D. & SCHROEDER, G. N. 2022. The Legionella pneumophila Dot/Icm type IV secretion system and its effectors. *Microbiology (Reading),* 168.

PATKOWSKI, J. B., DAHLBERG, T., AMIN, H., GAHLOT, D. K., VIJAYRAJRATNAM, S., VOGEL, J. P., FRANCIS, M. S., BAKER, J. L., ANDERSSON, M. & COSTA, T. R. D. 2023. The F-pilus biomechanical adaptability accelerates conjugative dissemination of antimicrobial resistance and biofilm formation. Nat Commun, 14, 1879.

Acknowledgements

First and foremost, I would like to take this opportunity to thank Dr Tiago Costa for being an incredible and supportive supervisor. I would not be the scientist I am today without your teachings. It has been a great honour to watch the Costa lab grow, and it makes me sad to leave, which is a testament to the friendly work environment you have nurtured. Thank you so much for being my mentor; I will continue coming to you for advice even after this PhD.

A special thank you to the Costa Lab, both past and present members. You all have no idea what your friendship has meant to me, and I have always been excited to come into work because it meant working alongside you all. Thank you to Dr Alejandro Peña-Ontalvilla for his kindness, always helping me find focus, and for showing me that nothing is impossible. A huge thank you to my best lab buddy Simona Draganova for her unwavering friendship, thoughtfulness, and belief in my abilities as a scientist. A massive thank you to Jonasz Patkowski and Ambre Bexter-The Crumble Gang! Jonasz, thank you for making me laugh hysterically and for being my lab bestie. Ambre, thank you for your words of encouragement, sisterhood and introducing me to Gilmore Girls. A big thank you to the members of the Dionne Lab for making the office and lab space so fun to work in. Furthermore, thank you to the members of the Filloux Lab for their insight during group meetings, especially Prof Alain Filloux for his guidance and, Dr Patricia Paracuellos-Torrecilla for her friendship and for teaching me everything I know about the ÄKTA.

I would like to extend my appreciation to my PRP committee, Prof Gad Frankel and Dr Harry Low, for their expert guidance. I would like to acknowledge my undergraduate and postgraduate students for their hard work on preliminary experiments. I am also hugely grateful to our EM facility manager, Paul Simpson, for teaching me everything I know about the electron microscope. A big thank you to our collaborator Dr Aravindan Ilangovan without whom the structure of the F-OMCC could not have been possible.

Of course, none of this could have been possible without the support of my family and friends. They have been a constant source of love and motivation throughout the years. Shoutout to my brother Kandarp Amin for his support and for being such a brilliant role model. Most importantly, to my partner Arjun Sharda – thank you so much for your unwavering belief in me, emotional support, love and for everything you do. I truly am very blessed to have you all in my life.

Finally, I'd like to thank my Viva examiners for finding time to read my thesis.

This thesis is dedicated to my parents, Sunil and Falguni Amin. You both are my greatest source of inspiration and my role models.

Abstract

Type IV secretion system (T4SS) are versatile nanomachines that enable the efficient transport of substrates in bacteria. In general, they are formed from two major membrane embedded subassemblies: an outer membrane core complex (OMCC) and an inner membrane complex (IMC). The conjugative T4SS encoded by the F plasmid is of particular interest due to its clinical relevance as it facilitates the spread of antibiotic resistance amongst bacterial population. Despite its importance, atomic details of the F-T4SS structure and protein-protein interactions were rudimentary which in turn precludes thorough understanding of how conjugation is orchestrated. Therefore, this thesis aimed to improve knowledge on the F-T4SS by studying the structure of the F-OMCC and investigating other proteins the complex may interact with.

After optimising the detergent solubilisation of the F-OMCC expressed from the pED208 F-like plasmid, and improving the purification of the complex, a cryo-EM dataset was collected. Using single particle analysis, the structure was solved with an overall resolution of 3.3 Å. The F-OMCC is formed from two concentric rings which have two distinct symmetries. The outer ring adopts 13-fold symmetry whereas the inner ring showed 17-fold symmetry, together they form a 2.1 MDa complex. The atomic models of TraB, TraK and TraV were built into the structure, and they revealed a unique stoichiometric arrangement. Interestingly, TraV and TraK proteins were found to adopt two different conformations within the outer ring. TraV and TraB were found to accommodate the symmetry mismatch by existing in both F-OMCC rings, and also appeared to confer flexibility. This makes the F-OMCC a dynamic complex which is likely to have important implications in the pilus and T4SS activity during conjugation.

The interactions between the F-OMCC and other Tra/Trb proteins were also investigated to decipher how the concerted dynamics of the pilus may be connected to the complex. A potential interaction between F-OMCC and the proteins TraH and TraN was observed by pull-down assays. Furthermore, initial work on TraG found that it seems to assemble as a high order oligomer in solution. The results are reminiscent of a hexameric protein which may be functionally important.

Together, the findings of this thesis reveal novel insights into the F-T4SS and its subassemblies. The approach used to purify the F-OMCC and study the interactions will act as the basis of future work on the F-T4SS and is directly applicable to the other protein complexes within the conjugative nanomachine.

Contents

List of Figures

List of Tables

Abbreviations

2D	Two dimensional
3D	Three dimensional
AAD	all-α-domain
ApR	Ampicillin resistance
ASU	Asymmetric unit
bp	base pair
Cag PAI	Cag pathogenicity island
Choline	n-dodecylphosphocholine
CMC	Critical micelle concentration
Cryo-EM	Cryogenic electron microscopy
Cryo-ET	Cryogenic electron tomography
CTD	C-terminal domain
CTF	Contrast transfer function
ddH_2O	Double distilled H_2O
DDM	n-Dodecyl β-D-Maltoside
DED	Direct electron detector
DMNG	Decyl Maltose Neopentyl Glycol
DNA	Deoxyribonucleic acid
DSS	Disuccinimidyl suberate
Dtr	DNA transfer and replication factor proteins
Eex	Surface exclusion
ER	Endoplasmic Reticulum
FA	Formaldehyde
FT	Flow through
FSC	Fourier Shell Correlation
GGI	Gonococcal genetic island
GDN	Glycol-Diosgenin
HGT	Horizontal gene transfer
ICE	Integrated conjugative elements
IHF	Integrated host factor
IM	Inner membrane
IMC	Inner membrane complex
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
LB	Lysogeny broth
LCV	Legionella containing vacuole
LDAO	Lauryldimethylamine-N-oxide
LonCEM	London Consortium Electron Microscopy
Mpf	Mating pair formation
Mps	Mating pair stabilisation
MWCO	Molecular weight cut-off
NEB	New England Biolab
NS-EM	Negative stain electron microscopy
NTD	N-terminal domain
OBG	n-Octyl-β-D-Glucoside
OD_{600}	Optical density at 600 nm

OM	Outer membrane
OMCC	Outer membrane core complex
$OMCC_{OR}$	Outer membrane core complex outer ring
$OMCC_{IR}$	Outer membrane core complex inner ring
OmpA	Outer membrane protein A
oriT	Origin of transfer
PCR	Polymerase chain reaction
PG	Peptidoglycan
PR	Periplasmic ring
RT	Room temperature
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Sfx	Surface exclusion
SmR	Spectinomycin resistance
T4CP	Type IV coupling protein
T1SS	Type I secretion system
T2SS	Type II secretion system
T3SS	Type III secretion system
T4SS	Type IV secretion system
T5SS	Type V secretion system
T6SS	Type VI secretion system
T7SS	Type VII secretion system
T8SS	Type VIII secretion system
T9SS	Type IX secretion system
Tat	Twin-arginine
TBS	Tris buffered saline
Tra	Transfer
Ti	Tumour inducing
UA	Uranyl acetate
Vir	Virulence
X-Tfe	Xanthomonadales-like T4SS effectors
XVIPCD	*Xanthomonas* VirD4 interacting protein conserved domain

Chapter 1: Introduction

1.1 Bacterial secretion is vital for adaptation and survival

Essential to life is the ability to secrete macromolecules across the cell membrane. This phenomenon is particularly vital for bacteria to enable them to adapt to the environment and survive. While some molecules may diffuse across the membrane in a passive manner, more complex macromolecules such as polypeptides and genetic material are transported across the bacterial membrane using dedicated nanomachines termed secretion systems (Costa et al., 2015). These systems are crucial to bacteria and are involved in various different processes. Some machines provide physiological support by enabling scavenging for nutrients and motility (Green and Mecsas, 2016). Others are used by bacterial pathogens to secrete effectors that aid invasion and colonisation by manipulation of the host, and toxins which directly intoxicates the target cell. Therefore, these secretion systems have been implicated in an array of diseases. Dedicated secretion systems also provide a competitive advantage within an environmental niche by secretion of toxins that can kill other microorganisms (Green and Mecsas, 2016). There are different types of secretion systems, and their architecture differs depending on whether the substrate they translocate crosses across a single or double membrane (Maffei et al., 2017) (Figure 1.1).

In diderm (Gram-negative) bacteria, secretion is achieved either by one step or by two steps (Maffei et al., 2017). The machines involved in the one step secretion, span the entire envelope, and enable the transport of the substrate across both the inner and outer membranes from the cytoplasm. This includes the Type 1, 3, 4, and 6 secretion systems (T1SS, T3SS, T4SS and T6SS) which transport molecules either to the extracellular space or even into host cells. Alternatively, some substrates are delivered across the membrane envelope via two separate steps, which firstly requires the delivery of the molecule across the inner membrane and to the periplasm by either the Sec or twin-arginine (Tat) secretion systems. From here, a trans-envelope apparatus facilitates the second transfer step by delivering the substrate across the outer membrane. This consists of the following: Type 2, 5, 7 and 9 secretion systems (T2SS, T5SS, T7SS and T9SS). In monoderm (Gram-positive) bacteria, extracellular secretion is achieved by the Sec/Tat machineries. Additionally, some Gram-positive microorganisms employ separate secretion apparatuses, such as the T7SS which was first characterised in *Mycobacterium tuberculosis* (Stanley et al., 2003). Despite the metabolic burden, several bacteria express numerous secretion systems emphasising the evolutionary benefits of being able to secrete substrates.

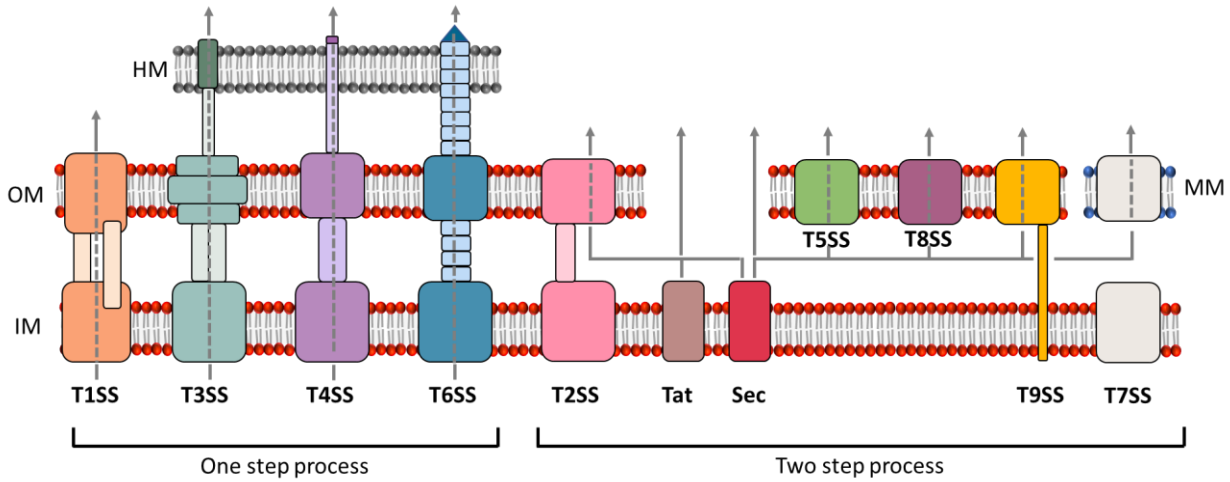

Figure 1.1: Overview of bacterial secretion systems. Substrates are secreted either by a one step or two step process. The Sec and Tat pathways are involved in transporting substrates across the inner membrane (IM) in Gram-negative bacteria and across the cytoplasmic membrane in Gram-positive bacteria. The T1SS, T3SS, T4SS and T6SS can deliver substrates across both the IM and outer membrane (OM) in one step, with T3SS, T4SS and T6SS also injecting molecules into another cell across the host membrane (HM). Alternatively, the T2SS, T5SS, T8SS and T9SS secrete substrates from the periplasm to the extracellular matrix. The T7SS is found in mycobacteria and molecules are transported across the mycomembrane (MM).

Amongst the 11 currently established types of secretion systems, the T4SSs are particularly versatile. Along with being involved in the secretion of toxins and effectors, T4SSs have the unique ability to transport DNA across the membrane envelope and into target cells (Costa et al., 2015). This encourages bacterial genome plasticity and promotes the emergence and persistence of antibiotic resistance. T4SS play a central role in disease. It is worth noting that in this Chapter, the term toxin has been utilised to describe potent molecules which harm other bacterial or host cells, while the term effector has been used to denote a molecule which can affect host biological processes. Apart from their role in spreading antibiotic resistance, as effector translocators, these machineries are also responsible for the cause of a range of different disease by pathogenic bacteria such as Legionella pneumonia (*Legionella pneumophila*), Q fever (*Coxiella burnetii*), whooping cough (*Bordetella pertussis*) and gastric cancer (*Helicobacter pylori*) (Sgro et al., 2019). Thus, highlighting the importance of studying the T4SSs. Consequently, T4SSs will be the focus of this Chapter.

1.2 The functional classification of T4SSs

The T4SS are incredibly versatile machines and are involved in a variety of different functions. They can be broadly divided into groups based on their function which gives rise to four major functional classes: conjugation, effector translocation, toxin delivery and lastly DNA export and import (Figure 1.2).

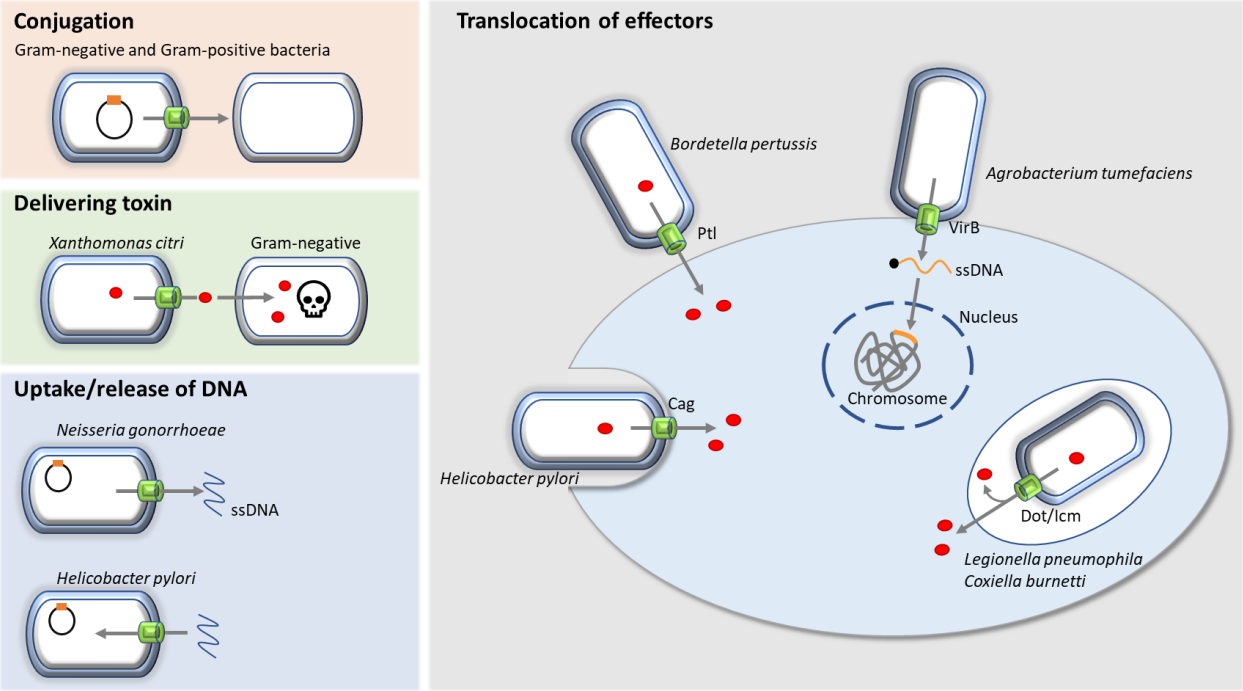

Figure 1.2: The functional classes the T4SS can be grouped into. The T4SS are involved in conjugation, toxin delivery, uptake and release of DNA, and translocation of effectors.

1.2.1 Conjugative transfer machines

Bacterial conjugation describes the phenomenon whereby single stranded DNA (ssDNA) is transferred from a donor to a recipient cell in a unidirectional manner. It is one of the major forms of horizontal gene transfer (HGT), alongside transduction and transformation via natural competence. While the latter depends greatly on the involvement of the recipient cell, conjugation is driven by the donor cell which participates in the search and recognition of a suitable recipient cell, followed by the transfer of the ssDNA (de la Cruz et al., 2010). It represents the largest functional class of T4SSs, as it is present in almost all bacterial species (Gram-negative and Gram-positive), and homologs of two ATPases in the T4SS (VirB4 and VirD4) have even been identified in some archaeal species (Wagner et al., 2017). These conjugative apparatuses are encoded not only on self-transmissible plasmids, but also on integrated conjugative

elements (ICE) and conjugative transposon (Frost and Koraimann, 2010). Overall, the conjugative elements accumulate virulence determinants and resistance to antibiotics and heavy metals.

A key feature of the conjugative elements is the expression and assembly of a T4SS. Prominent examples of such T4SSs includes those encoded by the conjugative plasmids F, R388 and pKM101. The VirB/D4 system from *Agrobacterium tumefaciens* can also be regarded as a conjugative machinery as it is involved in the transfer of oncogenic DNA to plant cells. Nonetheless, due to its role in pathogenesis and host-pathogen contact dependence, it has also been seen grouped with the effector translocation machineries (Grohmann et al., 2018).

The mechanism underpinning conjugation can be described by four general steps. The first requires the elaboration of an extracellular filament called the pilus which contacts the recipient cell and retracts to bring it close to the donor cell. The second step is mechanistically conserved. It involves the formation of the relaxosome which occurs when the relaxase enzyme and accessory factors, collectively known as the DNA transfer and replication factor proteins (Dtr), bind to the origin of transfer (*oriT*) sequence on the conjugative plasmid. This results in the nicking of the plasmid and the relaxase binding to the 5'-end of the DNA. In the third step, the relaxase bound DNA is recruited by a specialised ATPase called the Type IV coupling protein (T4CP) and is shuttled to the membrane embedded T4SS. In the last step, powered by the ATPases associated with the T4SS, the ssDNA is transported through the machinery and into the recipient cell.

1.2.2 The effector translocation machinery

The T4SS involved in effector translocation are used by bacteria to deliver effectors directly into the cytosol of eukaryotic cells which aid bacterial colonisation and survival inside the host (Alvarez-Martinez and Christie, 2009). The bacteria that utilise these T4SSs are medically relevant pathogens. For example, *Helicobacter pylori* (gastric cancer), *Legionella pneumophila* (Legionnaire's pneumonia), *Bordetella pertussis* (whooping cough) and *Coxiella burnetti* (Q fever) (Backert and Meyer, 2006). Described only in Gram-negative bacteria, the effector translocation T4SSs lack the Dtr proteins which are involved in the DNA processing. Nevertheless, most of them still encode for a T4CP that recruits the cytosolic substrates. An exception to this is the PtI T4SS assembled by *B. pertussis* which lacks a T4CP and instead uses the general secretory pathway. Furthermore, as mentioned in section 1.2.1, the *A. tumefaciens* VirB/D4 system is sometimes grouped into this functional group which does encode for Dtr proteins. The number

of effectors that the T4SS secretes greatly varies. For instance, the Cag T4SS assembled by *H. pylori* has only one identified effector protein whereas the Dot/Icm T4SS encoded by *L. pneumophila* delivers over 300 effectors (Isaac and Isberg, 2014, Cover et al., 2020). This results in diversity amongst the effector translocation machineries and enables establishment of host specific replicative niches.

1.2.3 The toxin delivery machines which allow interbacterial competition

Machines that are involved in toxin delivery between bacteria were first discovered in 2015 when it was observed that *Xanthomonas citri* can kill neighbouring bacteria. This function was attributed to the encoded T4SS (Souza et al., 2015). In a contact-dependent manner, the *X. citri* T4SS transports toxic effectors into the target prokaryotic cell which includes both Gram-negative and Gram-positive bacteria. This results in cellular lysis and thereby, provides the bacteria releasing the toxin a competitive growth advantage. A related T4SS has since been identified in *Stenotrophomonas maltophila* (Bayer-Santos et al., 2019). Recently, another T4SS involved in the delivery of toxins has been discovered in *Pseudomonas putida* strain IsoF (Purtschert-Montenegro et al., 2022). The encoded T4SS not only allows *P. putida* IsoF to kill bacterial competitors, but also enables the bacteria to invade existing biofilms and in a T4SS-dependent manner can protect tomato plants from infection from the plant pathogen*, Ralstonia solanacearum*.

1.2.4 The DNA export and import machines

While one may consider conjugative T4SSs as export machines, the systems found in this functional class are completely contact-independent. There are two main T4SSs that are found in this category: the *H. pylori* ComB system and the *Neisseria gonorrhoeae* DNA release system (Alvarez-Martinez and Christie, 2009). The ComB system carries out natural transformation by mediating importing exogenous DNA from the extracellular milieu (Hofreuter et al., 2001). It is believed to function independently of the effector translocating Cag T4SS (Fischer et al., 2020). The *N. gonorrhoeae* DNA release system transports ssDNA into the extracellular environment, which contributes to the genetic diversity in this species (Grohmann et al., 2018). It is encoded on the gonococcal genetic island (GGI) and is considered as a F-like T4SS due to similarity between proteins needed for DNA donation (Kohler et al., 2013, Grohmann et al., 2018).

1.3 Other T4SS classifications

Regardless of the functional class a T4SS belongs to, in Gram-negative bacteria, diversity also exists at the level of genetic organisation with differences in the number and arrangement of shared homologous genes. Therefore, the T4SSs have been broadly grouped into two main classes: type IVA (T4ASS) and type IVB (T4BSS) (Christie and Vogel, 2000). The T4ASS represent machines that are made of proteins similar in number and composition to the paradigmatic *A. tumefaciens* VirB/D4 system. The T4ASS is composed of 12 proteins and therefore has also been referred to as minimised systems (Costa et al., 2021). This includes the T4SS encoded by *Xanthomonas citri* (X-T4SS) and the conjugative plasmids R388 ($T4SS_{R388}$) and pKM101 ($T4SS_{pKM101}$). The prototype for the T4BSS is the *L. pneumophila* Dot/Icm apparatus. Due to the complexity imparted by the large number of proteins involved, T4BSS are also referred to as expanded systems (Costa et al., 2021).

Another classification scheme that exists is based originally on the incompatibility of the conjugative plasmid that encodes the T4SS thereby, giving rise to three main groups: F-type (IncF), P-type (IncP) and IncI (I-type) (Lawley et al., 2003). The P-type T4SS can also be grouped as T4ASS, while the I-type systems have been associated with T4BSS (Juhas et al., 2008).

1.4 The minimised T4SS

1.4.1 The VirB/D system of *Agrobacterium tumefaciens*

The *A. tumefaciens* VirB/D system acts as the paradigm for the minimised T4ASS and historically has been the most studied T4SS. *A. tumefaciens* is a plant pathogen that uses the T4SS to transfer oncogenic T-DNA into the plant cells causing tumorigenesis and formation of crown gall. The secretion system is encoded on the virulence *(vir)* region of the large 194-kb tumour-inducing (Ti-) plasmid. The translocation machinery is formed from a total of 12 proteins, with 11 VirB proteins encoded by the genes *virB1 – virB11* arranged in a single operon and the remaining protein, VirD4 being encoded on a separate *virD* operon. All of these proteins, except for VirB1, are required for infectivity in most plant cells, alongside nine other proteins that are expressed on four other operons on the Ti-plasmid, *virA, -C, -E* and *-G* (Berger and Christie, 1994, Zhu et al., 2000).

The T4SS is employed to infect the plant cells with a single stranded fragment of the Ti-plasmid, referred to as the T-strand, alongside several protein effectors that are needed for pathogenicity. Therefore,

despite its archetypal position, the VirB/D T4SS acts as both a conjugative and effector translocating apparatus. Once inside the plant cell cytoplasm, the T-DNA is targeted to the nucleus, guided by VirD2 and VirE2 proteins which contain a nucleus localisation signal (Howard et al., 1992, Citovsky et al., 1997). It is then integrated into the host's genome via nonhomologous recombination. The expression of the T-DNA results in the synthesis of opines and plant hormones involved in crown gall tumour formation. Opines are small compounds that act as a valuable nutrient source for the bacterium. They have also been shown to act as signalling molecules that lead to an increased conjugative transfer frequency of the Ti-plasmid (Genetello et al., 1977, White and Winans, 2007).

The expression of the *vir* genes is transcriptionally activated in response to plant damage. Upon plant cell wounding, phenolic compounds are released, which along with low pH levels (5.0 -5.5) and low phosphate levels, are detected by the membrane embedded sensor kinase, VirA (Zhu et al., 2000). VirA phosphorylates the cytoplasmic response regulator VirG, resulting in the activation of the genes in the *vir* operon by the phosphorylated VirG (Stachel and Zambryski, 1986). This initiates the T-DNA processing, the T4SS assembly and culminates in the transport of the substrates into the plant cell. The phosphorylated VirG also activates the *virA* and *virG* genes which establishes a positive feed-back loop.

The manner in which the T-DNA is transferred to the plant cell closely resembles interbacterial conjugation. The steps broadly involve DNA processing, transfer through the machinery culminating in the integration of the DNA into the host's chromosome. The T-DNA is processed by the DNA transfer and replication (Dtr) proteins VirD1 and VirD2. VirD2 is a relaxase and, in the presence of VirD1, nicks the T-DNA forming the T-strand (Pansegrau et al., 1993, Scheiffele et al., 1995). This reaction is reversible and the reverse reaction has been suggested to be important post-transfer for the chromosomal integration of the T-DNA (Tinland et al., 1995). The nucleoprotein complex is recruited to the translocation machinery by the coupling protein, VirD4, resulting in the transfer of the T-strand to the plant cell. This is supported by the attachment of the bacterium to the recipient plant cell which is mediated by the extracellular filament called T-pilus (Lai et al., 2000). Independently to the DNA substrate, the T4SS also delivers VirE2 into the plant cell and necessary to the export is the molecular chaperone VirE1 (Sundberg et al., 1996, Deng et al., 1999). Once inside the cell, VirE2 coats the T-strand and shields it from attack by the host's nucleases. Together with VirD2, VirE2 pilots the DNA substrate to the nucleus for chromosomal integration. A third protein that is transported by the VirB/D apparatus is VirF which mediates the proteolysis of the VirE2 coating (Magori and Citovsky, 2011, Gelvin, 2012).

Diversity also exists within the minimised T4ASS. In the case of *A. tumefaciens*, the core machinery proteins (VirB1 to VirB11 and VirD4) facilitate the transfer on oncogenic DNA to plant cells. However, the same canonical set of 12 structural components are utilised by *X. citri* in bacterial warfare.

1.4.2 The bacterial killing VirB/D system of *Xanthomonas citri*

The phytopathogen, *X. citri* is the causative agent of citrus canker, a disease that has a devastating impact on crops and therefore is of global economic importance. The *X. citri* genome was found to encode a T4ASS, on the chromosomal *vir* locus, which was demonstrated to be involved in the novel role of interbacterial antagonism (Alegria et al., 2005, Souza et al., 2015). This system has been termed Xanthomonadales-like T4SS (X-T4SS) and homologous loci have been found in several Xanthomonadales species (Sgro et al., 2019). A system homologous to X-T4SS was identified on the chromosome of the common opportunistic human pathogen, *Stenotrophomonas maltophilia* which enables killing of bacterial cells and appears to modulate apoptosis in human cells. (Nas et al., 2019, Bayer-Santos et al., 2019). It has been demonstrated, using time-lapse microscopy, that bacterial killing by *X. citri* and *S. maltophilia* occurs in T4SS and in a cell-cell contact dependent manner, and also that they can use their T4SS to duel one another (Souza et al., 2015, Bayer-Santos et al., 2019). Thus, highlighting the advantage that the T4SS confers to the bacteria for interbacterial competition.

The X-T4SS enables bacterial killing by secreting protein effectors called Xanthomonadales-like T4SS effectors (X-Tfes). It was the discovery of these substrates that provided the initial clues into the physiological function of the X-T4SS following a yeast two-hybrid assay using the coupling protein, VirD4 as bait (Alegria et al., 2005). The *X. citri 306* genome encodes for 13 potential X-Tfes. A common feature in all the X-Tfes is the presence of a conserved C-terminal region that is typically 120 residues called *Xanthomonas* VirD4 interacting protein conserved domain (XVIPCD), which is needed for interaction with VirD4 and is absolutely required for the secretion of the effector (Alegria et al., 2005, Souza et al., 2015). The XVIPCD is distinguishable by an N-terminal region which has conserved sequence motifs and a glutamine-rich C-terminus. It was recently shown that the N-terminus of the XVIPCD from a particular X-Tfe called XAC2609 adopts an ααβββ fold and this is sufficient for interaction with the VirD4 (Oka et al., 2022). Furthermore, the C-terminus was found to be important for efficient X-Tfe translocation to the target cell, however, the precise role of the glutamine residues in this region remains a mystery. The N-terminal region of X-Tfes appear to vary in length and architecture and have been predicted to contain catalytic domains that are responsible for toxicity of the effector. The domains that have so far been

predicted suggest that X-Tfes act as putative peptidoglycan (PG) binding proteins, PG glycohydrolases, lytic transglycosylases, PG peptidases or lipases (Souza et al., 2015). To protect from self-intoxication, the genes which encode for the X-Tfes are found with upstream genes that encode for the cognate immunity protein.

1.4.3 The P-type conjugative plasmids pKM101, RP4 and R388

A T4ASS has also been found to be encoded by conjugative plasmids R388, RP4 and pKM101 which are broad host range plasmids that confer antibiotic resistance. R388 is a 34 kb plasmid which harbours resistance to sulfonamide and trimethoprim antibiotics, while pKM101 is a 35 kb plasmid that encodes for penicillin resistance (Llosa et al., 1994). Both plasmids have a 15 kb transfer region which encodes for the T4SS with the genes being homologous to those from the VirB/D system from the Ti plasmids of *A. tumefaciens* and even maintaining the same genetic organisation. Therefore, while the genes from the pKM101 and R388 plasmids are often referred to with the prefix *tra* and *trw* respectively, for the rest of this Chapter, the VirB/D4 nomenclature will be used. Comparably, RP4 is a 60 kb plasmid which encodes resistance for several antibiotics including ampicillin, kanamycin and tetracycline (Pansegrau et al., 1987).

All conjugative plasmids contain the two regions which the conjugation genes are divided between, this consists of the Mating pair formation proteins (Mpf) and DNA transfer and replication proteins (Dtr). The Mpf consists of the proteins that form the T4SS with the operon encoding for *virB1 – virB11*. The neighbouring operon to this is the Dtr. The Dtr proteins play an important role in the initial steps of conjugation by assembling at the *oriT* to form the relaxosome. In its most simple form, the Dtr encodes for three proteins as represented by the R388 plasmid: a VirD4 like protein, an accessory protein and a relaxase/helicase (Christie, 2016). The Ti plasmids, encode for an additional Dtr accessory protein, VirC1 which has been demonstrated to promote the nicking activity of the relaxase, VirD2, and also is thought to be involved in the recruitment of the relaxosome to the cell pole to coordinate the docking of the substrate to the VirB/D4 machinery (Atmakuri et al., 2007, Christie et al., 2014).

The plasmids also encode specific functional genes. The bacteria conjugative plasmids encode for a protein that is involved in entry exclusion (Eex). In the case of pKM101 and R388 this is encoded on the *eex* genes, whereas in RP4 this function is carried out by protein encoded by *TrbK*. These proteins serve to prevent redundant DNA transfer between two donor cells. The non-conjugative Ti plasmid does not encode for an Eex protein as such a function is not required.

1.4.4 The general architecture of T4ASS

There has been exciting progress in deciphering the architecture and translocation pathway of T4ASS. Central to this progress has been the structural studies which provided an understanding of the architecture of the apparatus (Figure 1.3).

The first insight into the general structure came from a negative stain electron microscopy (NS-EM) reconstruction of the R388 T4SS which showed that the machinery within the bacterial cell is formed from four subcomplexes: the outer membrane core complex (OMCC) , inner membrane complex (IMC), a stalk that bridges the two membrane complexes, and lastly, arches that connect the stalk and the IMC (Low et al., 2014). Approximately 8 years later, a high-resolution structure of the R388 apparatus was solved using single particle cryo-EM which provided details on the location of the individual VirB proteins (Mace et al., 2022). While it confirmed the overall structure that was observed with NS-EM, the cryo-EM reconstruction revealed a different IMC architecture which will be discussed further on in this Chapter. This reconstruction also furthered the findings of previously solved structures of various subunits of the machinery which includes studies using X-ray crystallography and electron microscopy.

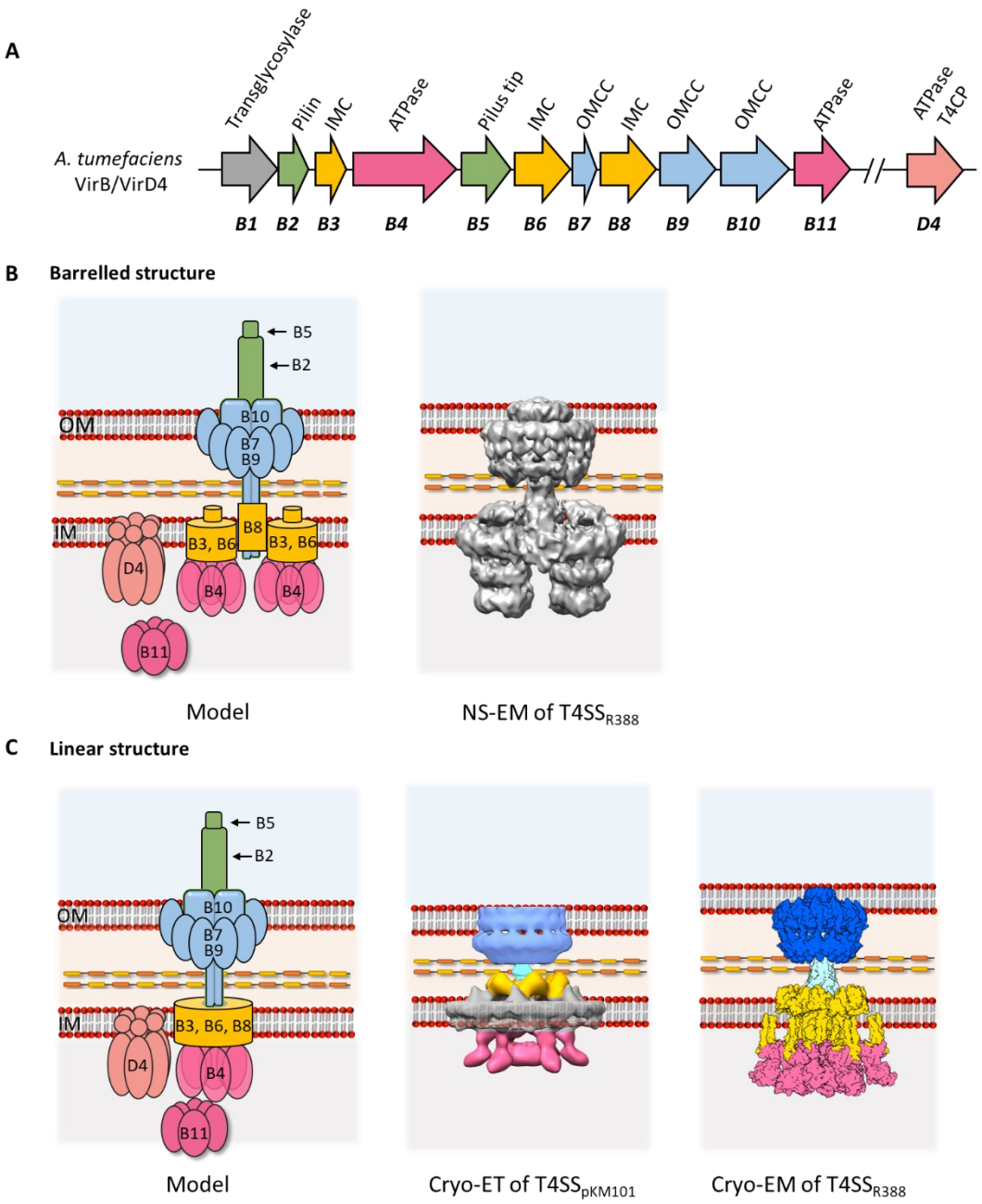

Figure 1.3: The minimal T4SS. (A) The organisation of the genes required for the formation of the VirB/D4 apparatus encoded by *A. tumefaciens*. The 12 VirB/D4 genes required are shown with the genes encoded labelled below the arrows and the enzymatic functions listed above. The genes are coloured depending on the complexes the encoded protein forms. **(B)** The barrelled organisation of the apparatus shown by schematic representation and negative stain electron microscopy (NS-EM) reconstruction of the R388 encoded machinery (EMD-2567). **(C)** The linear structure of the minimal T4SS shown by cartoon representation, and the cryo-electron tomography (cryo-ET) and the cryo-electron microscopy (cryo-EM) of the VirB/D4 apparatus encoded by the plasmid pKM101 (EMD-24100 and EMD-24098) and R388 (EMD-12707, EMD-12708, EMD-12715) respectively.

1.4.4.1 The T-pilus

Found at the cell surface, conjugative pili are helical, extracellular filaments elaborated by T4SS and are composed of repeating pilin subunits. There are two prevailing morphological classes of pili (Bradley, 1980a). The first is characterized by long and flexible pili which is exemplified by the F-pilus and will be discussed further on in this Chapter. The second group consists of pili that are short and rigid, such as the T-pilus belonging to the VirB/D4 T4SS. The T-pilus is formed from the major pilin subunit VirB2 and the tip protein VirB5 (Lai and Kado, 1998, Aly and Baron, 2007). The VirB2 is a homologue of the major structural pilin subunit TraA of the F-pilus. Similar to TraA, it is processed to form the mature pilin (Shirasu and Kado, 1993). VirB2 is synthesised as a 12.3 kDa pro-pilin and, following cleavage of the signal sequence, forms a 7.2 kDa mature product. Through two hydrophobic domains, the mature VirB2 is inserted into the inner membrane where they form a pool of monomers before being recruited by the T4SS machinery to build the pilus. VirB5 has been postulated to be important for T-pilus nucleation and may also facilitate attachment to the plant cell receptors (Yuan et al., 2005). A similar role in receptor binding has been demonstrated for CagL, a VirB5 homologue in the *H. pylori* Cag T4SS, where it enables infection of mammalian cells by binding to β-integrin on the host cell surface (Backert et al., 2008). The assembly of the T-pilus requires most of the same proteins that are needed to form the translocation machinery with the exception of the transglycosylase VirB1 which is needed for pilus assembly but not substrate translocation, and VirD4 ATPase which is important for substrate transfer but not pilus formation (Berger and Christie, 1994, Zupan et al., 2007).

Two independent studies have reported the structure of the T-pilus (Amro et al., 2023, Kreida et al., 2023). The T-pilus is a five start helical filament with an external diameter of approximately 76 Å and a luminal diameter of 22 Å. Interestingly, both structures did not provide evidence of cyclic VirB2. This contradicts early work using mass spectrometry which proposed that VirB2 undergoes cyclisation before recruitment to build the pilus (Eisenbrandt et al., 1999). Furthermore, while the two studies yielded similar structures, discrepancies exist in the lipid detected. Density corresponding to a phospholipid moiety was detected near the VirB2 subunit. This was identified as phosphatidylglycerol in the structure by Kreida and colleagues, while Amro and colleagues found phosphatidylcholine in their structure (Amro et al., 2023, Kreida et al., 2023). Interestingly, both studies found positively charged residues extending towards the centre of the T-pilus lumen. This contributes to the T-pilus lumen having a positive charge, which raises the question of whether the pilus is a conduit for ssDNA transport. Nevertheless, both studies concluded that VirB2 and the phospholipid form the T-pilus with stoichiometric ratio of 1:1.

1.4.4.2 The outer membrane core complex from T4ASS

Since it self-assembles into a stable complex, the OMCC from T4SS has been at the forefront of structural research. For the T4ASS, the crystal structure of a section of the *E. coli* pKM101 core complex along with the cryo-electron microscope structures of the OMCC belonging to pKM101, *X. citri* and very recently the high-resolution structure of the R388 machinery have provided details of the architecture of the core complex (Fronzes et al., 2009, Chandran et al., 2009, Rivera-Calzada et al., 2013, Low et al., 2014, Sgro et al., 2018, Mace et al., 2022). The core complex from T4ASS is formed from three proteins: the lipoprotein VirB7, VirB9 and VirB10. Initial structural work showed the pKM101 core complex having a barrel-like shape, approximately 185 Å in diameter and height and, formed from 14 copies of the three proteins (Fronzes et al., 2009). This 14-fold symmetric arrangement was confirmed by subsequently solved core complex structures, and the presence of a single symmetry became a characteristic of the T4ASS OMCCs. However, recently a high-resolution structure of the R388 T4SS was solved which showed the complex containing two different symmetries.

While VirB7 and VirB9 are contained within the outer membrane, VirB10 is a remarkable protein in that it spans both the inner and outer membrane. The importance of the OMCC is highlighted by the observation that mutations that affect VirB10 from associating with the VirB9 and VirB7 proteins to form the complex abrogates pilus formation and transfer ability (Jakubowski et al., 2009). This makes the core complex indispensable for the functioning of the T4SS. Early insight into the specific location of the OMCC proteins came from a limited proteolysis study coupled with imaging of the resulting complex using electron microscopy (Fronzes et al., 2009). The structure of the core complex can be divided into two layers; the outer layer (O-layer) consists of VirB7 and the C-terminal domains (CTD) of VirB9 and VirB10, while the inner layer (I-layer) comprises of the N-terminal domains (NTD) of VirB9 and VirB10. The crystal structure of the isolated O-layer revealed that VirB10 forms the interior wall of the complex surrounded by VirB7 and VirB9 (Chandran et al., 2009). Within the tetradecamer, a central ring is formed by the VirB10 CTD adopting a two helix bundle, which has been referred to as the antenna region, and is thought to be embedded within the outer membrane, and thereby provides a channel across it. This hypothesis is supported by the observation that when a FLAG epitope is inserted between the antenna α-helices, the tag is exposed extracellularly (Chandran et al., 2009). This chambered architecture of the core complex appears to be a common feature along with the central localisation of VirB10 amongst the T4SS OMCCs.

A similar protein arrangement was observed for the *X. citri* OMCC with the O-layer being reminiscent of the equivalent in the pKM101 T4SS (Figure 1.4) (Fronzes et al., 2009, Sgro et al., 2018). A striking difference

in the *X. citri* OMCC is that the VirB7 lipoprotein is considerably larger compared to its counterpart in the conjugative VirB/D4 machinery due to the presence of a C-terminal N0 domain (Souza et al., 2011). N0 domains are also found on the N-termini of secretins from T2SS and T3SS, as well as being present on the lipoprotein DotD from the *L. pneumophila* Dot/Icm T4SS. Thus, highlighting remarkable structural similarities among proteins involved in transport of substrates across the bacterial outer membrane. The additional domain adopts a globular fold which extends outward from the central structure. This contributes to the *X. citri* OMCC having a "flying-saucer" shape as opposed to the barrel-like form of the pKM101 and R388 OMCC. The resolution of the *X. citri* complex also meant that features of the I-layer became visible which was not resolved in previous structures. Linkers were observed between the VirB9 and VirB10 allowing for considerable flexibility among the two layers.

Recently, a high resolution cryo-EM structure of the R388 T4SS showed that the OMCC consists of two symmetries; the O-layer has 14-fold symmetry while the I-layer adopts 16-fold symmetry (Figure 1.4 and Figure 1.5) (Mace et al., 2022). This means that there are 16 VirB10 and VirB9 at the I-layer of which 14 extend to form the O-layer. This leaves two heterodimeric VirB10-VirB9 complexes at the I-layer without an obvious connection to the O-layer. While this symmetry mismatch was unexpected since it has not been seen in minimised T4ASS OMCCs yet, it is a common characteristic of the core complexes from expanded T4SS (Chung et al., 2019, Durie et al., 2020).

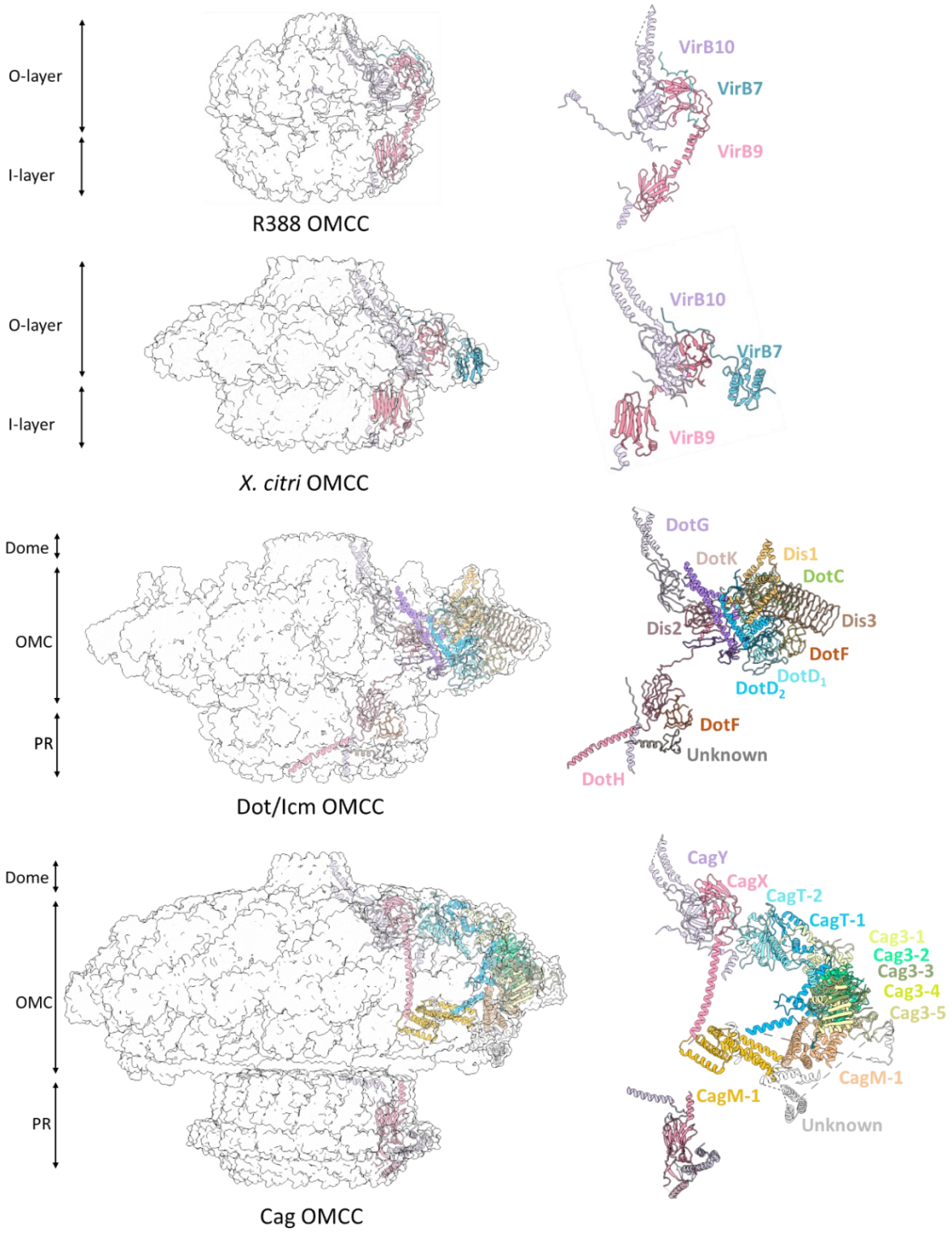

Figure 1.4: The high-resolution cryo-EM structures of different T4SS outer membrane core complexes (OMCC). The cryo-EM map and the asymmetric units are shown for the OMCCs of: R388 (EMD-12707, EMD-12708, PDB-7O3J, PDB-7O3T), *X. citri* (EMD-0089, PDB-6gyb), Dot/Icm (EMD-22068, EMD-22069, PDB-7MUD, PDB-7MUC), and Cag (EMD-22081, EMD- 20021, PDB-6X6J, PDB-6X6S). O-layer (Outer layer), I-layer (Inner layer), OMC (Outer membrane complex), PR (Periplasmic ring).

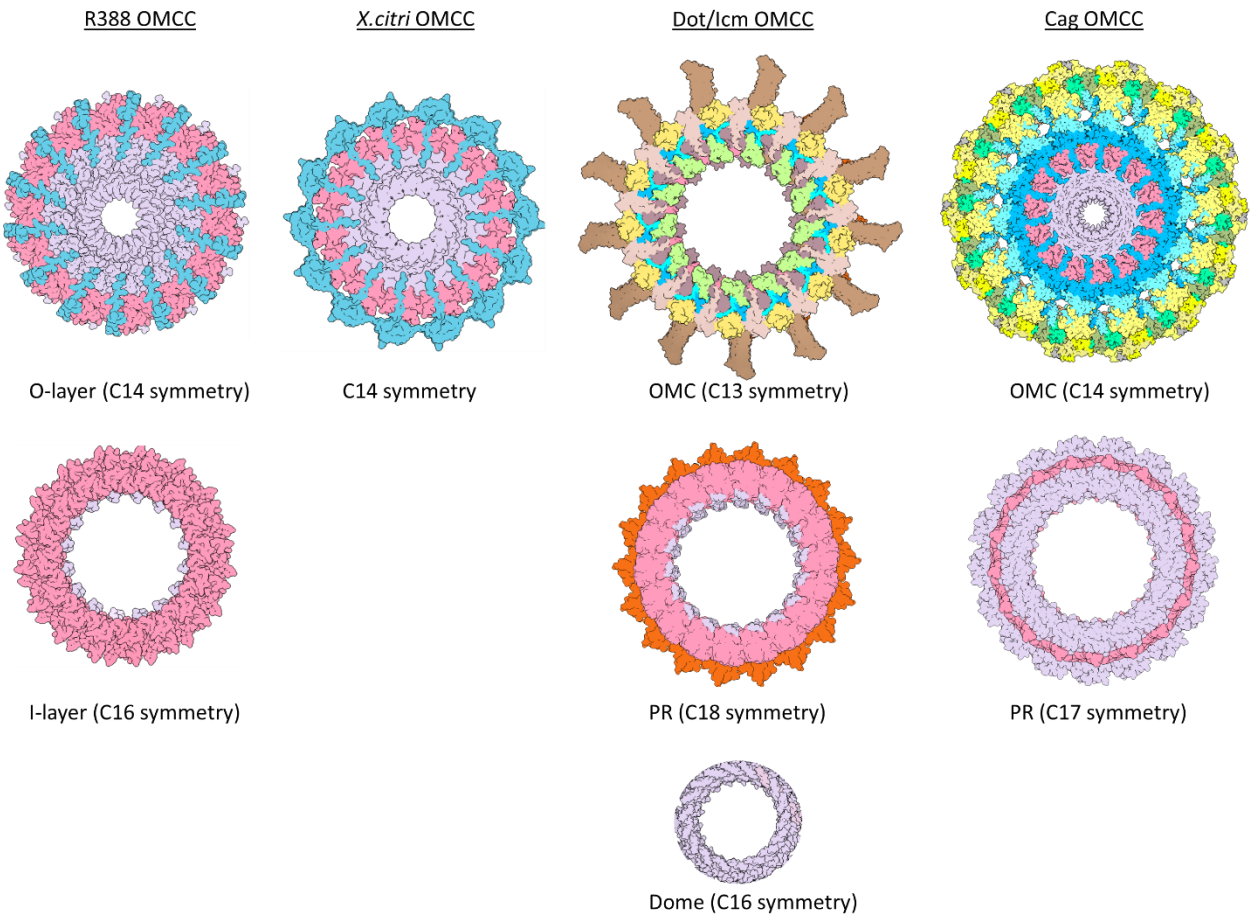

Figure 1.5: The symmetry of different T4SS outer membrane core complexes (OMCC). The cryo-EM map and the symmetry of various components are shown for the OMCCs of: R388 (EMD-12707, EMD-12708, PDB-7O3J, PDB-7O3T), *X. citri* (EMD-0089, PDB-6gyb), Dot/Icm (PDB-7MUC, PDB-7MUD, PDB-7MUE), and Cag (PDB-6X6J, PDB-6X6S). O-layer (Outer layer), I-layer (Inner layer), OMC (Outer membrane complex), PR (Periplasmic ring). The colour key of the proteins is the same as Figure 1.4.

While the OMCC is often considered to be the scaffold for the channel through which substrates through, there is increasing evidence which alludes to the OMCC playing a more active role in conjugation. VirB10 has been reported to be involved in energy sensing and transducing with the protein undergoing a structural transition following substrate binding and ATP hydrolysis by the ATPases VirD4 and VirB11 (Cascales and Christie, 2004a, Cascales et al., 2013). While direct contact of DNA with VirB10 was not observed in a formaldehyde (FA) cross-linking assay, VirB10 was required for DNA to establish FA-cross-linkable contacts with VirB2 and VirB9 (Cascales and Christie, 2004b). This suggests that VirB10 may dynamically be involved in coupling the cytoplasmic/inner membrane ATP activity to the gating of the transfer machinery at and or proximal to the outer membrane. In line with this mode of action, a G272R

mutation in VirB10 was identified to be a gating defect mutation, with VirB10 being locked in an active conformation even in an energy-depleted state which conferred leakiness to the channel, with the unregulated secretion of VirE2 to the cell surface independently of target cell contact (Banta et al., 2011). Using molecular simulations, the conformation of VirB10 caused by the G272R mutation has been shown to result in a more rigid complex compared to the wildtype, leading to reduced conformational flexibility (Darbari et al., 2020). This may cause the complex to become unresponsive to intracellular signals thereby impacting the gating of the channel by VirB10. Together with the inherent flexibility observed within the OMCC structure, it suggests that the core complex not only acts as a structural scaffold, but through the activity of VirB10, it can undergo structural rearrangement to allow passage of molecules in response to substrate engagement and ATP hydrolysis by VirD4 and VirB11.

1.4.4.3 The inner membrane complex (IMC), arch and stalk

Following on from the OMCC, the second membrane complex which structures the T4SS is the IMC. The VirB10 N-terminus extends down from the OMCC and forms part of the IMC along with multiple other proteins. This includes VirB3, VirB4 and VirB8. While additional proteins have also been associated with the IMC, namely VirB5, VirB6 and the ATPases VirB11 and VirD4, the recent structure of the almost entire R388 T4SS shows that the four proteins (VirB5, VirB6, VirB11 and VirD4) form components which are distinct from the IMC (Figure 1.6) (Mace et al., 2022). Therefore, in this Chapter, those proteins will be discussed separately to the IMC.

The first structural insight into the complex spanning the inner membrane came from a negative stain reconstruction of the R388 T4SS (Low et al., 2014). The IMC was found to be asymmetric with a dimension of 255 Å in width and 134 Å in height. A striking feature of this complex was that the AAA^+ VirB4 ATPase formed two side by side hexameric barrels that extended down into the cytoplasm (Figure 1.3). Contradictory to this, the *in situ* cryo-electron tomography (cryo-ET) structure of the pKM101 T4SS showed VirB4 adopting a central hexamer of dimers (Figure 1.3) (Khara et al., 2021). This arrangement has been further supported by the high resolution cryo-EM structure of the R388 T4SS which displayed six dimers of VirB4 assembling to form a hexamer of dimers (Figure 1.6) (Mace et al., 2022). This oligomeric conformation has also been visualised in the expanded T4SSs (Chetrit et al., 2018, Hu et al., 2019a, Hu et al., 2019b). However, within the cryo-EM dataset, a few of the 2D classes showed side-views that were reminiscent of the double-barrelled architecture observed in the negative stain reconstruction. This suggests that a small number of particles existed that had the double-barrelled shape. Therefore, it

may be plausible that the two different IMC architectures represent different stages in the assembly of the T4SS.

The cryo-EM R388 T4SS structure also provided the first detailed insight into the protein organisation that builds the IMC which is 1.32 MDa in size (Figure 1.6). Overall, the IMC adopts 6-fold symmetry with it being formed from 6 protomers. Each protomer consists of one VirB3, two VirB4 and three N-terminal portions of VirB8. The major protein component of the complex is VirB4 and is organised as such that one subunit in each of the 6 dimers (referred to as $VirB4_{central}$), forms a central hexamer while the other subunit of the dimer protrudes away from the centre (termed $VirB4_{outside}$) (Figure 1.6). The other two protein constituents of the IMC form part of the complex by interacting with one of the VirB4 dimer subunits. VirB3, which is a small, two-pass membrane protein, interacts solely with the $VirB4_{central}$ subunit of the dimer. Therefore, there are a total of 6 VirB3 proteins in the complete T4SS. On the other hand, VirB8 associates with the $VirB4_{outside}$ subunit of the dimer via an N-terminal portion which has been called $VirB8_{tails}$. There are 3 copies of the $VirB8_{tails}$ per $VirB4_{outside}$ subunit thus, making a total of 18 VirB8 proteins in the entire T4SS.

While the N-terminus of VirB8 forms part of the IMC, the rest of the protein form the arches at the periplasm (Figure 1.6) (Mace et al., 2022). Prior to this structure, the composition and architecture of the arches was not known. It is built from hexamers of homotrimeric units of VirB8 that arrange as a ring around the stalk with a diameter of 177 Å. The stalk is the central, coned shape structure that connects the two membrane spanning complexes of the T4SS. Rising from the inner membrane, the stalk is formed from a pentamer of VirB6 units. Mounted on top of the VirB6 base, is a pentamer of VirB5, such that the one VirB5 binds between two VirB6 proteins. Based on the structural findings, it has been proposed that during pilus biogenesis, 5 VirB2 subunits bind to the 5 VirB6 from the inner membrane. The pilin subunits are then levered up to the site of assembly in order to form the pilus, which in turns frees up the binding sites on VirB6 for a second cycle of VirB2 binding (Mace et al., 2022). Both the VirB proteins consist predominantly of long α-helices which enables the formation of a lengthy stalk, spanning a height of 216 Å, that bridges the IMC and OMCC.

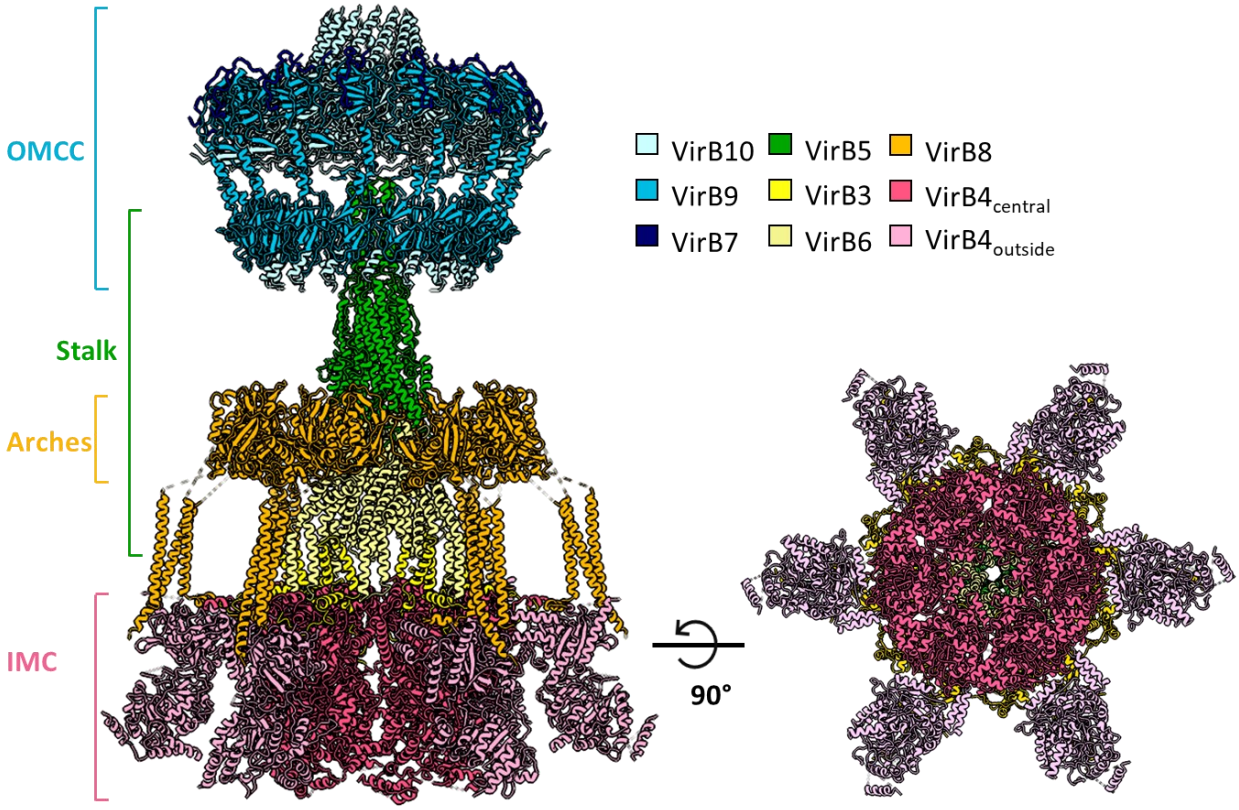

Figure 1.6: The high resolution cryo-EM structure of the R388 T4SS. The structure is formed from four general sections: OMCC (Outer membrane core complex), stalk, arches, and IMC (Inner membrane complex).

1.4.4.4 The ATPases VirD4 and VirB11

Alongside VirB4, the ATPases VirD4 and VirB11 form the energy centre which is situated at the base of the translocation machinery. However contrary to VirB4 which is stably attached to the T4SS, VirD4 and VirB11 are more loosely associated. VirD4 is also known as the type IV coupling protein (T4CP) and functions as a receptor for DNA and protein substrate binding and, is involved in shuttling them to the translocation apparatus. Therefore, VirD4 acts as the connecting protein between the translocating substrates and the T4SS. This has been proposed to occur via participation from VirB11 as VirB11 has been observed to interact with both VirD4 and VirB4 (Ripoll-Rozada et al., 2013). VirB11, which is often referred to as the traffic ATPase, belongs to the AAA^{+} ATPase family. Unlike VirB4 which is a ubiquitous component of T4SSs, VirB11 is not present in all machineries such as the T4SS encoded by the F-plasmid. The VirB11 localises to the cytoplasmic face of the inner membrane and has been shown to co-fractionate with both the cytosol and inner membrane fraction (Rashkova et al., 1997). This is indicative of VirB11 interacting dynamically with the IMC.

A prevailing controversy in the field surrounding the ATPases is how VirD4 mediates transfer of the bound nucleoprotein complex (ssDNA-relaxase) to the membrane embedded machinery. There are currently two main models which have been proposed to address this. According to one model, VirD4 engages with the nucleoprotein complex and then coordinates with the ATPases VirB11 and VirB4, to process and transport the molecule through the T4SS (Figure 1.7). This is supported by the observation that VirB11 has been shown to interact with both ATPases and further by the findings of immunoprecipitation assays defining the T-DNA translocation pathway (Cascales and Christie, 2004b, Atmakuri et al., 2007, Ripoll-Rozada et al., 2013). The pathway through the T4SS was proposed based on the ability of various proteins to form formaldehyde cross-links with the T-DNA, and contacts were observed with VirD4 T4CP, the ATPase VirB11, IMC proteins VirB6 and VirB8, the OMC protein VirB9, and lastly, the pilin VirB2. While VirB4 was not found to cross-link with the T-DNA, deletion of the ATPase resulted in arrested transfer between VirB11 to VirB6/VirB8 (Cascales and Christie, 2004b). Thus, suggesting that VirB11 and VirB4 are actively involved in the T-DNA transport. The alternative model suggests that VirD4 recruits the nucleoprotein complex and directly guides it to the periplasm, where it enters the T4SS through contact with IMC proteins VirB6 and VirB8, and is subsequently transported through the machinery (Larrea et al., 2017). This has also been suggested for the expanded Dot/Icm T4SS from *L. pneumophila* (Meir et al., 2020). Given that there is compelling evidence in support of both models, it makes it difficult to distinguish between the two. This is further exacerbated by the lack of a structural studies showing an intact T4SS structure with either VirD4 or VirB11 bound. Therefore, this remains as one of the most poorly understood aspects of the T4SS field.

While the high resolution structure of the T4SS with the VirD4 ATPase bound remains to be solved, structures of VirD4 alone and also components of it associated with the cognate T4SS is available. As a T4CP, the VirD4 consists of a N-terminal transmembrane domain which has been implicated in establishing contact with the nucleoprotein complex, and a large, cytoplasmic C-terminus (Llosa and Alkorta, 2017). The C-terminal region is formed from two distinct domains: a nucleotide binding domain (NBD) which comprises of the canonical Walker A and Walker B nucleotide binding motifs involved in ATP hydrolysis, and an all-α-domain (AAD). The crystal structure of the functional soluble domain of the VirD4 from the R388 plasmid showed a globular hexameric assembly (Gomis-Ruth et al., 2001). Each monomer consists of two major structural domains which includes a NTD and an AAD. The six protomers come together to form a ring which has a central channel with a diameter of approximately 20 Å that narrows to 8 Å at the channel entrance. The overall shape resembles that of the F_1-ATPase. A similar hexameric arrangement has also been observed in the Dot/Icm T4CP with DotL connected to a larger bell-shaped

complex (Kwak et al., 2017). A structure of VirD4 cross-linked with the T4SS was probed using negative stain electron microscopy. This revealed two dimers of VirD4 integrated between the VirB4 hexamers, however, this has been speculated to represent an incompletely assembled apparatus (Redzej et al., 2017). Unfortunately, VirD4 was absent in the recently solved R388 T4SS structure which hinders understanding of how the ATPase interacts with the machinery and suggests that perhaps due to the dynamic nature of the protein, cross-linking is required to capture it (Mace et al., 2022).

Insight into the structure of VirB11 comes from the homolog in *B. suis* (Hare et al., 2006). The protein was found to form a hexameric double stacked ring, and a similar arrangement has also been reported for the equivalent protein in the Cag T4SS (Hu et al., 2019b).

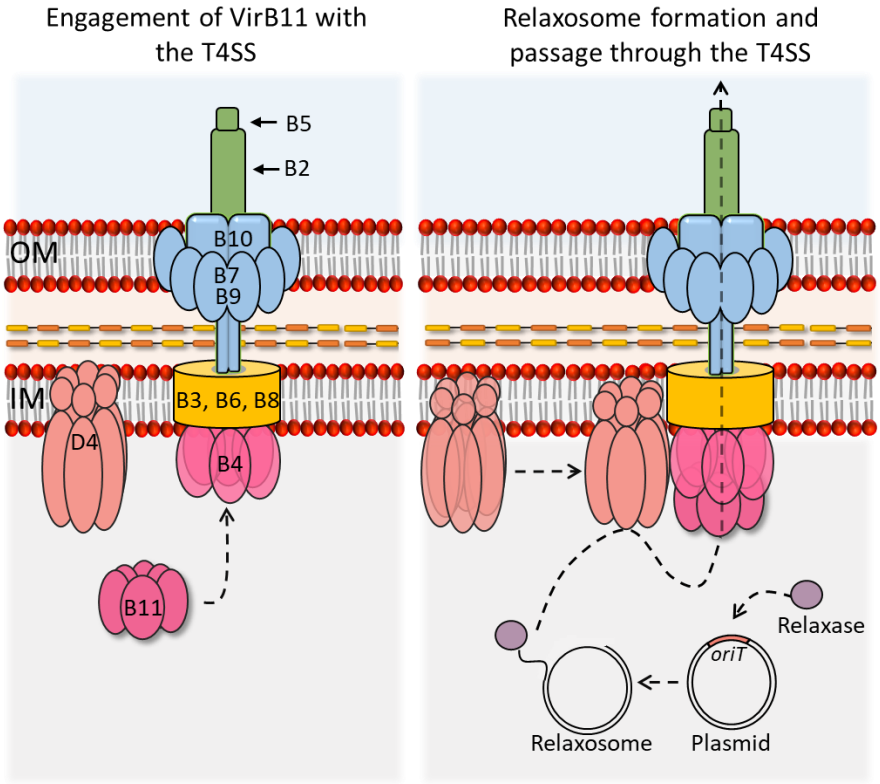

Figure 1.7: A working model for substrate translocation through the VirB/D4 T4SS. This figure illustrates one of the prevailing models for substrate transfer through the machinery. This involves the docking of the VirB11 hexamer with the VirB4 hexamer. Then, through the involvement of VirD4., the nucleoprotein is shuttled through the translocation apparatus. *oriT* (Origin of tranfer)

1.4.4.5 The lytic transglycosylase VirB1

VirB1 is not a structural component of the T4SS machinery but is involved in lysis of the peptidoglycan layer. It is a member of lytic transglycosylases family of proteins that are lysozyme-like virulence factors (Chandran Darbari and Waksman, 2015). VirB1 aids the activity of the T4SS by lysing the peptidoglycan which alleviates the structural restraints that are imposed by the polysaccharide layer and thereby, enabling the construction of the machinery across the cellular envelope. Prior to its lytic activity, in the periplasm, VirB1 is cleaved into two components in *A. tumefaciens*. This includes the NTD which contains the lytic activity and a C-terminal region, VirB1*, that is secreted to the extracellular milieu following cleavage, and is proposed to be involved in pilus formation (Baron et al., 1997, Llosa et al., 2000, Zupan et al., 2007). Interestingly, in *A. tumefaciens*, it has been shown that VirB1 is not required for assembly of the T4SS but is needed for pilus biogenesis (Berger and Christie, 1994, Zupan et al., 2007). This dispensability of VirB1 for channel formation is surprising given that the lysis of the peptidoglycan layer should supposedly be necessary for the extension of the T4SS across the two membranes.

1.5 The expanded T4SS

Compared to the minimised T4SS, the expanded T4SSs are more complex, both in the number of proteins involved in structuring the apparatus, and also in terms of the overall architecture. Also known as T4BSS, the expanded machineries are composed of the core proteins that are found in the minimised systems, but also contain numerous others which are often system specific, which presumably extends functional capacity. Key examples of the expanded T4SS includes the *L. pneumophila* Dot/Icm, *H. pylori* Cag and the T4SS encoded by the F-plasmid. Recently, a strain of *P. putida* was found to encode for a T4BSS that is involved in bacterial killing by translocating toxin effectors into the target cell (Purtschert-Montenegro et al., 2022).

1.5.1 The Dot/Icm system of *Legionella pneumophila*

Legionella pneumophila is the causative agent of Legionnaires' disease, a severe form of pneumonia (Fraser et al., 1977, McDade et al., 1977). It is a human respiratory pathogen and infects host macrophages. A factor that is central to *L. pnuemophila*'s virulence is the Defective in organelle trafficking/Intracellular multiplication (Dot/Icm) T4SS (Figure 1.2) (Marra et al., 1992, Berger and Isberg, 1993). The Dot/Icm T4SS is built at the cell poles and enables secretion of effectors into the host cell

(Jeong et al., 2017). Over 300 effectors are transported which allows the bacteria to manipulate the host cellular processes (Isaac and Isberg, 2014). Firstly, the T4SS and its effectors allows evasion of the host immune system and facilitates assembly of a protective shelter called Legionella containing vacuole (LCV) (Figure 1.2) (Lockwood et al., 2022). The LCV avoids endocytic maturation processes such as fusion with lysosyme and resists acidification (Horwitz, 1983, Horwitz and Maxfield, 1984). The LCV adopts endoplasmic reticulum (ER)-like properties and serves as a replicative niche. The importance of the machinery is highlighted by the observation that mutants deficient in a functional Dot/Icm T4SS are avirulent (Isaac and Isberg, 2014). The Dot/Icm translocated effectors control various stages of infection by regulating signal transduction, membrane trafficking and cytoskeleton dynamics.

Early work on effectors suggested the presence of a translocation signal at the C-terminus (Nagai et al., 2005). The sequence consists of 20 – 30 amino acid residues that have specific biochemical properties, for example small charged residues (Burstein et al., 2009). Alternatively, as observed in a large number of effectors, the translocation sequence can also be a glutamate-rich E-block motif (Huang et al., 2011). Some substrates have even been identified to contain an internal translocation signal that has been reported to contact the adaptor proteins IcmS and IcmW (Cambronne and Roy, 2007, Sutherland et al., 2012). This highlights the variations that exist amongst the large repertoire of Dot/Icm effectors. Despite, the increasing number of substrates being discovered, a universal consensus sequence has still not been determined (Lockwood et al., 2022). The absence of such a definitive sequence suggests that perhaps for different effectors there may be different modes of recruitment and transport through the Dot/Icm apparatus.

The Dot/Icm machinery acts as a paradigm for the T4BSS. The machinery consists of approximately 27 proteins whose genes are organised mostly on 2 large gene clusters (Figure 1.8) (Voth et al., 2012). However, 3 additional proteins affiliated with the Dot/Icm T4SS has been recently identified (Sheedlo et al., 2021). A T4BSS similar to the Dot/Icm machinery is also utilised by the zoonotic pathogen *Coxiella burnetii*, thus highlighting the significance of this apparatus in the virulence of intracellular pathogens (Segal and Shuman, 1999, Sexton and Vogel, 2002).

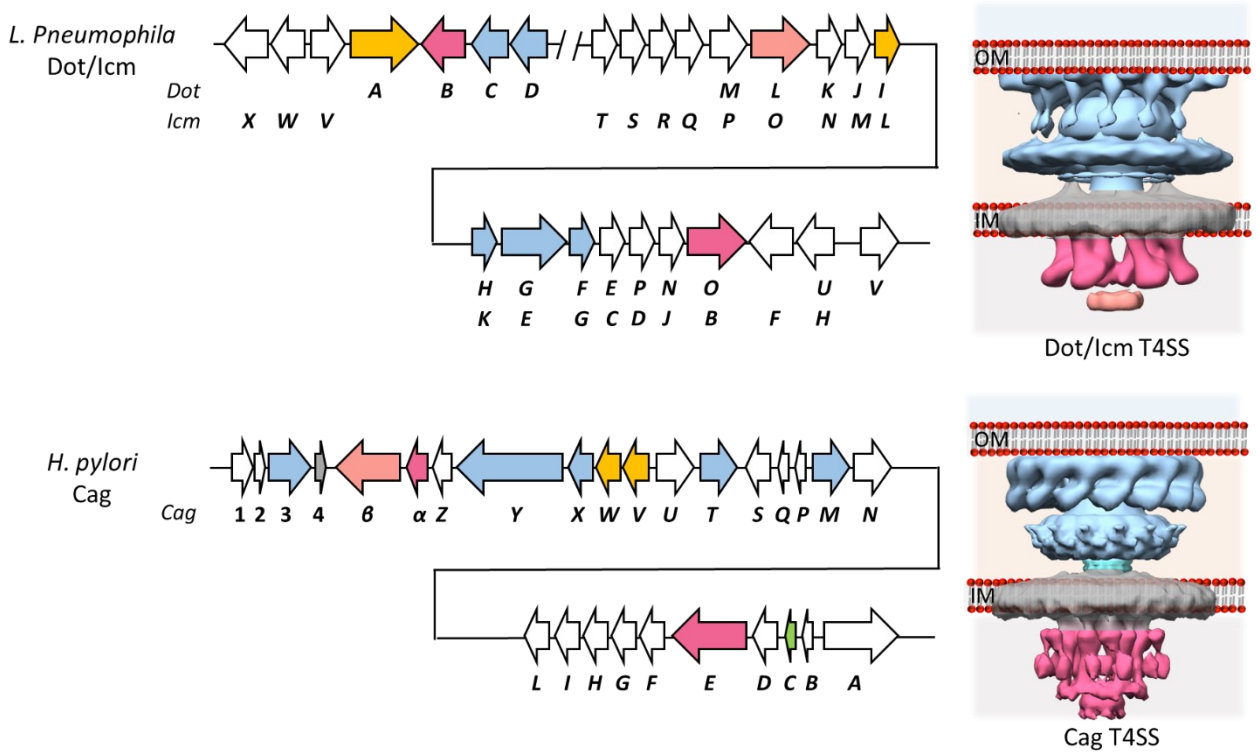

Figure 1.8: The expanded T4SS. The organisation of the genes involved in the formation of the Dot/Icm and the Cag T4SSs. The genes are coloured depending on the complexes the encoded protein forms and follows the same scheme as Figure 1.3: yellow (IMC), pink (ATPase), salmon (T4CP ATPase), blue (OMCC), green (pilus related). Homologous genes are shown by same colour coding. The cryo-ET structures of the *L. pneumophila* Dot/Icm T4SS (EMD-7611 and EMD-7612) and *H. pylori* Cag T4SS (EMD-0634 and EMD-0635).

1.5.2 The Cag T4SS of *Helicobacter pylori*

Helicobacter pylori is a Gram-negative bacterium that colonises the stomach in approximately half of the human population. The presence of the bacteria in the stomach is the greatest known risk factor of developing gastric cancer. Present in the highly virulent *H. pylori* strains is the *cag* pathogenicity island (*cag*PAI) which encodes for the Cag T4SS and the effector protein, CagA (Blaser et al., 1995, Cover, 2016). Individuals harbouring *H.pylori* strains that have the *cag*PAI have a greater risk of developing cancer compared to those who have *H.pylori* strains lacking this genetic island (Cover, 2016). It should be noted that alongside the Cag T4SS, three additional T4SSs have been described that are assembled by *H. pylori* (Fischer et al., 2020). One such T4SS is the ComB system which facilitates natural transformation and was discussed in section 1.2.4. The remaining two secretion machines are Tfs3 and Tfs4, the function of which are presently not well understood (Cover et al., 2020). Nevertheless, the focus of this section will be the

Cag T4SS which is the first characterised T4SS in *H. pylori* and due to its complexity is categorised as a T4BSS expanded apparatus.

The Cag T4SS is chromosomally encoded on the 40 kb *cag*PAI region. The *cag*PAI region encompasses approximately 27 genes, of which 17 are needed for substrate transfer, one encodes for the oncoprotein CagA, and several have unknown function (Figure 1.8) (Cover et al., 2020). Currently, CagA is the only known protein to be secreted by the Cag T4SS and is injected into gastric epithelial cells where it elicits cancer causing activities. The T4SS causes several cellular changes that are attributed to the activity of CagA. Examples include modified cell polarity and cell adhesion, enhanced cell proliferation, increased motility of cells and altered cellular shape (Hatakeyama, 2004). The Cag T4SS is also involved in causing cellular alterations, which occurs independently of CagA. One such action is the production of a strong pro-inflammatory response in gastric cells by encouraging interleukin-8 production through a mechanism involving NF-kB and transfer of lipopolysaccharide metabolites such as ADP-glycero-β-D-manno-heptose into the target cell (Backert and Naumann, 2010, Cover et al., 2020). Furthermore, it has been demonstrated that Cag T4SS is directly involved in the contact-dependent activation of Toll-Like receptor 9 (TLR-9) which is driven by the transfer of *H. pylori* DNA which actively engages with the host receptor (Varga et al., 2016). Therefore, the Cag T4SS plays a central role in promoting development of gastric cancer.

1.5.3 The general architecture of the expanded T4SS

Despite the complexity and vast number of proteins involved in both the Dot/Icm and Cag T4SS, the overall architecture is reminiscent of the minimised VirB/D4 machineries (Figure 1.8). Insight into the structural organisation of both these expanded machineries in intact bacterial cells has been revealed using cryo-ET (Figure 1.8) (Chang et al., 2018, Chetrit et al., 2018, Ghosal et al., 2019, Hu et al., 2019b). When visualised using this approach, the Dot/Icm and Cag T4SSs consists of two membrane spanning complexes (IMC and OMCC) which are bridged by a central stalk. Nevertheless, clear differences exist between the two expanded T4SS and that of the VirB/D4 T4SS. One specific difference is the absence of a definitive pilus associated with the translocation apparatus. Presently, a pilus connected to the Dot/Icm apparatus has not been observed which is further supported by bioinformatic analyses of the T4SS failing to identify an appropriate pilus gene cluster together with the absence of a pilus in the cryo-ET structures (Bock et al., 2021). A pilus-like component has been observed on *H. pylori* cells which are thought to be produced in response to contact with the host cell. However, whether this filamentous structure is indeed a pilus

elaborated by the Cag T4SS has sparked controversy. A pilin subunit has not been conclusively identified. On the other hand, *in vivo* structural analysis of *H. pylori* Cag T4SS showed rod-like structures which may represent the pilus. However, this was not detected in direct connection with the OMCC (Chang et al., 2018). Therefore, it is still uncertain whether a pilus is elaborated by the Cag T4SS.

1.5.3.1 The outer membrane core complex

Initial work surrounding the Dot/Icm and Cag T4SS suggested that the core complexes are made from five proteins. The ability to successfully isolate the complexes has allowed analysis of its structure and protein composition. This has indicated the intricacy that exists within the core complex architecture. The Cag T4SS OMCC is composed of five components: CagY, CagX, CagT, CagM, and Cag3 (Figure 1.4). While CagM and Cag3 are specific to the Cag T4SS, CagT, CagX and CagY are homologous to VirB7, VirB9 and VirB10 respectively. Visualisation of the isolated complex with negative stain electron microscopy showed that the OMCC is formed by an inner ring which is connected to an outer ring by spokes (Frick-Cheng et al., 2016). Complexes formed from a mutant strain lacking *cag3* or *cagT* retained the inner ring but lacked a well-formed outer ring, suggesting that these two proteins are located at the periphery. This was confirmed for Cag3 by immunogold labelling (Frick-Cheng et al., 2016).

Similarly, early work looking into the subcellular location of Dot/Icm components indicated the presence of a putative five-protein complex (Vincent et al., 2006). This was identified to be DotC, DotD, DotH, DotG and DotF. DotF and DotG are inner membrane proteins and join the OMCC in the presence of DotC, DotD and DotH which are crucial for formation of the complex (Vincent et al., 2006). DotH is associated with the outer membrane through interaction with the two lipoproteins DotC and DotD. The crystal structure has been solved for DotD which demonstrated that following a disordered region at the N-terminus, DotD adopts a globular fold which resembles the N0 domain that has also been characterised in the *X. citri* VirB7 (Nakano et al., 2010, Souza et al., 2011). Furthermore, DotH and DotG are homologs of VirB9 and VirB10 respectively. Thus, highlighting the remarkable structural conservation that exists amongst the T4SS.

The high-resolution structures of both core complexes have been recently solved using single particle cryo-EM which have been further resolved by addition of extra cryo-EM data (Figure 1.4) (Chung et al., 2019, Durie et al., 2020, Sheedlo et al., 2020, Sheedlo et al., 2021). The *H. pylori* core complex can be divided into three subsections: an outer membrane cap (OMC), periplasmic ring (PR) and a stalk. A symmetry mismatch was found to exist within the core complex with the OMC having 14-fold symmetry

and the PR adopting 17-fold symmetry (Figure 1.5) (Chung et al., 2019, Sheedlo et al., 2020). As observed in the minimised T4SS OMCCs, the Cag OMC also contains both an O- and I-layer. While all five core complex components form the OMC, only CagY and CagX extend downwards to the PR and structure it (Figure 1.4). The symmetry mismatch appears to be bridged by CagX whereby 14 rods of helical density was seen connecting the OMC and PR in an unsymmetrised cryo-EM map (Sheedlo et al., 2020). Therefore, it is thought that a total of 17 CagX proteins exist in the entire structure with all 17 forming the PR via the NTD, of which 14 CTDs extend upwards to form part of the OMC. This leaves 3 domains within the PR without a clear connection to the OMC which is currently unexplainable. The stalk was visualised at the base of the core complex, and the composition of this is yet to be confirmed but is thought to consist of CagY.

The Dot/Icm core complex structure is comparable to the equivalent complex in the Cag T4SS (Durie et al., 2020, Sheedlo et al., 2021). A symmetry mismatch also exists in the Dot/Icm complex which can be divided into regions which includes a 13-fold symmetric OMC and a PR which has 18-fold symmetry (Figure 1.5). Strikingly, the Dot/Icm OMC consists of an additional region, a 16-fold symmetric dome, which is embedded in the outer membrane (Figure 1.5) (Sheedlo et al., 2021). The overall structure resembles the Cag core complex in width (400 Å) but is more compact in terms of the height (164 Å). The dome is constructed from the CTD of DotG which is line with the observation that in a DotG deletion background, the complex lacks both the dome and the PR (Sheedlo et al., 2020, Sheedlo et al., 2021). This appears to be a conserved feature of VirB10 and its counterpart and has been observed in all VirB10 and homologous proteins resolved to date. A surprising difference between DotG and VirB10 is that absence of DotG does not abrogate formation of a complex (Vogel et al., 1998). Although, in this context, a central channel is missing from the core complex which analogous to the T4ASS OMCCs where VirB10 forms the central channel by lining the walls of the complex. While it was previously shown that the core complex is formed from five proteins, the cryo-EM structure revealed additional components which were formerly unknown (Sheedlo et al., 2021). The OMC is formed from the core Dot proteins, DotC, DotD, DotH, DotF and DotG along with other proteins (Figure 1.4). This includes DotK and three unrecognized T4BSS related components: Dis1, Dis2, and Dis3 which are also referred to as Lpg0657, Lpg0823, and Lpg2847 respectively. From all those proteins, only DotG, DotF and DotH appears to extend downwards to the PR. Here, a short loop belonging to DotG makes contact with a globular portion of DotH in a manner that has been reported between CagX/CagY and VirB9/VirB10 (Sgro et al., 2018, Sheedlo et al., 2020). Like in the Cag core complex, the VirB9 counterpart, DotH bridges the symmetry mismatch between the OMC and the PR. This is facilitated by a total of 18 DotH proteins existing in the entire core complex with all 18

forming the PR via their NTDs. Of the 18, 13 DotH CTDs extend upwards to partake in structuring the OMC. Remarkably, the remaining 5 DotH domains extend halfway and stretch into the surrounding space. Here, the 5 CTDs have been reported to be flexible given that these regions are not static in relation to the OMC and PR (Sheedlo et al., 2021). Even though a similar bridging of the symmetries was observed for the Cag core complex via CagX, the presence of similar densities corresponding to the remaining 3 domains in the intervening space was not observed. Subsequently, it has been hypothesised, that this difference and the flexibility that is attributed by it may be how the Dot/Icm machinery secretes such a large array of substrates (Sheedlo et al., 2021).

The cryo-EM structures of the Dot/Icm and Cag core complexes also provide insight into how these core complexes are so large. The Cag core complex protein exist at a stoichiometry of 1:1:2:2:5 (CagY:CagX:CagT:CagM:Cag3) with entire structure consisting of 14 copies of CagY and CagX, 28 copies of CagT and CagM and 70 copies of Cag3 (Chung et al., 2019, Sheedlo et al., 2021). Thus, a total of 154 proteins forms the Cag OMCC. The stoichiometry of the Dot/Icm core complex is 13:13:13:13:13:18:18:26:32 (DotC:DotK:Dis1:Dis2:Dis3:DotG:DotH:DotD:DotF), meaning that the complex is comprised of 159 polypeptides (Sheedlo et al., 2021). Together, this highlights the intricacy that exists within the core complexes from the expanded T4SS and helps explain the large size they adopt.

1.5.3.2 The inner membrane complex and the energy centre

The inner membrane complex (IMC) provides passage of the substrate across the inner membrane. The complex also plays an important role in coordinating the transfer as the ATPases associated with the machinery are located at or near the inner membrane.

Cryo-ET coupled with mutational analyses have provided insight into the Cag T4SS IMC and energy centre. A preliminary cryo-ET structure showed 4 barrel-like densities protruding into the cytoplasm which were proposed to be side-by-side hexamers of the VirD4 homolog CagE, as had been initially reported for the R388 T4SS (Low et al., 2014, Chang et al., 2018). Nevertheless, a better resolved structure established that the IMC adopts 6-fold symmetry and has a 3-ringed architecture (Hu et al., 2019b). The 3 concentric rings have been denoted as the I-, M- and O-ring which have a diameter of approximately 120, 220, 360 Å respectively. Several Cag components have been associated with the inner membrane. This includes the following: three ATPases (VirB4-like CagE, VirB11-like Cagα and VirD4-like Cagβ), inner membrane proteins with VirB homologs (VirB6-like CagW and VirB8-like CagV) and Cag specific inner membrane

proteins such as CagF, CagU and CagH (Fischer, 2011). Nevertheless, it remains elusive whether these Cag specific inner membrane proteins form part of the core secretion machinery or have other functions.

To decipher the contribution of certain Cag proteins to the IMC, mutant machineries were visualised by cryo-ET (Hu et al., 2019b). This approach provided insight into the involvement and structure of the three ATPases. Furthermore, this work was ground-breaking as it was the first time the energy centre was visualised for a T4SS where the three ATPases are present and contribute to the structuring of it. The VirB4-like CagE assembles as a hexamer of dimers at the cytoplasmic channel entrance. Furthermore, this ATPase was found to be integral for the assembly of the cytoplasmic region of the IMC as this was completely absent in the ΔCagE mutant. The VirB11-like Cagα appears to form a hexamer that docks at the base of the CagE and by establishing direct contacts with it, Cagα forms part of the I-ring. Lastly, the deletion mutant of VirD4-like Cagβ resulted in the loss of density for the O-ring as well as parts of the M- and I-rings. Considering that it is unlikely that VirD4 constitutes for all that density, it has been suggested that other proteins also form part of the IMC and may be stabilised by VirD4. Plausible protein candidates that shape the IMC are CagV and CagW, which are counterparts of VirB proteins (Fischer, 2011).

Strikingly, in the mutant machineries lacking either one of the three ATPases, the Cag OMCC was still present, implying no effect on the assembly or stability of the core complex (Hu et al., 2019b). Together, the results support an out-to-in assembly of the Cag T4SS. The OMCC assembles first and acts as the foundation for the formation of the inner membrane platform, composed of CagV and CagW. The energy centre establishes at the cytoplasmic side of the IMC. This occurs through the sequential recruitment of the ATPases, whereby CagE bind first, followed by the docking of Cagα, and in-turn recruitment of Cagβ (Hu et al., 2019b).

Like the Cag T4SS, the IMC belonging to the Dot/Icm machinery is important for recognition and transfer of effectors, and also, provides a platform for the docking of the Type IV coupling complex (T4CC). Three ATPases are located here which consists of DotO, DotB, and DotL which are homologous to VirB4, VirB1, and VirD4 respectively. DotO integrates with the IMC and recruits DotB, which is mostly cytosolic (Sexton et al., 2005, Chetrit et al., 2018). Strikingly, DotB shares higher homology with the ATPases PilT and EspE from the Type IV pilus system and T2SS respectively, than it does with VirB11 (Prevost and Waksman, 2018). This alludes to the Dot/Icm apparatus having a mosaic nature and also hints towards an evolutionary connection (Christie, 2016).

The Dot/Icm inner membrane proteins, DotU and IcmF, have been shown to be important for the polar targeting of the secretion system and are thought to provide a stable scaffold on which the T4SS is assembled (Ghosal et al., 2019). Moreover, the inner membrane proteins IcmT, IcmV, IcmQ, DotA, Vir8-like DotI, DotU, DotJ and DotE are required for DotB to be recruited to the cell pole (Chetrit et al., 2018). Out of these proteins, DotU and DotI in particular have been found to be required for polar localisation of DotO. Together, this suggest a successive assembly of the IMC whereby an IMC subcomplex is initially constructed which then provides a platform for the formation of the energy centre.

Once DotO and DotB have associated with the IMC, they provide important structural features to the T4SS. Initially, DotO densities protruding into the cytoplasm were considered to be the ATPase forming side-by-side hexamers, in a manner reminiscent to the VirB4 from R388 and CagE from the Cag T4SS (Low et al., 2014, Ghosal et al., 2017, Chang et al., 2018). However, with higher resolution, the DotO was found to arrange as a hexamer of dimers at the base of the T4SS apparatus (Chetrit et al., 2018, Park et al., 2020). This arrangement of the ATPase is becoming increasing prevalent as it has also been seen for the VirB4-like CagE, and also recently for VirB4 from both R388 and pKM101 (Hu et al., 2019b, Khara et al., 2021, Mace et al., 2022). An additional density was observed below the V shaped DotO which was assigned to DotB (Chetrit et al., 2018). DotB associates directly with DotO as a hexamer. Together, DotO and DotB shape a cylindrical structure at the cytoplasmic channel entrance of the Dot/Icm T4SS. This docking of DotB to DotO is important for effector transfer as it has been shown to cause a conformational change that results in the opening of the channel (Park et al., 2020).

The effectors are delivered to the Dot/Icm machinery by the cytoplasmic VirD4-like ATPase, DotL (Vincent et al., 2012). DotL also plays a significant role in recognition of the substrate by forming the T4CC along with additional proteins (Lockwood et al., 2022). It appears that the composition of this complex may depend on the type of effector being shuttled to the T4SS. One such T4CC displayed the C-terminus of DotL contacting a heterodimer composed of the adaptor proteins IcmS and IcmW (IcmSW) (Figure 1.9A) (Sutherland et al., 2012). Two additional proteins, DotN and LvgA, were found to be associated with this complex resulting in a structure consisting of DotL-DotN-IcmSW-LvgA (Figure 1.9B). Using the hexameric structure of the VirD4 from R388 (TrwB) as the template, a hexameric model of the DotL-DotN-IcmSW-LvgA complex was constructed and referred to as the holocomplex (Kwak et al., 2017, Kim et al., 2020). The holocomplex assumes an elongated bell-shaped structure with the DotL ATPase membrane-proximal and the adaptor proteins membrane-distal. The adaptor proteins participate in the recognition of the

effectors. The substrate SidF has been shown to interact with IcmSW, with other effectors such as VpdB, binding to LvgA in the holocomplex (Figure 1.9C) (Xu et al., 2017, Kim et al., 2020).

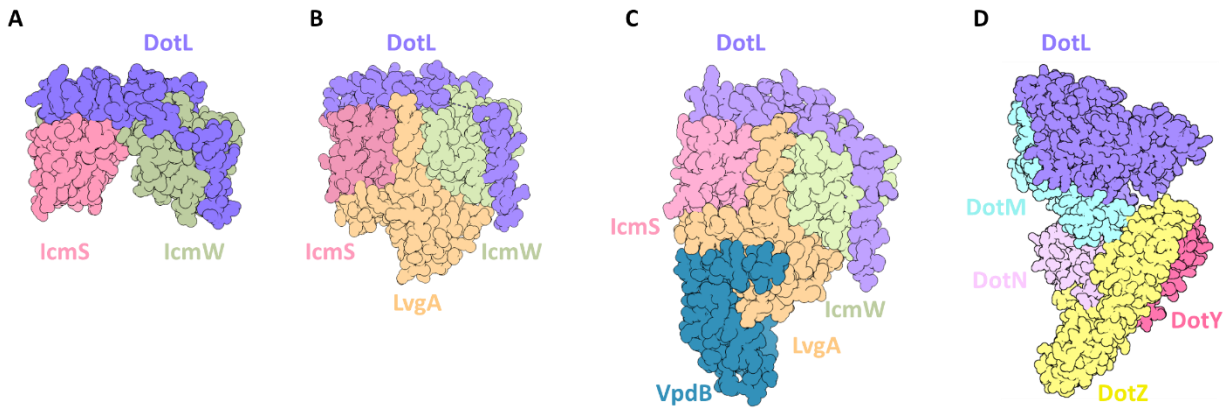

Figure 1.9: Different structures of the Dot/Icm Type IV Coupling Complex involved in effector recruitment. **(A)** The DotL-IcmSW complex, PDB 5XNB **(B)** The DotL-IcmSW-LvgA complex, PDB 5X90 **(C)** The DotL-IcmSW-LvgA-VpdB complex, PDB 7BWK **(D)** The DotLMNYZ complex, PDB 6SZ9.

Alternatively, a second effector recognition complex has been described which includes DotM. This has been suggested for effectors that contain the Glu-rich/E-block motif based on the observation that the crystal structure of DotM exhibited regions of positively charged residues which may provide binding sites for the negatively charged motifs (Meir et al., 2018). A T4CC formed by DotL, DotM, DotN, IcmSW and two additional proteins, DotY and DotZ, was isolated (Meir et al., 2020). The structure for the DotLMNYZ complex was solved using cryo-EM (Figure 1.9D). The hetero-pentameric complex was modelled as a hexamer with the resulting structure resembling a starfish. Inside the complex, a cavity was detected that was outlined by DotM, DotN and DotY. Therefore, effectors that are transported by DotM may enter the T4CC by associating with DotM and traversing through the cavity to get to the DotL channel. Thus, there appears to be two possible models at present, whereby the substrates are recruited to the T4CC at the cytoplasm via interaction with IcmSW, LvgA and/or DotM. These potential modes of effector capture may not be mutually exclusive. Furthermore, how the T4CC then facilitates the transfer of the effector to the DotL and then the T4SS still remains elusive. Two possible scenarios have been possible: (1) the effectors pass through the channel formed by the DotL central channel and is delivered directly to the T4SS, (2) the T4CC recruits the effectors which are then transferred to the DotO-DotB complex at the cytoplasmic base

of the translocation apparatus (Lockwood et al., 2022). Nonetheless, the exact connection of DotL to the core T4SS machinery remains to be addressed.

1.6 The F-type T4SS

The T4SS encoded by the iconic F-plasmid is also classified as an expanded system or T4BSS.

1.6.1 The F plasmid and its role in antibiotic resistance

Described in 1946, Edward Tatum and Joshua Lederberg first demonstrated conjugation by showing the ability of bacteria to transfer DNA in a unidirectional manner (Lederberg and Tatum, 1946). Initially referred to as the fertility factor (F-factor), it was soon established that the F-factor is a replicative extranuclear structure, and the term plasmid was coined.

The F plasmid is a large 100-kb conjugative vector found in *E. coli*. Over 200 different F-like plasmids have been sequenced and have been identified from closely related enterobacterial genera including *Klebsiella* and *Shigella* (Koraimann, 2018). In general, the F and F-like plasmids consist of four functional modules: (i) DNA transfer genes, (ii) cargo genes, (iii) a region containing genes involved in replication and stability and (iv) leading region (Figure 1.10A). Particularly noteworthy is the cargo region which contains the mobile genetic element, such as transposons and ICE, and confers phenotypic traits to the bacteria. The genes encoded in this region are called the cargo genes, and they provide survival advantages and increases genetic diversity within the bacterial population. Examples of traits that cargo genes found in F and F-like plasmids encode for include metal resistance, virulence, protein toxins such as colicin, and antibiotic resistance (Koraimann, 2018).

From a global health perspective and with the rapid dissemination and persistence of antibiotic resistance, arguably the most concerning aspect of F and F-like plasmids is the presence of antibiotic resistance genes. This is exacerbated by the presence of multiple resistance genes accumulating on the same plasmid, often observed in F-like plasmids. This results in multi-drug resistance strains. Furthermore, F-like plasmids are present in pathogenic bacteria that are responsible for human diseases such as urinary tract infection and typhoid fever. The threat posed by resistant bacteria which harbour F and F-like plasmids is demonstrated by the appointment of various enterobacteria as critical priority for the research and development of new antibiotics (World Health Organisation, 2017). A prominent example is the global dissemination of

carbapenem resistant *K. pneumoniae* which has been largely attributed to a F-like plasmid (Lee et al., 2016). Thus, demonstrating the pivotal role of conjugation. Furthermore, resistance of *E. coli* to the last resort antibiotic, colistin, has been reported in clinical isolates (McGann et al., 2016). The F-like plasmid conferring the resistance, also contained seven other resistance genes. Therefore, alarmingly, it appears that pan-drug resistance is becoming increasingly prevalent. The impact of antibiotic resistance is highlighted by an estimated 4.95 million deaths in 2019 attributed to resistant bacteria (Murray et al., 2022). With this having been predicted to rise to estimates of 10 million deaths per year by 2050 due to drug-resistant diseases, antibiotic resistance will not only be detrimental to health worldwide, but it will also have a catastrophic effect on the global economy with the devastation being compared to that of the 2008-2009 financial crisis (World Health Organisation, 2019). This highlights the importance of developing new therapeutics in an effort to tackle antibiotic resistance. One such tactic would be to understand new components that are essential for conjugation to occur.

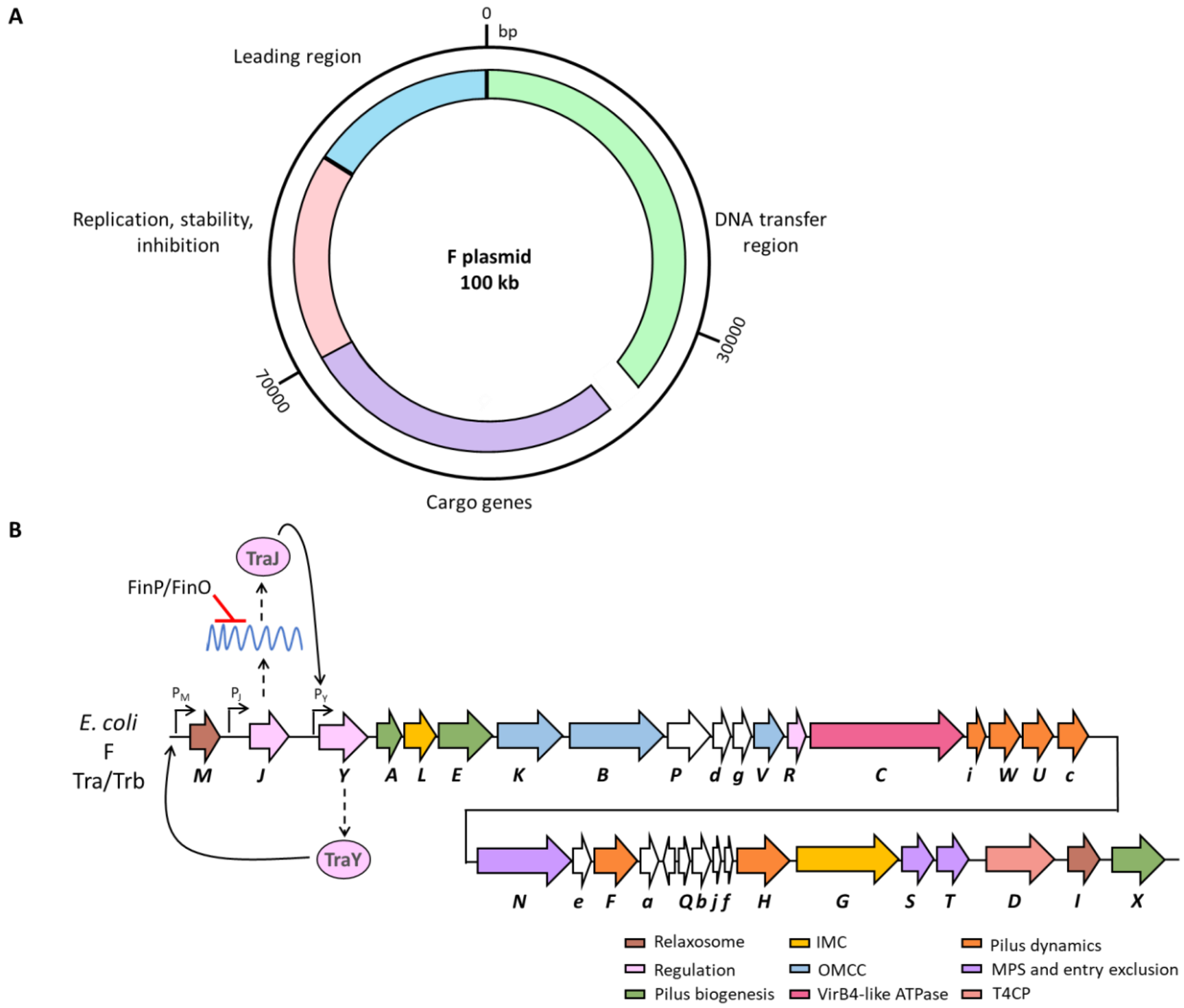

Figure 1.10: The F plasmid. (A) Schematic of the F plasmid showing the four functional modules. **(B)** The genetic organisation and expression of the F plasmid encoded T4SS. The genes encoding for the apparatus are shown in arrows coloured according to the function the protein is involved in. The small arrows indicate the promoters which drive the expression of the genes following the arrow. OMCC (Outer membrane core complex), IMC (Inner membrane complex), T4CP (Type IV coupling protein), Mps (Mating pair stabilisation)

1.6.2 The expression of the *tra* operon

Necessary for the spread of the antibiotic genes is the F-T4SS, which is encoded on the DNA transfer (Tra) region that is approximately 33 kb in size (Figure 1.10A) (Frost and Koraimann, 2010). The genes expressing the proteins which form the F-T4SS are organised within the *tra* operon and expressed under the control of one promoter, P_Y (Figure 1.10B). Upstream of the *tra* operon is the *traJ* and *traM* genes which are under the control of independent promoters. TraJ is the key transcriptional activator for the

expression of the *tra* operon by activating the P_Y promoter (Figure 1.10B). Following action of TraJ, the first protein expressed is TraY which is a regulator and promotes TraM expression. TraM is required for DNA processing during the early stages of conjugation. A key regulatory element of TraJ is the FinOP repressor system which downregulates TraJ translation. This occurs through the direct binding of FinP to the stem loop structures of TraJ mRNA; resulting in the masking of the ribosome binding site (Virolle et al., 2020). FinO associates with FinP and protects it from degradation by RNase. Interestingly, in some plasmids the FinOP system is disrupted and has been termed the super spreader mutation (Virolle et al., 2020). A notable example is the F plasmid which contains an IS3 insertion sequence in the *finO* gene resulting in almost undetectable levels of FinP (Cheah and Skurray, 1986, Frost and Koraimann, 2010). This naturally occurring mutation results in the F plasmid being derepressed and having increased transfer efficiency due to the overexpression of the *tra* operon, compared to the other F-like plasmids which have a functioning FinOP systems such as R100 and R1. The F-like plasmid pED208, which is becoming the model system to study components of the conjugative apparatus, is also derepressed. However, due to the presence of an additional insertion sequence located within *traY*, the *tra* operon is constitutively overexpressed resulting in the bacteria having a multi-piliated phenotype (Lu et al., 2002). Therefore, this confers the ability to increase expression of the *tra* operon and presumably assemble more T4SS by overriding regulatory mechanisms, exemplified by a more piliated phenotype c. Thus, it isn't surprising that the pED208 encoded F-like T4SS has become the ideal system in which to study conjugation and the transfer machinery, especially from a structural standpoint (Costa et al., 2016, Hu et al., 2019a, Amin et al., 2021, Liu et al., 2022).

The expression of the *tra* operon is also regulated by other factors such as the activation of P_Y, which is silenced by the chromosomally encoded histone-like nucleoid structuring protein (H-NS). This is counteracted by TraJ and the host encoded protein ArcA (aerobic respiration control or anoxic redox control). Both proteins directly activate P_Y in a cooperative fashion (Lu et al., 2019). The levels of H-NS within the cell varies according to growth with the highest being in the exponential phase and lowest at the stationary phase (Virolle et al., 2020). Consequently, this results in the expression of the F-T4SS being growth-phase dependent.

1.6.3 The protein components of the F-T4SS

Spanning from *traY* to *traX*¸ the *tra* operon encodes over 30 proteins of which 9 are homologous to the protein found in the VirB/D4 system (Figure 1.10B). Based on those that have been characterised, the proteins can be categorised into two groups: (i) core T4SS proteins (TraA, -L, -E, -K, -B, -V, -C and -G) which form a structural component of the translocation channel and (ii) auxiliary proteins (TraF, -H, -N, -U, -W and TrbC, -I) which are exclusive to the F-T4SS and confer characteristics unique to the F conjugative apparatus (Lawley et al., 2003). This includes pilus extension and retraction which enables dynamic movement of the pilus, and mating pair stabilisation (Mps) which facilitates efficient transfer of DNA. Overall, the proteins encoded on the *tra* operon can be divided into the following structural and functional groups: the core T4SS apparatus, pilus dynamics, relaxosome, regulation (described in section 1.6.2), Mps and lastly, entry exclusion (Figure 1.11). As shown by the cryo-ET structure, the core apparatus can be further split into the following sections: the pilus, OMCC and IMC (Figure 1.11) (Hu et al., 2019a).

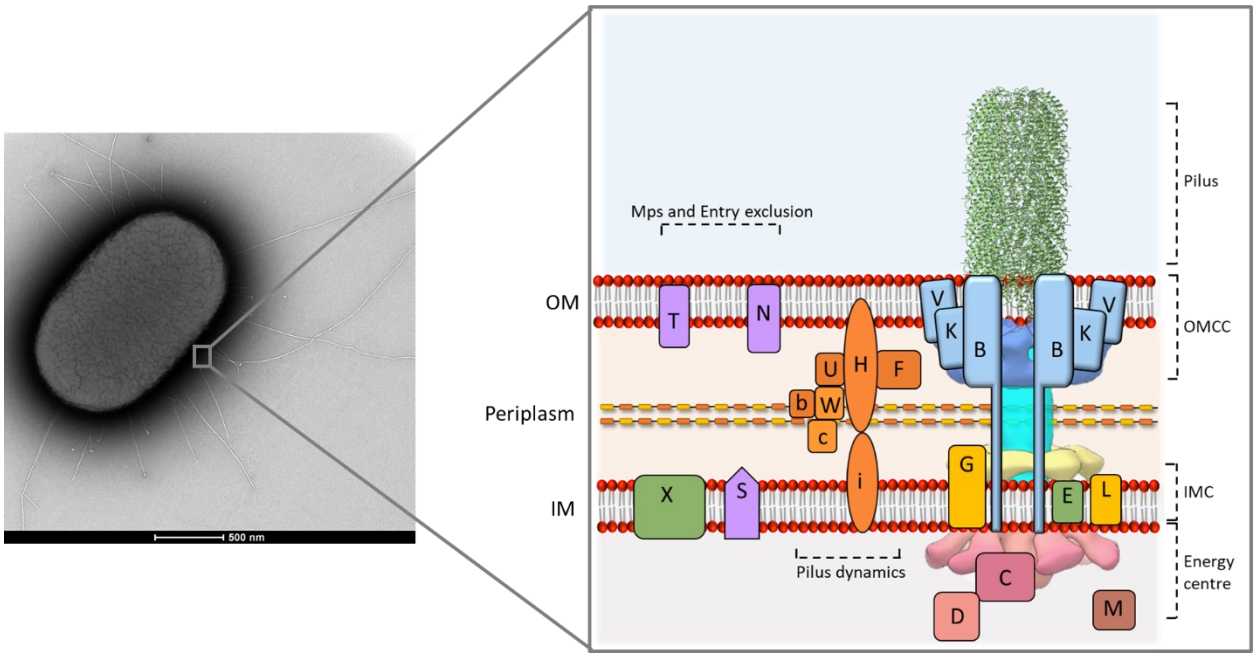

Figure 1.11: The F Type IV Secretion System (F-T4SS). A micrograph of an *E. coli* cell harbouring the pED208 plasmid leading to the expression of pili on the surface. At the base of the pilus is the F-T4SS. The schematic highlights protein localisation within the F-T4SS and the proteins are coloured as in Figure 1.10. The available structures have been shown which includes the cryo-electron tomography structure of the T4SS (EMD-9344) and the cryo-electron microscopy model of the pilus (PDB: 5LEG). The scale bar is 500 nm.

1.6.3.1 The F-pilus

The conjugative pilus is a hallmark of conjugation and is the extracellular filament which is observed on the bacterial cell surface (Figure 1.11). The pilus is used by the donor cell to contact a potential recipient cell. Thus, enabling establishment of a mating junction via cell-to-cell contact which is resistant to sodium dodecyl sulphate (SDS) and shear forces (Arutyunov and Frost, 2013). It is a highly dynamic filament and studies using real-time fluorescence microscopy have shown that the pilus undergoes cycles of extension and retraction (Clarke et al., 2008). Strikingly, this seems to occur without an obvious trigger event which, coupled with the flexibility of the F-pili, would enable efficient sampling of the surrounding space in efforts to contact a potential recipient cell. Once contact is established, the force produced by the retraction of the pilus is sufficient to pull the cells towards one another.

The precise function of the pilus continues to be at the centre of controversy in the field. Alongside the role of bringing the donor and recipient cells together, it has been proposed that the F-pilus may act as a conduit for the passage of the ssDNA in a manner that is cell-to-cell contact independent. Evidence in support of this hypothesis comes from an experiment using fluorescently labelled DNA which showed that the DNA could be transferred between cells that are spatially apart (Babic et al., 2008). An explanation for this has been that the DNA is transported via the pilus. Nonetheless, this phenomenon was seen in a small proportion of the cells analysed which suggests that the main mode of DNA transfer is via the cell-to-cell contact between the donor and recipient cells. Furthermore, parallel to its role in conjugation, the F-pilus has also been implicated in biofilm formation. *E. coli* strains harbouring the F-plasmid were found to be adept at forming biofilms, however, this ability was lost in the absence of the plasmid (Ghigo, 2001). This capability to form biofilms was attributed directly to the F-pilus that is encoded on the plasmid, indicating these filaments can act as adhesion factors. Furthering this, a recent study has demonstrated that the production of biofilm is greater when the donor and recipient cells are combined compared to when the two cells are separated or the recipient cell is absent (Patkowski et al., 2023). This provides a direct connection between two roles that involve cell-to-cell contact: conjugation and biofilm formation.

The F-pilus is composed of mature TraA which polymerises as a helical filament. The maturation of TraA involves the pro-pilin being transported to the inner membrane via TraQ (Hospenthal et al., 2017). Following cleavage of the leader peptide from the pro-pilin by LepB peptidase, the mature TraA pilin then undergoes acetylation by TraX at the N-terminus. During maturation, TraA pilin binds to a phospholipid, and this protein-phospholipid complex is then assembled into helical filament to form the pilus (Hospenthal et al., 2017). The structure of the pED208 and pOX38 pili have been solved which

demonstrated that both pili are almost identical (Costa et al., 2016). The structure of the pED208 pili was resolved at a higher resolution of 3.6 Å (Figure 1.11). The pilus is a five-start helical filament with an overall dimension of 87 Å in diameter and with an internal lumen diameter of 28 Å. The TraA pilin was found to fold into an all α-helical structure, containing three α helices. The TraA is orientated as such within the filament that a loop connecting α2 to α3 protrudes inside the lumen of the pilus. Interestingly, this particular loop has been hypothesised to be capable of making contact with DNA (Silverman, 1997). Within the vicinity of TraA, an additional density was observed which has been attributed to phospholipids, specifically phosphatidylglycerol species (Costa et al., 2016). The TraA and phosphatidylglycerol interact forming a complex with a stoichiometry of 1:1. The presence of the phospholipid units is integral to the pilus. Firstly, it impacts the electrostatic potential of the pilus lumen by making it moderately electronegative. Without the phosphatidylglycerol units, the lumen would otherwise be overwhelmingly positively charged. The charge within the lumen together with the positioning of the loop support the notion that the pilus may also act as a channel for the delivery of the ssDNA. Secondly, the phospholipids provide great structural stability to the filament. Using molecular dynamics stimulations, it has been demonstrated that in the absence of the phospholipids, the pilus is more fragile and succumbs to structural failure upon exposure to pulling forces (Patkowski et al., 2023). Other conjugative pili that have been solved are also composed of pilin-phospholipid units that exist with a stoichiometry of 1:1, thus highlighting the conservational significance (Zheng et al., 2020, Amro et al., 2023, Kreida et al., 2023).

The *in situ* cryo-ET study of the pED208 encoded F-T4SS has provided insight into the assembly of the pilus with several F-encoded structures being reconstructed (Hu et al., 2019a). In one such structure, called F2-channel/pilus, the pilus is attached to the outer membrane but does not extend across it. Coupled with mutational analyses, it has been suggested that the pilus is built at the outer membrane. This distinguishes the F-pilus from the type IV pili and the needle complexes from the T3SS, as the assembly of both are nucleated at the inner membrane (Chang et al., 2016, Hu et al., 2017). This observation was surprising as it was previously thought that the polymerisation of the pilus commences at the inner membrane platform given that this is the location where the pilin pre-processing occurs. In an effort to provide an explanation for this, the model that has been proposed by Hu and colleagues suggests that at the inner membrane, the TraA assembles to form a thin protofilament (Hu et al., 2017). The protofilament then extends upwards through the T4SS machinery by the sequential addition of TraA at the base. As it passes through the outer membrane, the protofilament polymerises into a helical arrangement that is associated with the mature pilus. Another notable structure that was observed showed the pilus without a T4SS at the

base. While this only represented a small percentage of the structures observed (5%), it may correspond to pili that are involved in biofilm formation, given that conjugative properties require the presence of the T4SS machinery.

1.6.3.2 The F outer membrane core complex

The cryo-ET reconstruction of the F2-channel/pilus structure provided the first insight into the general architecture of the F outer membrane core complex (F-OMCC) (Hu et al., 2019a). Strikingly, the complex has 13-fold symmetry and overall dimensions of 250 Å in width and 115 Å in height with the shape resembling a "flying-saucer". This symmetric arrangement for the OMCC has only been observed in the Dot/Icm machine so far (Ghosal et al., 2017). While the resolution was not sufficient enough to resolve the structural details of individual proteins, an understanding of which proteins comprise the F-OMCC has been shaped by protein-protein interaction studies. The F-OMCC is composed from three proteins, TraV, TraK and TraB which are homologous to VirB7, VirB9 and VirB10 respectively (Harris et al., 2001). Through evidence provided by yeast two-hybrid studies, it has been established that the three Tra proteins form a membrane spanning complex, with the lipoprotein TraV predicted to be a membrane anchor (Doran et al., 1994). In an interaction similar to that observed in the *X. citri* OMCC, TraK was found to be the bridging protein between TraV and TraB (Sgro et al., 2018). Thus, highlighting remarkable structural conservation amongst OMCCs that are implicated in different functions. Nevertheless, a high-resolution structure is required to better understand the stoichiometry and interactions between the F-OMCC proteins. This is important not only to understand the function of the F-OMCC and how the structure enables it, but also to decipher how steps of conjugation proceed and the role of the complex within this.

1.6.3.3 The periplasmic complex

Aside from the F-OMCC, a second group of interacting proteins have been identified by yeast two-hybrid analyses. This interacting group is formed from the proteins TraH, -U, -F, -W, TrbC and the inner membrane protein TrbI (Harris and Silverman, 2004). While TraU, -F, -W and TrbC are located in the periplasm, TraH is thought to be peripherally associated with the outer membrane (Arutyunov et al., 2010). For simplicity, this group of proteins will be referred to as the periplasmic group. The proteins TraH, -F, -W, -U and TrbC are involved in pilus extension, while TrbI and TraH have been implicated in pilus retraction (Moore et al., 1981, Anthony et al., 1999). Within the periplasmic group, TraH was the most

connected, thereby suggesting that this protein plays a key role in the function of the group. Members of the periplasmic group have been suggested to interact with the F-OMCC (Harris and Silverman, 2004, Arutyunov et al., 2010, Koch et al., 2020, Kishida et al., 2022). Since the pilus polymerises at the outer membrane and that the periplasmic proteins are involved in pilus associated functions, an interaction between members of this group and the F-OMCC may be how pilus extension and retraction are mediated.

1.6.3.4 The inner membrane complex

The F-OMCC surrounds a cylindrical channel that extends downwards to the inner membrane and connects with the IMC which adopts 6-fold symmetry (Hu et al., 2019a). The VirB4 homolog, TraC assembles as a dimer of hexamers at the cytoplasmic face of the IMC. Currently, other members of the IMC have not been experimentally confirmed, however a likely candidate is TraG. TraG appears to be a fusion of VirB6 and VirB8, which are IMC components in the VirB/D4 T4SS. The periplasmic C-terminus of TraG is homologous to VirB8 whereas the N-terminus is homologous to VirB6 (Audette et al., 2007). Nonetheless, this is yet to be proven and possible contributions of TraG towards structuring the IMC remains elusive.

As seen in the VirB/D4, Dot/Icm and Cag T4SSs, associated with the IMC and at the periplasm, a collar-like structure is present. The collar is also hexameric. In the R388 encoded VirB/D4 T4SS, a similar structure was constituted by VirB8. Therefore, it may be possible that the equivalent collar in the F-T4SS is formed by TraG.

1.6.3.5 TraD and the relaxosome

In the F-plasmid, TraD is the VirD4-like T4CP. While TraD was absent in the cryo-ET structure of the F-T4SS due to its transient nature of interaction with the machinery, all T4CPs share a similar domain arrangement (Hu et al., 2019a). Also observed in TraD, this entails a N-terminal transmembrane domain and a large cytoplasmic C-terminus which consists of an AAD and a NBD (Bragagnolo et al., 2020). TraD is presumed to arrange as a hexamer, in a similar fashion to the R388 VirD4 (Gomis-Ruth et al., 2001).

Before involvement of TraD, the DNA is processed by the Dtr proteins which comprise of the relaxase TraI and accessory proteins. This complex is also referred to as the relaxosome. The accessory proteins modulate activity of the relaxase, and includes the plasmid encoded TraM and TraY, along with another

accessory protein called integrated host factor (IHF). The proteins bind at various sites within the *oriT*, and notably, two TraI molecules have been demonstrated to bind either side of the nic site (Ilangovan et al., 2017, Waksman, 2019). TraI is a bifunctional protein. It functions as both a helicase, to unwind the plasmid DNA, and also as a transesterase that catalyses the nicking of the DNA. Modulated by the other Dtr proteins, the nicking of the DNA by the transesterase domain of TraI is facilitated by a conserved tyrosine residue. Following nicking, TraI remains covalently bound to the 5' end of the ssDNA.

The relaxosome is recruited to the T4SS transfer machinery by TraD. This is facilitated by TraD interacting with TraM in a highly specific manner (Lu et al., 2008, Wong et al., 2011). The C-terminus of TraM is responsible for tetramerisation of the protein. Through its C-terminal tail, TraD interacts with a hydrophobic pocket created by 3 out of the 4 TraM protomers (Lu et al., 2008). Surrounding the hydrophobic pocket are positively charged residues which interact with the carboxylate groups in TraD. Therefore, the contacts between TraM and TraD are governed by both hydrophobic and hydrophilic interactions, which is conserved amongst the F-plasmids. Furthermore, this interaction between the two proteins was found to be important for efficient conjugation to occur; presumably because it enables attachment of the relaxosome to the core T4SS (Bragagnolo et al., 2020). Notably, two regions that serve as translocation signals have been identified on TraI which may aid transport of the TraI to the T4SS (Lang et al., 2010). However, whether TraD binds to these signal sequences has not yet been shown.

1.6.3.6 Mating pair stabilisation and entry exclusion protein

Early steps of conjugation involve precautionary measures to ensure efficient DNA transfer. One such step is Mps. Mps occurs following pilus retraction and enables stabilisation of the contacts formed between the donor and recipient cells. This enables efficient DNA transfer by providing resistance to shear forces and harsh reagents such as SDS (Achtman et al., 1978). The two proteins which are important in Mps within the donor cells are TraN and TraG which are found located in the outer membrane and inner membrane respectively. TraN and TraG have been proposed to interact, which would span across both membranes and act as a clamp to secure contact with the recipient cell. Nonetheless, such an interaction remains to be proven experimentally. On the recipient cells, Mps is mediated by the outer membrane protein A (OmpA). TraN has been shown to interact with cognate OmpA and specificity has been attributed to a tip region on TraN (Klimke et al., 2005, Low et al., 2022).

Alongside Mps, exclusion proteins act to stop redundant plasmid transfer to a cell which already harbours the conjugative elements. The first preventative step to occur is surface exclusion (Sfx), which relies on the outer membrane protein TraT (Achtman et al., 1978). If TraT is present on the membrane of the recipient cell, mating pair formation is halted, and conjugation is arrested. If Sfx is unsuccessful, and Mps occurs, a second exclusion step is depended on to prevent redundant substrate transfer. This is Eex and is mediated by the donor cell TraG. This mechanism relies on the expression of TraS by the recipient cell which can only occur if it also contains an F-plasmid. TraG has been proposed to recognise the TraS found at the inner membrane (Audette et al., 2007). However, how an interaction between two inner membrane proteins across two cellular envelopes is achieved has not yet been explained. Strikingly, TraG appears to be at the centre of both Mps and Eex steps. The sparsity of structural details available for TraG limits the ability to decipher how one protein is involved in various different roles.

1.6.4 The importance of studying the F-T4SS

With the rapid rise of antibiotic resistance and emergence of multiple resistant pathogenic bacteria, attention has been drawn to secretion systems, specifically the T4SS, both as a facilitator for the dissemination and as a promising target for the development of therapeutics. Furthermore, unlike traditional antibiotic treatments, those targeting bacterial secretion systems, in principle, would be selective for pathogenic strains. Thus, maintaining the mammalian bacterial flora. One such area of research has focused on discovering and developing conjugation inhibitors. A high-throughput screening approach focused on the *Brucella suis* VirB8, has yielded a small molecule inhibitor that prevents protein dimerisation (Smith et al., 2012). A similar inhibitor for the F-T4SS has not yet been established. A contributing factor is the lack of understanding regarding the assembly of the F-T4SS and the structural details regarding the individual protein components.

Due to the extensive use of antibiotic and the developing resistance, another treatment strategy that has gained renewed interest is phage therapy. As viruses that infect bacteria, bacteriophages provide a promising alternative strategy against antibiotic resistant bacteria. Some of these phages use pili mediated by the T4SS to attach to the bacteria during infection and therefore, is the first point of entry for the virus (Arutyunov and Frost, 2013, Costa et al., 2016). The phages which specifically attach to the pilus, also referred to as sex pilus specific phages, pose as exciting, prospective antibacterial agents (Spankie et al., 2020). They affect bacteria which contain pili expressing transmissible plasmids, which often also harbour antibiotic resistance. Subsequently, infection of these bacteria with bacteriophages

may reduce the carriage of such plasmids in the bacterial population. In theory, this will not only result in infection of the bacteria carrying such plasmids, but if the bacteria were to become phage resistant, this would occur due to the loss of plasmid. Thus, making the bacteria susceptible to antibiotics if the plasmid contained resistance genes (Lusiak-Szelachowska et al., 2022). This has been shown to be effective using the F-plasmid specific phage MS2 which was found to kill ampicillin resistant *E. coli* and select for plasmid loss (Colom et al., 2019). As sex pilus specific phages continue to be explored as antibacterial agents, broadening understanding of pilus assembly, dynamics, and the involvement of the F-T4SS will be invaluable as it may provide ways to shift sensitivity to phage infection.

1.7 Cryo-electron microscopy

Cryo-EM enables study of large protein complexes at high resolution. Prior to its advancement, X-ray crystallography was one of the main techniques employed for structural determination of proteins. While X-ray crystallography has been useful in structural biology, there are limitations to the sort of protein that is suitable for study by this technique. The protein is needed at high yields and is required to form well ordered and well diffracting crystals. This is difficult to achieve for proteins or protein complexes that are large, or membrane embedded. For such samples, cryo-EM can be an viable option for structural investigation. Within cryo-EM there are two main streams: single particle analysis (SPA) and cryo-electron tomography (cryo-ET). Cryo-ET allows for the structural determination of proteins *in situ*, with the protein being studied in thinned areas of the cell. Alternatively, SPA relies on purified protein being vitrified in ice, and involves collection of thousands of 2D images of the protein, which is then aligned, averaged and refined to reconstruct the 3D object. In this thesis, SPA cryo-EM was used, therefore this will be the focus of this section.

1.7.1 The transmission electron microscope and image formation

1.7.1.1 The basic anatomy of the TEM

Due to their short wavelength (2.25 pm at 300 keV) and strong interaction with biological matter, imaging with electrons provides the benefit of achieving high resolution (Orlova and Saibil, 2011). Within the transmission electron microscope (TEM), electrons are accelerated from an electron gun, situated at the

top of the microscope, down the microscope column. Since electrons interact strongly with matter, to prevent undesirable scattering by gas molecules, the column is maintained under high vacuum (Orlova and Saibil, 2011). A series of lenses, termed the condenser system, focus the electron beam onto the specimen (Figure 1.12A). Upon contact with the specimen, the electrons are scattered. The objective lens system refocuses the scattered electrons onto the back focal plane and also provides the primary magnification. The image is further magnified by the projector lens system, before the electrons reach the detector.

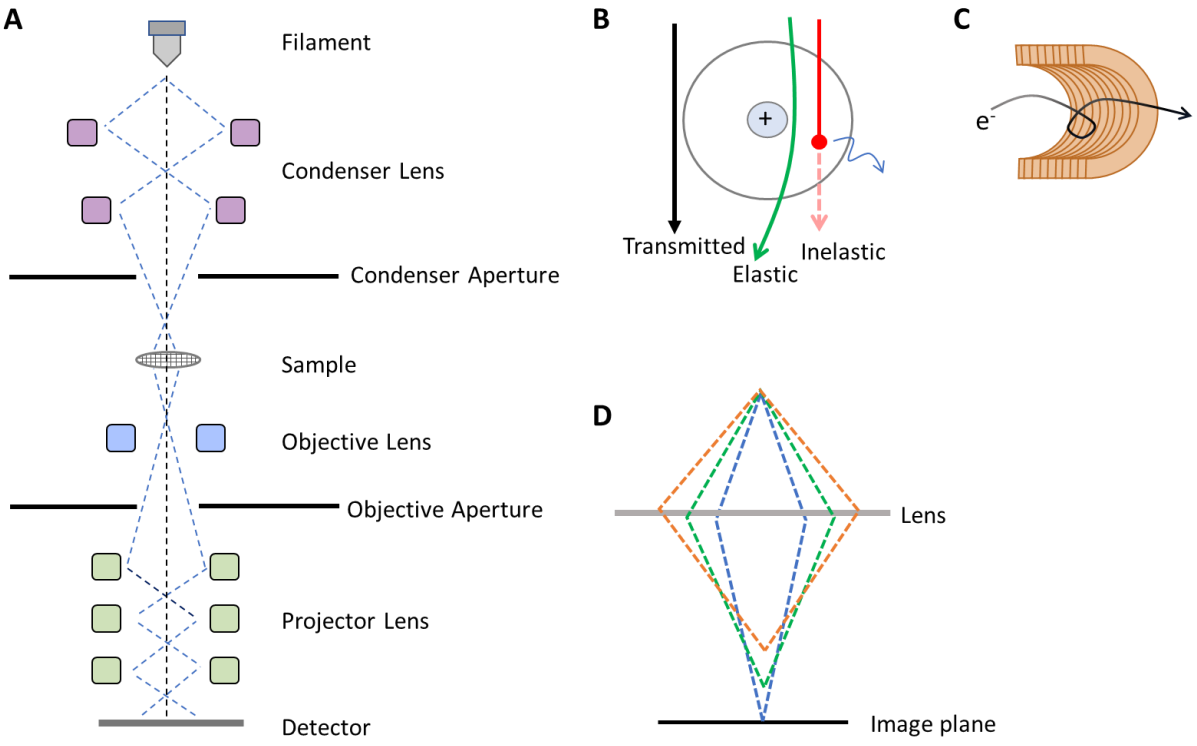

Figure 1.12: The transmission electron microscope (TEM) and image formation. (A) A schematic representation of an TEM. **(B)** A schematic of electrons interacting with the specimen. If the electron interacts with the atoms in the specimen, it is either elastically scattered whereby its path is altered without energy loss, or it is scattered inelastically whereby it deposits energy into the sample. **(C)** Electrons are focussed by electron lenses that are made from large copper wires. **(D)** A schematic representing the spherical aberration suffered by electron lenses.

1.7.1.2 The interaction of electrons with the specimen

Upon interaction with the sample, there are three possible outcomes which determine the path of the electron (Figure 1.12B). In the first scenario, the electron does not come into contact with any atoms within the specimen, and therefore pass straight through without being scattered (Orlova and Saibil, 2011). The second and third scenario result from scattering events which change the path of the electron. As negatively charged particles, the travelling electrons can be deflected by the electrostatic field of the positively charged nuclei of the atoms in the specimen. Of these affected electrons, some are scattered elastically whereby their path is altered without loss of energy. However, electrons can also be scattered inelastically in which they deposit energy into the sample and cause radiation damage. Roughly 75% of electron scattering events are inelastic (Henderson, 1995). The loss of energy due to the inelastic scattering of the electron, also causes a change in its wavelength. Consequently, the electron is incorrectly focused and contribute only noise. The deflected electrons can be refocused by the lens systems, thereby contributing useful signals to the final image.

1.7.1.3 Electron lenses

Within the TEM, electrons are focused by electron lenses (Franken et al., 2020). These are large coils of copper wires, through which a current is passed. This creates an electromagnetic field. As the scattered electron passes through the lens, it encounters a force directed at the centre of the lens. This results in the spiralling of the electron as it accelerates through the lens, and eventually refocuses the electron onto the image plane (Figure 1.12C).

Due to imperfections, the electron lenses also have defects, namely spherical and chromatic aberrations, and astigmatism (Orlova and Saibil, 2011, Franken et al., 2020). Spherical aberration describes the situation where electrons that have been scattered at varying spatial frequencies will not all be focused on the same image plane. This means that highly scattered electrons compared to those with smaller scattering angles, are focused with a sharper angle onto the imaging plane, which results in the over-focussing of this electrons by the lens (Figure 1.12D). Chromatic aberrations arise due to electrons with different wavelengths being focussed onto different focal planes which causes blurring of the image (Orlova and Saibil, 2011). Astigmatism in electron lenses is caused by an asymmetric magnetic field in the lens which manifests as an oval shaped beam. The astigmatism is usually corrected prior to imaging by changing the microscope setup parameters that can compensate for the astigmatism using stigmator coils.

1.7.1.4 Detectors

The detector lies at the bottom of the EM column. The development of direct electron detectors (DEDs) has contributed greatly to improving the resolution that is now attainable by cryo-EM (Kühlbrandt, 2014). These DEDs improved the quality of the data in two major ways. Firstly, the first readout time enables collection of multiple shorter exposures. This allows for correction of beam induced movement, as the collected frames can be computationally aligned and averaged (Nogales and Scheres, 2015). Ultimately, improving the signal-to-noise ratio and reducing blur. Secondly, the DEDs have a better quantum efficiency which enables lower cumulative electron doses to be used. Thus, reducing radiation damage.

The maximum theoretically achievable resolution is determined by the sampling rate of the detector which translates to the magnification used for the data collection and the resulting pixel size. An optimal resolution would describe a situation where the number of particles per field-of-view is good, and the achievable resolution is high. While higher magnifications mean the images have a finer pixel size, the number of particles per field-of-view is compromised. On the other hand, lower magnifications produce images with a larger field-of-view, but the pixel size is coarser which limits the resolution. This limit is known as the Nyquist frequency; whereby the maximum resolution that can be achieved is twice the pixel size of the images at a given magnification (Feathers et al., 2021). This is because highest frequency wave that is able to be sampled, completes one full oscillation over two pixels.

1.7.2 Contrast and The Contrast Transfer Function

In biological TEM, there are two major types of contrast: amplitude and phase contrast (Orlova and Saibil, 2011). Amplitude contrast occurs because specimen atoms which have high atomic numbers, such as those stained with heavy metals, will absorb electrons or scatter them at high angles. Therefore, these electrons do not contribute to the final image, and consequently, these regions appear darker in the final image (Franken et al., 2020). Hence, amplitude contrast is particularly observed in stained biological material. Amplitude contrast can also be increased by the removal of inelastically scattered electrons by the use of objective lens aperture and energy filter.

As biological specimens consist of light atoms (C,H,O,N), they do not absorb electrons from the incident beam (Orlova and Saibil, 2011). Therefore, very little amplitude contrast is produced. Instead, contrast is provided by phase contrast. This is possible because the specimen directly scatters the electrons at varying

angles. The electrons scattered at different spatial frequencies experience different phase shifts relative to the incident beam. As the electrons are recombined on the image plane, their interference pattern governs the final image. At focus, due to the weak scattering of electrons by the sample, there is no contrast (Saibil, 2000). To improve the phase contrast, micrographs are collected under-focus or at defocus. While this results in greater contrast, this limits high resolution information. Subsequently, during cryo-EM data collection, micrographs are collected at a range of defocus values, with some values closer to focus.

Phase contrast is also provided by the spherical aberration of the objective lens (Franken et al., 2020). Electrons which approach the lens away from the optical axis are bent more strongly. This adds another phase shift to the scattered electron waves. Therefore, phase contrast is provided by the combination of defocus and spherical aberration (Saibil, 2000). Taking this into account, the modulation of a recombined image can be described by the contrast transfer function (CTF). The CTF is an oscillating sine function which represents the contrast transfer as a function of spatial frequency. At some spatial frequencies, the contrast is zero representing loss of information. This is displayed as black Thon rings on the Fourier Transform of the micrograph (Figure 1.13A). The rings oscillate between positive and negative contrast (Figure 1.13B). To circumvent the loss of information at zero, data is collected over a range of defocus values. Each defocus value has its own zero crossing, and therefore, missing information can be obtained by averaging multiple micrographs with different defocus values (Figure 1.13C) (Franken et al., 2020). Another characteristic of CTF is that it is a diminishing function, owing to chromatic aberrations of the EM lenses. Furthermore, as a consequence of temporal and spatial incoherence by the electron beam, there is a dampening of the CTF at high spatial frequency.

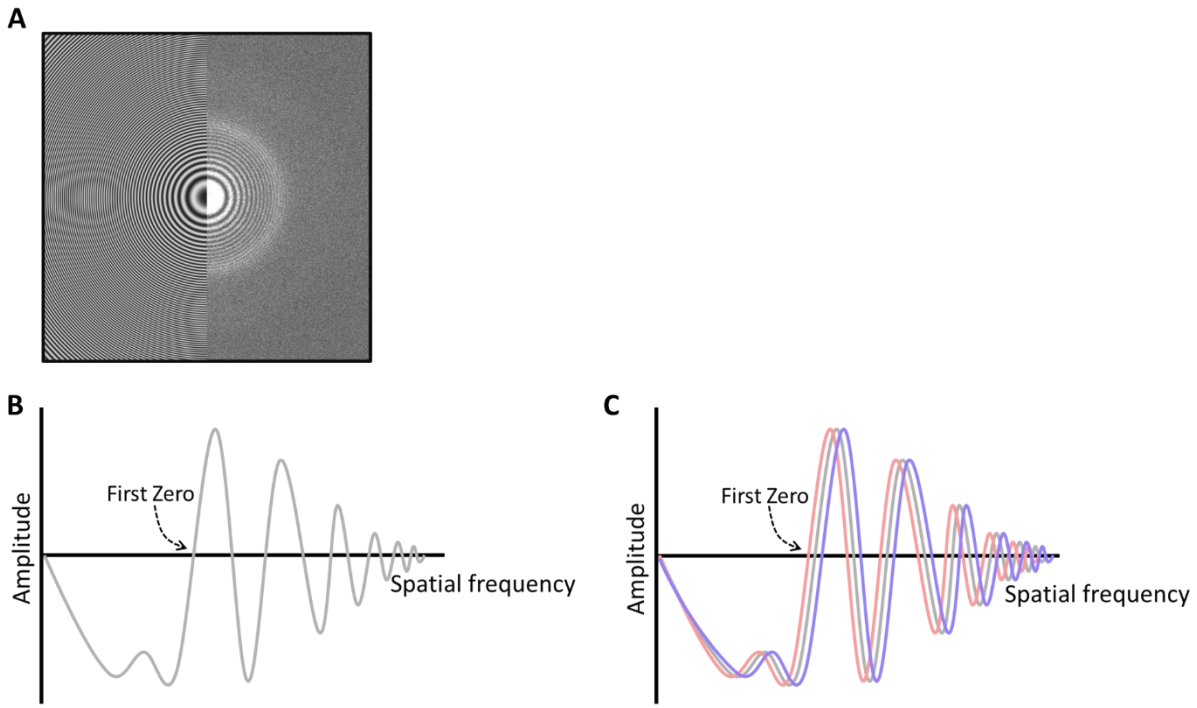

Figure 1.13: The Contrast Transfer Function (CTF). (A) A representative power spectrum showing Thon rings. The oscillations of the CTF can be seen represented by the regions of high amplitude (white) and the zeros (black). **(B)** A 2D example of CTF as a function of spatial frequency. **(C)** To recover loss of information at the zero crossings, imaging is performed at ranging defocus values. The CTF in pink represents an image at high defocus, the CTF in purple describes an image at low defocus (closest to focus) and the CTF in grey represents a defocus value between the two.

1.7.3 Cryo-EM sample preparation

A key step in the workflow for structural determination by SPA cryo-EM is the sample and grid preparation. In general, the vitrification process is carried out via the following steps: the sample is applied to the grid, filter paper is used to blot away excess solution in order to create a thin aqueous layer, and then the sample is plunge frozen in liquid ethane. The plunge freezing is done rapidly to avoid formation of crystalline ice, and thereby, preserving the sample in amorphous ice (Dubochet et al., 1988).

Cryo-EM grids are generally composed of copper or gold, as these materials are able to conduct away extra charge that may deposit during imaging and are typically coated with perforated carbon film with holes in which the sample is frozen. As the grids tend to be hydrophobic, in order to ensure even distribution of the sample across the grid, the grids are glow-discharged. Glow discharging involves

ionising a chamber filled with gas, and the glow discharge in the air adds a negative charge onto the carbon layer of the grid.

As different protein samples will have different optimal freezing conditions, the vitrification process often needs optimising. The ice thickness needs to be thick enough that it encompasses the proteins, but thin enough that it doesn't cause excessive beam interference. For example, larger protein complexes will tend to require thicker ice than smaller complexes. The ice thickness can be optimised by testing different freezing conditions such as blot time and blot force.

1.7.4 Overview of cryo-EM single particle analysis data processing

1.7.4.1 Pre-processing

Once the vitrification of the sample in ice has been optimised, a large dataset is collected at high magnification in the form of movies, if the microscope was equipped with a DED. The multiple frames within the movie are first aligned and averaged through a process called motion correction. Motion correction works by tracking the movement of the particles through the movie frames of each exposure, aligning the frames with one another and then averaging them to obtain a single micrograph. This corrects for beam induced motion and subsequently minimises blur (Nogales and Scheres, 2015).

Following on from motion correction, the next step of pre-processing involves estimation of the defocus by fitting the CTF to the Thon rings of each micrograph. This not only enables computational correction of the CTF but also provides the opportunity to remove micrographs of poor quality from the dataset.

1.7.4.2 Particle picking, extraction and 2D classifications

From the pre-processed micrographs, individual particles need to be picked and extracted to build a 3D structure. During particle picking, the exact coordinates of individual particles within a micrograph are found and this can be done manually or in automated approaches, with the latter commonly favoured due to the large number of micrographs present in cryo-EM datasets. Once picked, the particles are extracted from the micrographs to create a particle stack.

Owing to the low signal-to-noise ratio which hinders proper assessment of the particles, the particle stack is subjected to 2D classification (Sigworth, 2016). Here, the particles are aligned relative to one-another

and then grouped depending on apparent similarities. As such, each 2D class is an average of many similar particles which, therefore, improves the overall signal. These 2D class averages can provide a great deal of information regarding the sample, including basic features such as dimensions, different conformations, and sometimes even secondary structure elements. They can also give an impression on structural heterogeneity or existence of orientation bias.

2D classification also provides a way to curate the particle stack. Bad quality particles or 'junk', consisting of denatured or aggregated particles, carbon edges or background noise, are sorted into a separate group from the better quality particles. The particles contained within the bad 2D classes can be removed from the particle stack. Several rounds of 2D classification are performed to refine the particle stack.

1.7.4.3 Generating 3D reconstructions from 2D images

In TEM, micrographs are 2D projections of 3D objects. A key concept by which structures are solved in EM is the "projection-slice theorem". This theorem describes that in reciprocal space, the Fourier transform of each of these 2D projections relates to a different slice through the Fourier transform of the 3D object. These projections overlap through a "common line" (Figure 1.14) (Van Heel, 1987). The 3D object can therefore be reconstructed by using the "common line" to deduce orientations of the projections relative to each other.

In order to determine the relative orientations between particles, a method known as projection mapping is often employed by cryo-EM softwares (Nogales and Scheres, 2015, Penczek, 2010). Specifically, projections of a 3D reference structure in multiple orientations are generated, and then the individual particles are compared and aligned to these reference projections (Figure 1.15A, B). Thus, giving an insight into the orientation of the particles. The quality of the reference structure can be improved by back calculating it. Furthermore, alignment of 2D projections with the reference is performed in an iterative manner which improves the alignment accuracy and should mean that it converges with the true structure of the particle being investigated. The 3D reference for the refinement can either be imported using a similar structure previously solved or can be determined *de novo* from the dataset using *ab initio* methods (Nogales and Scheres, 2015). After the orientations of the particles in the dataset are aligned, each particle is back projected into the 3D reconstruction.

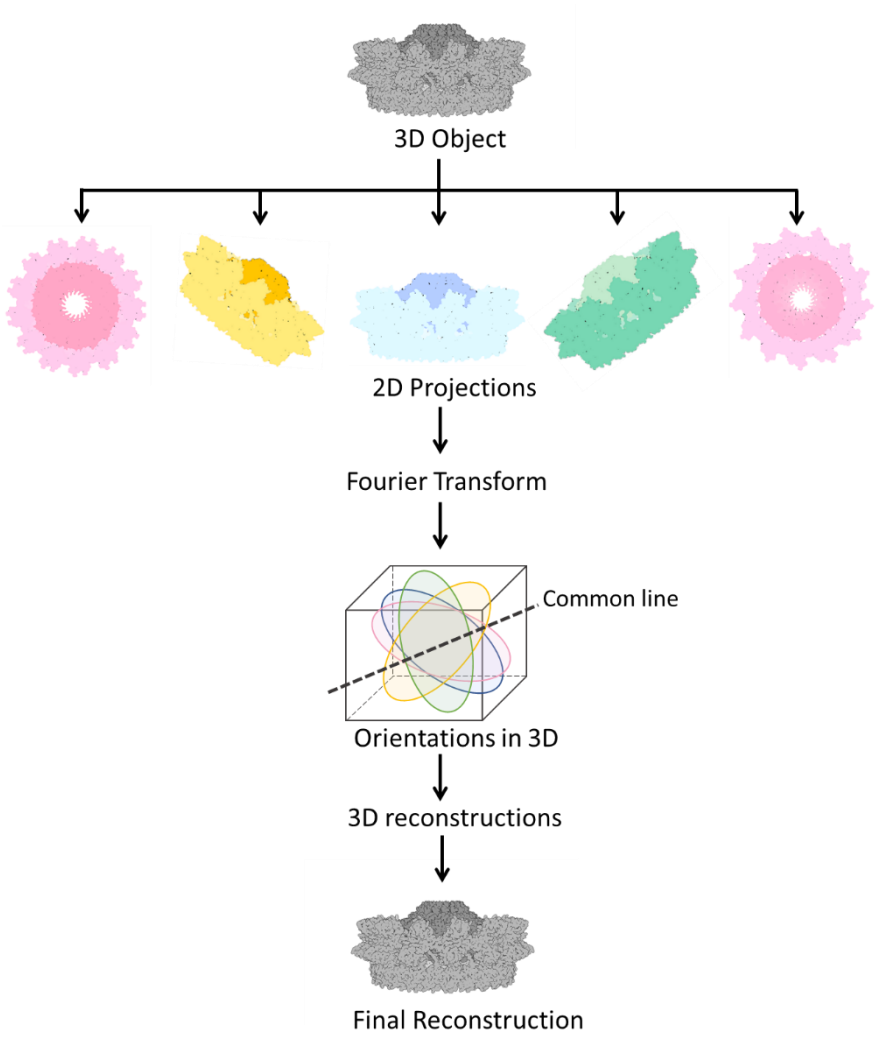

Figure 1.14: Schematic illustrating the basics of how the structure of a 3D object is solved.

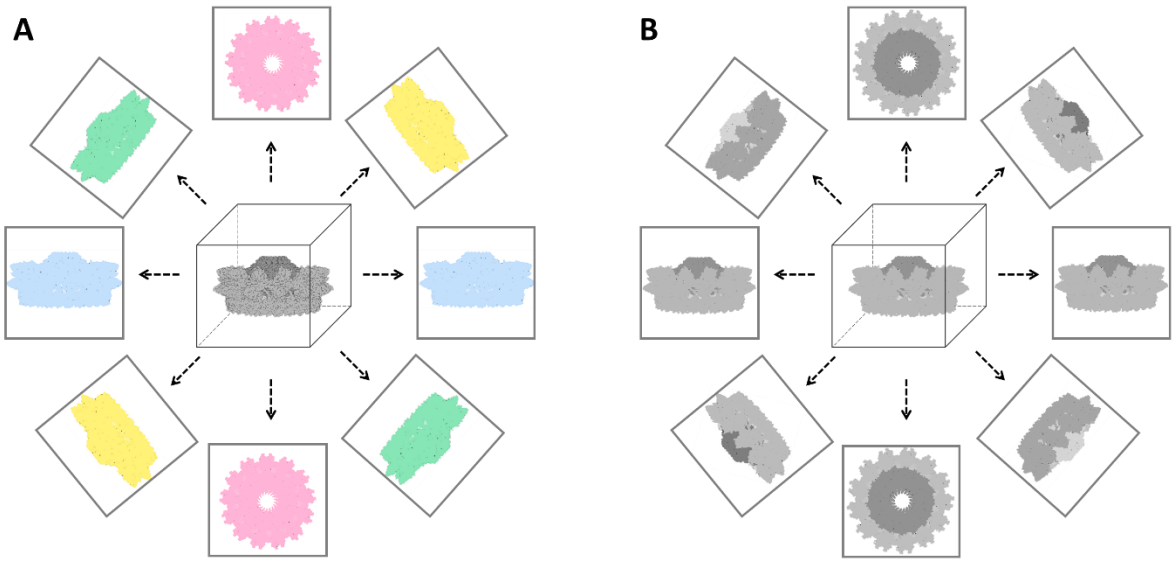

Figure 1.15: Schematic illustrating projection mapping during 3D reconstruction. (A) A 3D reference model is used to generate reference 2D projections in different orientations. **(B)** The experimental particles are compared to the reference 2D projections to determine their orientations.

1.7.4.4 Post-processing, structure refinement and validation

Having obtained a 3D reconstruction, this cryo-EM map can be further refined. This involves restoring contrast on the map which occurs due to loss of contrast at higher resolution and amplitude decay which is resolution dependent. This resolution dependence can be described by the B-factor (Rosenthal and Henderson, 2003). This value can be estimated from the Fourier representation of the reconstructed volume. In a process termed sharpening, the inverse B-factor can be applied to the map to improve contrast and amplify high resolution features, usually along with a weighting factor (figure-of-merit) to avoid enhancing noise (Rosenthal and Henderson, 2003). Once the model is built, to achieve an atomic structure that matches the experimental map to the greatest degree, further refinement can be performed (Beckers et al., 2021). This is done so using an energy function that boosts the model-to-map correlation while also maintaining the realistic features of the map which includes bond length and angles.

The validation of EM maps is important to avoid model bias and the over-fitting of data. This is assessed by dividing the dataset randomly into two, and they are refined independently (Beckers et al., 2021). The resulting 3D reconstructions are then compared. In the “gold standard” approach, the Fourier Shell

Correlation (FSC) between the two half maps is calculated for varying spatial frequencies. The resolution of a structure is classically reported as the number at the FSC value of 0.143.

1.7.5 The suitability of single particle cryo-EM to solving the structure of the F-OMCC

In the recent years, cryo-EM has undergone spectacular progress and the significant advancement in electron detectors and imaging software has led to a 'resolution revolution'(Kühlbrandt, 2014). The technique allows for structural determination of an extremely vast range of protein sizes; ranging from below 100 kDa proteins, as exemplified by the 3.2 Å structure of the 52 kDa streptavidin protein, to entire protein machineries as highlighted by the structure of the R388 encoded T4SS at near atomic resolution (Fan et al., 2019, Mace et al., 2022).

This technique is particularly advantageous for studying membrane proteins compared to the other leading structural determination techniques such as nuclear magnetic resonance (NMR) and X-ray crystallography. Firstly, ordered crystals are not required which is often hard to produce with membrane proteins and especially biocomplexes, such as the F-OMCC, which are composed of several proteins and therefore, are not readily amenable to crystallisation. Also, an important advantage is that in cryo-EM the protein sample is imaged in vitreous ice whereby they are preserved in a hydrated, near-native state in solution. This removes the requirement for tightly packed crystals and also often means that large protein concentrations are not needed which can be hard to achieve with membrane proteins.

The success of cryo-EM in deciphering the structure of complexes is demonstrated by the various structures of the outer membrane core complex that have been solved from the T4SS alone. This includes the OMCC belonging to pKM101, R388, the Cag machinery from *H. pylori* and the Dot/Icm apparatus from *L. pneumophila* (Low et al., 2014, Chaptal et al., 2017, Sheedlo et al., 2021, Chandran et al., 2009, Fronzes et al., 2009, Durie et al., 2020, Sheedlo et al., 2020, Chung et al., 2019, Sgro et al., 2018). While the *in situ* analyses of these T4SSs by cryo-ET have been invaluable in providing the overall architecture, the high resolution details have only emerged from single particle cryo-EM.

1.8 Aims of this project

The overall aim of this project is to investigate the biochemistry and structure of protein complexes from the F-T4SS. Specifically, this will involve:

- Solving the structure of the F-OMCC which will show the architecture and symmetry of the complex and will provide insight into how the F-OMCC fits in with the rest of the F-T4SS, especially in terms of the pilus. Steps towards this will firstly require optimisation of purification since it is a membrane embedded complex (Chapter 3), followed by structural determination by single particle cryo-EM (Chapter 4).
- Exploring possible interactions between the F-OMCC and other proteins encoded on the *tra* region to broaden understanding of how the T4SS functions using pull-down assays (Chapter 5)
- Investigation into the structure of TraG and whether it interacts with TraN using purification methods and pull-down assays (Chapter 6).

Chapter 2: Materials and methods

2.1 Chemicals and Bacterial strains

2.1.1 Water

Unless stated, all the reagents and solutions used in this study were prepared in ultrapure deionised MiliQ water.

2.1.2 Growth media and media additives

Lysogeny broth (LB) (Miller) solution and LB-agar (Miller) plates was prepared from powdered stocks from Thermo Scientific and Millipore respectively. The solutions were made as per manufacturer's instructions and were autoclaved prior to use. The media supplements have been listed in Table 2.1, which were dissolved in the solvent listed, filter sterilised using a 0.2 µm filter and stored at either 4 °C for short term use or at -20 °C for long-term storage.

Table 2.1: The media additives that were used in this study

	Media additive	Solvent	Concentration of stock	Concentration of working
	Ampicillin	ddH_2O	100 mg ml^{-1}	100 µg ml^{-1}
Antibiotic	Carbenicillin	ddH_2O	100 mg ml^{-1}	100 µg ml^{-1}
	Spectinomycin	ddH_2O	100 mg ml^{-1}	100 µg ml^{-1}
	L-Arabinose	ddH_2O	20 % (w/v)	0.2 % (w/v)
Inducer	Anhydrotetracycline	Ethanol	2 mg ml^{-1}	200 µg l^{-1}
	Isopropyl ß-D-1-thiogalactopyranoside (IPTG)	ddH_2O	1 M	2 mM
Other	X-Gal	DMSO	100 mg ml^{-1}	100 µg ml^{-1}

2.1.3 Bacteria strains and plasmids

The bacterial strains and plasmids used in this study are provided in Table 2.2 and Table 2.3 respectively with the relevant properties and source listed.

Table 2.2: The *E. coli* strains used in this study

E. coli strain	Properties	Source/ Reference
JE2571	Harbours the F-like plasmid pED208 encoding the F-T4SS	(Bradley, 1980b)
Top10	Strain used for cloning and protein expression. *F– mcrA Δ(mrr-hsdRMS-mcrBC) φ80dlacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)*	ThermoFisher
BL21 (DE3)	A T7 expression strain. *F– ompT gal dcm lon hsdSB(rB–mB–) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS)*	Filloux Lab

Table 2.3: The plasmids that was used in this study

Plasmid	Properties	Source/ Reference
pED208	Wild type conjugative plasmid which is a transfer derepressed mutant of the IncFV plasmid	(Lu et al., 2002)
pBAD24	A high-copy cloning vector which is ampicillin resistant (Ap^R) and provides controllable protein expression by altering arabinose concentration	Life Technologies (Guzman et al., 1995)
pASK3C	A high copy cloning vector, spectinomycin resistant (Sm^R), anhydrotetracycline inducible *tet* promoter	Life Technologies
pTAMAHISTEV	Expression vector, Ap^R, IPTG inducible T7 promoter, allows N-terminal tagging of proteins with hexahistidine tag and includes TamA signal peptide	(Moynie et al., 2019)

pBAD_TraK_TraB^{Strep}_TraV	The *traK* – *traV* fragment of pED208 cloned into pBAD such that it includes a C-terminal Strep-II tag on *traB*	Laboratory collection
pBAD_TraG^{Strep}	The *traG* gene of pED208 cloned into pBAD with a C-terminal Strep-II tag	This study
pBAD_TraK_TraB^{Strep}_TraV_TraH^{His}	The *traK* – *traV* fragment of pED208 and the *traH* gene cloned into pBAD such that it includes a C-terminal Strep-II tag on *traB* and C-terminal hexahistidine tag on *traH*	Postgraduate student Hayley Jackson for this study
pBAD_TraK_TraB^{Strep}_TraV_TraF^{His}	The *traK* – *traV* fragment of pED208 and the *traH* gene cloned into pBAD such that it includes a C-terminal Strep-II tag on *traB* and C-terminal hexahistidine tag on *traF*	Postgraduate student Hayley Jackson for this study
pASK_TraW^{His}	The *traW* gene of pED208 cloned into pASK with C-terminal hexahistidine tag	This study
pASK_TraW^{His}_TrbC	The *traW* and *trbC* gene of pED208 cloned into pASK with C-terminal hexahistidine tag on *traW*	This study
pASK_TrbC^{His}	The *trbC* gene of pED208 cloned into pASK with C-terminal hexahistidine tag	This study
pASK_TrbI-TraG_TraN^{His}	The *trbI* – *traG* fragment of pED208 cloned into pASK with C-terminal 10 x histidine tag on *traN*	This study
pBAD_TrbI-TraG_TraH^{His}	The *trbI* – *traG* fragment of pED208 cloned into pBAD with C-terminal 10 x His-tag on *traH*	Postgraduate student Ambre Bexter for this study
pBAD_TrbI-TraG_TraH^{Strep}	The *trbI* – *traG* fragment of pED208 cloned into pBAD with C-terminal Strep-II tag on *traH*	Postgraduate student Ambre Bexter for this study

pTAMAHISTEV_TraN	The *traN* fragment of pED208 cloned into pTAMAHISTEV	This study

Schematic of the operon structure of the following plasmids pBAD_TraK_TraB^{Strep}_TraV, pBAD_TraG^{Strep}, pBAD_TraK_TraB^{Strep}_TraV_TraH^{His}, pASK_TraW^{His}, pASK_TraW^{His}_TrbC, pASK_TrbC^{His}, pASK_TrbI-TraG_TraN^{His}, pTAMAHISTEV_TraN are shown in Figure 3.1B, 6.1C and Appendix figures A.3C, A.5C and A.6C respectively.

2.1.4 Preparation of chemically competent *E. coli*

To prepare chemically competent *E. coli* strains for the transformation, an overnight culture of the appropriate cells were sub-cultured by adding 1 ml to 100 ml LB. The cells were left to grow till OD_{600} of 0.5 – 0.6 at 37 °C. The culture was chilled immediately on ice for 30 min and then centrifuged at 4500 rpm, 10 min at 4 °C. The cells were resuspended in 5 ml of pre-chilled 100 mM $MgCl_2$ and then incubated on ice for 90 min. The cells were pelleted by centrifugation at 4500 rpm, 4 °C for 10 min. The pellet was resuspended in 1.6 ml of pre-chilled 100 mM $CaCl_2$ containing 15% (w/v) glycerol and then aliquoted into 50 µl samples and subsequently flash frozen in liquid nitrogen. The chemically competent cells were stored at -80 °C.

2.2 Molecular Biology techniques

2.2.1 Isolation of plasmid DNA

Plasmids were isolated using the QIAprep® Spin Miniprep Kit (Qiagen) and as per manufacturer's instructions. The DNA was eluted in autoclaved double distilled water (ddH_2O) and the concentration was measured using a Nanodrop™ spectrophotometer.

2.2.2 Transformation of chemically competent *E. coli*

For transformation of competent TOP10 or BL21 (DE3) cells with plasmid, 50 – 100 ng of plasmid DNA was added to 50 µl of competent cells on ice. The cells were incubated on ice for 30 min, and then heat-shocked at 42 °C for 45 sec and followed by a 5 min recovery on ice. Then, 700 µl LB was added to the

mixture and left at 37 °C with gentle agitation. The cells were plated onto agar plates supplemented with the appropriate antibiotics and left to grow at 37 °C overnight.

2.2.3 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate DNA fragments based on their size. The gels were prepared by dissolving agarose in boiling 1X TAE (Tris, acetate, EDTA) buffer to obtain a 0.8 % (w/v) gel. For DNA visualisation, SYBR™ Safe DNA gel stain (ThermoFisher) was added from a 10, 000 X stock.

The samples were prepared for gel electrophoresis by adding 6X loading dye (0.25% bromophenol blue, 0.25% xylene cyanol FF, 30% glycerol). The agarose gel was submerged in tanks filled with 1X TAE. To allow determination of the size and concentrations of the DNA bands, GeneRuler™ 1kb DNA ladder (ThermoFisher) was used as the molecular marker. The gels were allowed to run for 50 min at 100 V. DNA was visualised using a BioRad Gel Doc 2000 transilluminator. DNA fragments were excised and purified from the agarose gel using QIAquick® Gel Extraction Kit (Qiagen) and eluted in autoclaved ddH_2O.

2.2.4 Polymerase chain reaction (PCR)

DNA fragments were amplified by polymerase chain reaction (PCR) using CloneAmp HiFi Polymerase (Takara). The typical thermocycling conditions are listed in Table 2.4. The annealing temperature was determined as the lowest melting temperature of the two primers, minus 5 °C. The reaction mix was set-up according to the manufacturer's instructions and in 1X CloneAmp HiFi PCR premix and using the primers detailed in Table 2.5.

Table 2.4: The thermocycling conditions for the PCR reaction

Thermocycling Step	Temperature (°C)	Time (sec)	Cycles
Initial denaturation	98	240	1
Denaturation	98	10	30
Annealing	Tm - 5°C	10	
Extension	72	10/ kb	
Final extension	72	240	1

2.2.5 Oligonucleotide primers

The primers that were used in this study are listed in Table 2.5. Newly obtained lyophilized primers were resuspended in ddH_2O to achieve a final concentration of 100 µM. The primers were stored at -20 °C.

Table 2.5: The primers used for the PCR reaction

Construct	Primer name	Primer Sequence (5' – 3')
All constructs using pASK as vector	pASK_fw	GCAGCGCATCACCATCACCATCACTGAGCTTGACCTGTGAAGTGAAAAATGG
	pASK_rev	TTGCCCTCGTTATCTAGATTTTTGTCGAAC
pASK_TraW^{His}	TraW_His_fw	AGATAACGAGGGCAAATGAAAAAAGGGCTGCTGACACT
	TraW_His_rev	TCAGTGATGGTGATGGTGATGCGCTGCTGGCTTCACCTCCGGCG
pASK_TrbC^{His}	TrbC_His_fw	AGATAACGAGGGCAAATGAAACGCGTACTGATATCGCTG
	TrbC_His_rev	TCAGTGATGGTGATGGTGATGCGCTGCTTGCACCTCCCCACGCAG
pASK_TraW^{His}_TrbC	TraW_His_TrbC_fw	GCAGCGCATCACCATCACCATCACTGAGTCACAGGGAGAAAAACGATGAAAC
	TraW_His_TrbC_rev	CTTCACAGGTCAAGCTCATTGCACCTCCCCACG
	TraW_His_TrbC_vector_fw	GCTTGACCTGTGAAGTGAAAAATGGCGCACATTG
pBAD_TraK_TraB^{Strep}_TraV_TraH^{His}/TraF^{His}	pBAD_TraH_F_FOMC_fw	GCAGCGCATCACCATCACCATCACTGAGTACCCGGGGATCCTCTA

	pBAD_TraH_F_FOMC_rev	TTAATTTACCCTGGCCGGAAGAAC
	TraH_FOMC_fw	TTCCGGCCAGGGTAAATTAACGCTCAGGGACAACAATAATG
	TraH_FOMC_rev	TCAGTGATGGTGATGGTGATGCGCTGCATTCACCCTGCTGAACTGGTA
	TraF_FOMC_fw	TTCCGGCCAGGGTAAATTAAATGCAGGAGAACGCGC
	TraF_FOMC_rev	TCAGTGATGGTGATGGTGATGCGCTGCGTAGTCAGGTGCAAAGTGGGTAAG
pBAD_TrbI-TraG_TraH^{Strep}	TraH_streptag_fw	TGGAGCCACCCGCAGTTCGAAAAGTGATATGGCCGTTGATACGATTTACAC
	TrbI_reverse	TTGGTCAGTGGTCATGGTGAATTC
	TrbI_forward	ATGACCACTGACCAAAAACTTACCCCTG
	TraH_streptag_rev	TCACTTTTCGAACTGCGGGTGGCTCCAACCGCCGCTGCCAGAA
pTAMAHISTEV_TraN	pTAMAHISTEV_fw	GGATCCGAATTCGAGCTCCG
	pTAMAHISTEV_rev	CATGGCGCCCTGAAAATACAG
	TraN_fw	TATTTTCAGGGCGCCATGAATGACCAGTTTAATCAGGGCATG
	TraN_rev	GAGCTCGAATTCGGATCCTCATGGCGCGTTCTCCTG
pBAD_TraG^{Strep}	pBAD_G_fw	GCAGCGGGTTCTGGCAGCGGCGGTTGGAGCCACCCGCAGTTCGAAAAGTGAGTACCCGGGGATCCTCTAG
	Pbad_G_rev	GGTGAATTCCTCCTGCTAGC
	pTraG_Strep_fw	TAGCAGGAGGAATTCACCATGGCCGTTGATACGATTTACAC
	pTraG_Strep_rev	TCACTTTTCGAACTGCGGGTGGCTCCAACCGCCGCTGCCAGAACCCGCTGCCACCATTATCACCGATCCAATCAT

pASK_TrbI-TraG_TraN^{His}	TrbI_fw	AGATAACGAGGGCAAATGACCACTGACCAAAA ACTTACCCCTGC
	TraN_his_rev	TCAGTGGTGATGATGGTGGTGATGGTGATGAT GACCGCCGCTGCCAGAACCTGGCGCGTTCTCCT GCATT
	TraN_his_fw	GGTTCTGGCAGCGGCGGTCATCATCACCATCAC CACCATCATCACCACTGAGCCGGAGGTGAAGCC ATGAGATGTATTGCCGGTGC
	TraG_rev	CTTCACAGGTCAAGCCTACACCATTATCACCGAT CCAATCA

2.2.6 In-Fusion cloning

All the constructs in this study were made using the In-Fusion cloning technique (Takara). This relies on the homologous recombination of PCR generated overlapping complementary regions, which are at least 15 base pairs (bp) in length, at the ends of the insert and vector. The primers were designed such that they would include a complementary region at the 5' end and also linearise the vector. To fuse the vector and insert, the fragments were combined at a 1:2 molar ratio (vector: insert). The reaction mix was bought up to 8 µl if needed using ddH_2O and 2 µl of 10X In-Fusion Reaction Mix was added. The reaction was allowed to proceed at 50 °C for 15 min. The mixture was transformed into chemically competent TOP10 cells and plated onto agar plates supplemented with the appropriate antibiotics.

2.2.7 Checking for insert and sequencing

Following In-Fusion, colonies were grown overnight, and the plasmid was isolated. To probe if it contained the expected insert, PCR was performed. Plasmids which appeared to be positive, were sequenced using GATC at Eurofins Supreme Sanger Sequencing using the primers listed in Table 2.6.

Table 2.6: The primers provided for sequencing

Primer name	Primer Sequence (5' – 3')	Targets
pBAD_ara	GACGCTTTTTATCGCAACTC	araBAD promoter
S1	TACCACTCCCTATCAGTGATAGAG	Reads into Tet promoter on pASK vector

S4	TCTGATGGGCGGAGGAAGAATGAC	*traK*
S5	AGCTGCGTCTCACTGGCAAAGTTC	*traB*
S8	ACGTCCGCAGGTGTAAACACAAAC	*traV*
S18	GAAGCGCAGCACTGCAAGAAGATG	End of *traN*
S19	CTGGACACCGCCATTCTGTATC	Middle of *traF*
S22	CCCATTCTCTCATGGATTGTG	End of *traH*
S25	GTGAATATGATACGCAACAGACAC	End of *traG*

2.3 Protein Biochemistry

2.3.1 Expression test

Prior to carrying out purification trial, for new constructs, an expression test was carried out to assess if the tagged protein is being expressed. Overnight cultures from transformants colonies were used to inoculate 50 ml LB such that a starting OD_{600} of 0.1 is achieved. The cells were left to grow at 37 °C at 180 rpm. When the cells reached OD_{600} of 0.6, a 1 ml 'before induction' sample was taken and the expression was induced with anhydrotetracycline or L-arabinose or IPTG for pASK, pBAD and pTAMAHISTEV respectively. The cultures were left to grow overnight at 18 °C at 180 rpm. A 1 ml 'after induction' sample was taken. Both the before and after cell samples were pelleted at 7000 x g for 5 min and then resuspended into 100 µl fresh LB for analysis.

2.3.2 Protein analysis by SDS-PAGE

Protein samples were prepared for sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) by adding 2X Laemmli sample buffer (ThermoFisher Scientific) and heating at 100 °C for 5 minutes. The 10-20% SDS-PAGE gel (ThermoFisher Scientific) was submerged in the tank filled with 1X Tris-glycine buffer running buffer made from 10X stock (144g Glycine, 30.5g Tris, 10g SDS to make 1L). To allow visualisation of protein molecular weight, 5 µl Colour Prestained Protein Standard, Broad range (NEB) was run alongside the sample as a molecular marker. The gels were allowed to run for 50 min at 180 V. Protein was detected using ReadyBlue™ protein gel stain (Merck).

2.3.3 Western blot

The presence of Strep- or His-tagged proteins in the cell lysates or purification samples was verified by performing a Western blot. The antibodies used in this study have been listed in Table 2.7. Proteins were transferred from the SDS-PAGE gel to nitrocellulose membranes at 0.4 A, 250 V, 50 W for 1 hr. The membranes were blocked for an hour in blocking buffer [5% (w/v) milk powder (Sigma) in 1 x TBS-T (50 mM Tris pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween20] and then incubated with the primary antibody in 1 x TBS-T supplemented with 1% (w/v) milk powder for 1 hr at room temperature or overnight at 4 °C. The membranes were then washed four times for 5 mins each with 1 x TBS-T, and then incubated with the anti-mouse secondary antibody for 1 hr in 1 x TBS-T supplemented with 1% (w/v) milk powder. The membranes were washed again four times for 5 mins each with 1 x TBS-T. For immunodetection, the membrane was incubated with the chemiluminescent substrate SuperSignal™ West Pico substrate (ThermoFisher Scientific) for 5 min prior to imaging with BioRad ChemiDoc XRS+.

Table 2.7: The antibodies used for Western blotting in this study

Antibody	Host	Serum	Dilution	Source
anti-StrepII	Mouse	Monoclonal	1:10000	Sigma Aldrich
anti-His_6	Mouse	Monoclonal	1:5000	Sigma Aldrich
anti-mouse	Rabbit	Polyclonal	1:5000	Sigma Aldrich

2.3.4 Purification of the F-OMCC

TOP10 cells were transformed with pBAD_TraK_$TraB^{Strep}$_TraV construct. Protein expression was induced with 0.2% (w/v) L-arabinose when the bacterial culture reached OD_{600} = 0.6, and the culture was left to grow overnight at 18 °C whilst shaking at 180 rpm. The cells were harvested at 7000 x g, 20 mins at 4 °C.

All the steps of purification were carried out at 4 °C. The bacterial cells were resuspended in lysis buffer [50 mM Tris-HCl pH 7.5, 200 mM NaCl, 1 mM EDTA, 0.2 mg ml^{-1} lysozyme, 1 cOmplete protease inhibitor cocktail tablet (Roche)] and lysed by sonication on/off for 2 mins at 60 Amp. The cellular debris was removed by centrifugation at 35000 x g, 20 mins and the resulting supernatant was used to isolate the membrane by ultracentrifugation at 120,000 x g, 60 mins. The F-OMCC was extracted and stabilised from the membrane by mechanical homogenising using the solubilisation buffer [50 mM Tris-HCl pH 7.5, 300 mM NaCl, 1 mM EDTA, 1% (w/v) n-dodecyl-β-D-maltoside (DDM, Anatrace) and 0.8%

Lauryldimethylamine-N-Oxide (LDAO, Anatrace)]. The solution was left solubilising with gentle agitation for 60 mins. The insoluble material was removed by centrifugation at 100,000 x g for 40 mins. The supernatant was loaded onto a 1ml StrepTrap™ HP (StrepTrap) column (GE Healthcare) which had been equilibrated in equilibration buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl and 0.23% w/v LDAO) and eluted in the same buffer supplemented with 10 mM D-desthiobiotin (IBA). Fractions which contained the complex were pooled and concentrated using a centrifugal concentrator with 100K MWCO (Merck Millipore). The sample was then further purified by loading onto a Superose 6 GL 10/300 column (GE Healthcare) equilibrated with equilibration buffer. The proteins were analysed by SDS-PAGE and Coomassie blue staining. The fractions from the shoulder peak on the gel filtration were pooled and concentrated using 100K MWCO (Millipore) to be further analysed by electron microscopy.

2.3.5 Detergent screen for solubilisation of F-OMCC and TraN

To probe which detergent was best at solubilising the F-OMCC from the bacterial membranes, a detergent screen was carried out. The steps to isolate the membranes were the same as above. At the solubilisation stage, different detergents were used at 5 x CMC to test efficiency of extraction, as listed in Table 2.8, in a base buffer composed of 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 1 mM EDTA. The membranes were mechanically homogenised, and the solution was left solubilising with gentle agitation for 60 mins at 4 °C. Before the insoluble material was pelleted, a sample was taken of the 'total protein' or also referred to as 'membrane homogenised'. Following centrifugation at 100,000 x g for 40 mins, a sample was taken of the supernatant termed 'solubilised protein'. A western blot was carried out as described in section 2.3.3. The same screen was used to assess which detergent was best at solubilising TraN.

Table 2.8: The detergents used for the screen

Detergent	CMC [% (w/v)]	5 X CMC [% (w/v)]	Source
Digitonin	0.0021	0.0105	Anatrace
Glycol-Diosgenin (GDN)	0.0021	0.0105	Anatrace
n-Dodecyl β-D-Maltoside (DDM)	0.0087	0.0435	Anatrace
Lauryldimethylamine-N-oxide (LDAO)	0.023	0.115	Anatrace
Cymal-5	0.12	0.6	Anatrace
Chaps	0.49	2.45	Anatrace

Decyl Maltose Neopentyl Glycol (DMNG)	0.0034	0.017	Anatrace
n-dodecylphosphocholine (Choline)	0.047	0.235	Anatrace
n-Octyl-β-D-Glucoside (OBG)	0.53	2.65	MyBioSource

2.3.6 Pulldown assays of TraW, TrbC with F-OMCC

TOP10 cells were co-transformed with the plasmids pASK_TraW^{His}_TrbC and pBAD_TraK_TraB^{Strep}_TraV, and plated onto agar plates containing both carbenicillin and spectinomycin. Protein expression was induced when the bacterial culture reached OD_{600} = 0.6 by addition of 0.2% (w/v) L-arabinose and 200 ug L^{-1} anhydrotetracycline. The culture was left to grow overnight at 18 °C whilst shaking at 180 rpm. The culture was harvested, and membranes were isolated as mentioned in section 2.3.4 but with the absence of EDTA in the buffers. The membranes were solubilised by mechanical homogenisation using the same solubilisation buffer as the F-OMCC purification but using 0.6% (w/v) Cymal-5 instead of DDM and LDAO. The sample was left solubilising for 1 hr at 4 °C, followed by ultracentrifugation at 100,000 x g for 40 mins to pellet insoluble material. The supernatant was either loaded onto a 1 ml HisTrap^{TM} HP column (HisTrap) (Cytiva), a 1 ml StrepTrap column or a double pulldown was performed. Both columns were equilibrated and washed with wash buffer [50 mM Tris-HCl pH 7.5, 300 mM NaCl and 0.12% (w/v) Cymal-5]. For the HisTrap purification, elution was done via a gradient using buffer B [50 mM Tris-HCl pH 7.5, 300 mM NaCl and 0.12% (w/v) Cymal-5, 500 mM imidazole]. For the StrepTrap purification, elution was carried out with the wash buffer supplemented with 10 mM desthiobiotin. The samples were analysed by SDS-PAGE, Coomassie staining and western blot.

For the double pull-down, the sample was loaded onto the 1 ml HisTrap column that was equilibrated and washed with wash buffer. The column was further washed with wash buffer supplemented with 10 % (v/v) buffer B. Before elution, the 1 ml StrepTrap column, which had been equilibrated in wash buffer, was attached directly below the HisTrap column. The sample bound to the HisTrap was eluted in one step with 90% (v/v) buffer B into the attached StrepTrap column. The HisTrap column was then unattached. The Strep column was washed with wash buffer and then eluted with the same buffer supplemented with 10 mM desthiobiotin. The samples were analysed by SDS-PAGE, Coomassie staining and western blot.

2.3.7 Pulldown assay of TraN with F-OMCC

The interaction of TraN with the F-OMCC was investigated using two different plasmids expressing TraN: pTAMAHISTEV_TraN and pASK_TrbI-TraG_TraN^{His}. For pASK_TrbI-TraG_TraN^{His} pull-down with pBAD_TraK_TraB^{Strep}_TraV, TOP10 cells were co-transformed with both plasmids as explained in section 2.3.6. The cells were grown, harvested, membranes generated, solubilised and the pull-down was performed with either a 1 ml HisTrap column or a 1 ml StrepTrap column, as described in section 2.3.6 except with 50 mM HEPES pH 7.5 as opposed to Tris-HCl.

For the pull-down assay of pTAMAHISTEV_TraN with pBAD_TraK_TraB^{Strep}_TraV, the two plasmids were incompatible and could not co-transform the same competent cells with them. Therefore, different compentent cells were transformed with the plasmids, whereby BL21 (DE3) cells were transformed with pTAMAHISTEV_TraN and TOP10 cells were transformed with pBAD_TraK_TraB^{Strep}_TraV. The competent cells were grown separately and F-OMCC expression was induced as mentioned above. TraN expression was induced by addition of 2 mM IPTG when cultures reached OD_{600} = 0.6. The cultures were left to grown overnight at 18 °C whilst shaking at 180 rpm and harvested separately the following day as previously mentioned. The TraN and F-OMCC cell pellets were lysed together using lysis buffer II (50 mM HEPES pH 7.5, 200 mM NaCl, 0.2 mg ml^{-1} lysozyme, protease inhibitor tablet) and by mechanical lysis using a cell disruptor with two rounds of disruption at 25 kpsi. From here onwards, the pull-down protocol followed in the same manner as described in section 2.3.6. The samples were analysed by SDS-PAGE, Coomassie staining and western blot.2.3.8 Purification of TraH and periplasmic protein complex

Purification of TraH along with other members of the periplasmic protein complex was done using the plasmids pBAD_TrbI-TraG_TraH^{His} and pBAD_TrbI-TraG_TraH^{Strep}. Separate TOP10 cells were transformed with each of the plasmids, grown, harvested as mentioned in section 2.3.4.

The membranes for the pBAD_TrbI-TraG_TraH^{Strep} purification were generated as mentioned in section 2.3.4. The membranes were solubilised by mechanical homogenisation using solubilisation buffer II [50 mM HEPES pH 7.5, 200 mM NaCl, 0.6% (w/v) Cymal-5], and the solution was left solubilising for 1 hr at 4 °C. The insoluble material was pelleted by ultracentrifugation at 100,000 x g for 40 mins at 4 °C. The supernatant was loaded onto a 1 ml StrepTrap column which had been equilibrated in wash buffer II [50 mM HEPES pH 7.5, 200 mM NaCl, 0.12% (w/v) Cymal-5]. The column was washed with wash buffer II and then eluted using the same buffer supplemented with 10 mM desthiobiotin. The fractions containing TraH were pooled and concentrated using a 10K MWCO (Millipore) before being injected onto a Superdex 200

10/300 GL column (GE Healthcare) which had been equilibrated with buffer wash buffer II. The samples were analysed by SDS-PAGE, Coomassie staining and western blot.

For the pBAD_TrbI-TraG_TraH^{His} purification, the harvested cell pellet was resuspended in lysis buffer II [50 mM HEPES pH 7.5, 400 mM NaCl, 1 cOmplete protease inhibitor cocktail tablet (Roche), 0.2 mg ml^{-1} lysozyme]. The membranes were generated and solubilised as mentioned in section 2.3.4 but using solubilisation buffer III [50 mM HEPES pH 7.5, 300 mM NaCl, 0.7% (w/v) DDM). The insoluble material was pelleted by ultracentrifugation at 100,000 x g for 40 mins at 4 °C. The supernatant was loaded onto a 5 ml HisTrap column which had been equilibrated in wash buffer III [50 mM HEPES pH 7.5, 300 mM NaCl, 0.03% (w/v) DDM]. The column was washed in the wash buffer III and then further washed with the wash buffer with 10 % (v/v) buffer C [50 mM HEPES pH 7.5, 300 mM NaCl, 0.03% (w/v) DDM, 500 mM Imdiazole]. The protein was eluted in one step with 100% (v/v) buffer C. The protein sample was dialysed overnight at 4 °C in wash buffer III using D-Tube^{TM} Dialyzer Midi, MWCO 6-8 kDa. For further purification, the sample was injected onto a Superdex 200 10/300 GL column (GE Healthcare) which had been equilibrated with buffer wash buffer III. The samples were analysed by SDS-PAGE, Coomassie staining and western blot.

2.3.9 TraG purification

The pBAD_TraG^{Strep} plasmid was used for the purification of TraG. TOP10 cells were transformed with the plasmid and grown, harvested and membranes generated as described in section 2.3.4 but with the exception that the cells were mechanically lysed by cell disruptor with two rounds of disruption at 25 kpsi. The membranes were solubilised in a similar manner but with solubilisation buffer III which contained either 0.7% (w/v) DDM or 1% (w/v) Octyl Glucose Neopentyl Glycol (OGNG)]. The StrepTrap purification was performed as mentioned in section 2.3.4 but with wash buffer III [50 mM HEPES pH 7.5, 300 mM NaCl, 0.03% (w/v) DDM or 0.15% (w/v) OGNG], and the elution with wash buffer III supplemented with 2.5 mM desthiobiotin. The fractions containing TraG were pooled, concentrated with 30K MWCO (Millipore), and then further purified by gel filtration. For the samples solubilised with DDM or DMNG, the sample was injected on a Superdex 200 10/300 GL column equilibrated with wash buffer III. For the samples solubilised with OGNG, the sample was injected on a Superose 6 GL 10/300 column equilibrated with wash buffer III. The fractions were analysed by SDS-PAGE, Coomassie staining and by electron microscopy.

For chemical cross-linking with disuccinimidyl suberate (DSS), a trial was carried out to determine the appropriate DSS concentration to cross-link TraG. The TraG sample following affinity purification as described above using the detergent OGNG, was concentrated to 0.2 mg ml^{-1} using a centrifugal concentration with 30K MWCO (Millipore). For the trial, the following final concentration of DSS (10 mM, Thermo Fisher Scientific) in DMSO was added to TraG: 0.4 mM, 0.8 mM, and 1.6 mM. The samples were left incubating for 1 hr at RT followed by quenching of the reaction upon addition of 50 mM Tris-HCl pH 7.5. The cross-linked TraG was analysed by SDS-PAGE. To assess the different oligomeric states that were present following cross-linking, the purification of TraG was carried out as mentioned above with OGNG, but before the sample was cross-linked before it was injected onto the gel filtration column. DSS was added at a final concentration of 2mM to 0.2 mg ml^{-1} of TraG and the reaction was left to proceed for 1 hr at RT before being quenched. The sample was injected on to a Superose 6 GL 10/300 column equilibrated with wash buffer III with 0.15 % (w/v) OGNG. The fractions were analysed by SDS-PAGE, Coomassie staining and by electron microscopy.

2.3.10 Pulldown assay of TraG with TraN

The plasmids pBAD_TraG^{strep} and pTAMAHISTEV_TraN were used to perform a pulldown assay for TraG and TraN. TOP10 cells were transformed with pBAD_TraG^{strep}, grown, and harvested as described in section 2.3.4. BL21 (DE3) cells were transformed with pTAMAHISTEV_TraN, grown, and harvested as mentioned in section 2.3.7. The cell pellets for TraG and TraN were combined and lysed together as mentioned in section 2.3.7. The cellular debris was pelleted by centrifugation at 35000 x g, 20 mins. The supernatant was used to isolate the membrane by ultracentrifugation at 120,000 x g, 60 mins. The membranes were solubilised in solubilisation buffer III with 0.6% (w/v) Cymal-5 instead of the DDM, and then left to solubilise and further ultracentrifuged as described in section 2.3.4. The supernatant was loaded onto a 1 ml HisTrap column or 1 ml StrepTrap column both equilibrated with wash buffer III except with 0.12% (w/v) Cymal-5 instead of DDM/DMNG/OGNG. Both columns were washed with buffer III containing 0.12% (w/v) Cymal-5. For the HisTrap elution, a one step elution was performed with 100% Buffer C except with 0.12% Cymal-5 instead of DDM. For the StrepTrap elution, wash buffer III with 0.12% (w/v) Cymal-5 supplemented with 2.5 mM desthiobiotin was used.

2.4 Electron microscopy procedures

2.4.1 Electron microscopy sample preparation

2.4.1.1 Negative staining of samples

Negative stain grids were prepared by applying 10 µl of protein sample onto glow discharged Copper 300 mesh grids with a continuous carbon film (Agar Scientific). After 2 mins of incubation, the sample was blotted off and then washed twice with 10 µl of dH_2O. The sample was then stained with 2% (w/v) uranyl acetate. After 1 min incubation, the stain was blotted off and grid was allowed to dry before imaging. The negative stained grids were screened on a FEI Technai12 Spirit transmission electron microscope.

2.4.1.2 Optimising cryo-EM grid type for F-OMCC

To optimise vitrification conditions for F-OMCC, three grids were tested: Ultrathin carbon lacey 300 mesh (Agar Scientific), C-Flat™ (Protochips), Quantifoil R2/2 on 300 mesh copper with 2 nm ultrathin carbon film (Agar Scientific). For freezing, 4 µl of the sample at approximately 0.04 mg ml^{-1} was deposited onto the glow discharged grids and left for a minute before being blotted with blot force 1 and blotting time of 3 secs, before being plunge frozen in liquid ethane using a Vitrobot Mark IV (FEI) operating at 4 °C with 95% humidity. The grids were screened using the FEI Tecnai12 Spirit transmission electron microscope.

2.4.1.3 Cryo-EM grid preparation for the F-OMCC structure

For cryo-EM, 4 µl of the F-OMCC sample at 0.1 mg ml^{-1}, was applied onto glow discharged Ultrathin carbon lacey 300 mesh (Agar Scientific). The sample was left on the grid for 1 min before being blotted with blot time of 2.4 – 2.8 secs. The grids were vitrified by plunge freezing using a Leica automatic plunge freezer EM GP2 operating under 85% relative humidity at 4 °C. The grids were screened to ensure adequate sample concentration and ice thickness on a Jeol 2100 plus electron microscope.

2.4.2 Electron microscopy imaging

2.4.2.1 Negative stain data collection

To generate an initial negative stain model of the F-OMCC, a dataset was collected on the FEI CM200 microscope operating at a voltage of 200 kV and equipped with a TVIPS camera. A total of 200 micrographs were collected at defocus values 2.5 and 3.0 µm.

2.4.2.2 F-OMCC Diamond dataset

A dataset was collected, for the complex purified in DDM and GDN. It was collected at the eBIC facility at Diamond Light Source using the FEI Titan Krios operating at a voltage of 300 kV and equipped with a K2 Summit detector (Gatan) with a nominal pixel of 1.048 Å/pixel. A total of 4406 movies were collected using the EPU software for automated collection. The total exposure time was 9 secs and accumulating a total dose of 57.96 e-/Å2 over 40 frames per image. The nominal defocus range was approximately -1.5 to -4 µm.

2.4.2.3 F-OMCC cryo-EM data collection for high resolution structure

The F-OMCC cryo-EM dataset was collected at LonCEM (London Consortium EM, Francis Crick Institute) using a FEI Titan Krios operating at 300 kV and equipped with Gatan K3 detector. A total of 11588 movies were recorded automatically using the EPU software (FEI) with the detector in counting mode with a nominal pixel side of 1.08 Å. The total exposure time was over 3.9 secs and accumulating a total dose of 50 e-/$Å^2$ over 38 frames per image, corresponding to a dose rate of 15 e-/px/s. The nominal defocus range was approximately −0.7 to −3.0 µm

2.4.3 Data processing for F-OMCC

2.4.3.1 Negative stain model

The images were processed using RELION 3.1 whereby 868 particles were manually picked and used to generate representative 2D class averages (Zivanov et al., 2018). These classes acted as templates for automated particle picking, using which, 40871 particles were auto-picked. Several round of 2D class averaging was performed to filter out bad particles and leaving only the good classes which included a total of 25300 particles. These particles were subjected to 3D classification with C13 symmetry imposed.

2.4.3.2 Initial cryo-EM model from the Diamond dataset

The 4406 movies that were collected were dose-weighted and then aligned using MotionCor2 implemented in Relion 3.141 (Zheng et al., 2017, Zivanov et al., 2018). The contrast transfer function (CTF) values were then calculated using CTFFIND4 built within Relion 3.141 (Rohou and Grigorieff, 2015). Image processing was carried out using Relion 3.141 and cryoSPARC (Punjani et al., 2017, Zivanov et al., 2018). Using Relion 3.141, a total of 240,713 particles were auto-picked. The particles were extracted and exported into cryoSPARC to enable 2D class averaging in both programmes. After several rounds of 2D class averaging to filter out bad particles, a stack of approximately 6440 (Relion 3.141) and 58175 (cryoSPARC) particles were kept. Upon visual inspection, the 2D class averages generated using cryoSPARC appeared more defined and therefore these particles were used to build an *ab initio* model with C13 symmetry imposed. This model was used as the basis for 3D refinement using only seven 2D class averages containing approximately 36300 particles with a lowpass filter of 30 Å. The map was sharpened by applying a B-factor to improve the overall features.

2.4.3.3 High-resolution cryo-EM image processing and reconstruction

All image processing was performed using CryoSparc40 (v3.1.0), unless otherwise stated (Punjani et al., 2017). Patch motion correction was used to sum and align the movie stacks and CTF was estimated using the patch CTF estimation function. Following an initial screen to check for good Thon rings and optimal ice thickness, a total of 10,911 micrographs remained in the dataset. Altogether, 3506 particles were manually picked and then classified to obtain 2D class averages as references for automated particle picking for the entire dataset. A total of 1,445,958 particles were auto-picked and extracted with a box size of 480 x 480 pixels. Following several rounds of 2D classification to filter out bad particles, a stack of 298,235 particles was obtained which showed define $OMCC_{OR}$ features. The $OMCC_{IR}$ features were enhanced after further round of 2D classification which resulted in a set of 74,956 particles. Upon visual inspection, it appeared that two symmetries existed within the complex with C13 for the $OMCC_{OR}$ and 17 for the $OMCC_{IR}$. Ab-initio maps were created and the stack of particles showing $OMCC_{OR}$ features were subjected to 3D homogenous refinement with C13 symmetry applied and a solvent mask of 5 pixels extension and 6 pixels of drop off which covered the $OMCC_{OR}$. In a similar manner, the stack of particles representing the $OMCC_{IR}$ were subjected to 3D homogenous refinement but with C17 symmetry applied and a solvent mask of 5 pixels extension and drop off of 6 pixels covering $OMCC_{IR}$. At the same time, 95,177 particles were used to create an unsymmetrised map which confirmed the presence of C13 and C17 symmetries in the complex. A B factor of -74, -93 and -304 were applied to the $OMCC_{IR}$, $OMCC_{OR}$, and

the OMCC C1 map improving the local resolution to 3.3 Å, 3.4 Å and 5.7 Å as estimated by the gold standard FSC at a 0.143 threshold (Figure 2.1). Relion 3.141 was used to estimate the local resolution for the $OMCC_{IR}$ and $OMCC_{OR}$ maps (Zivanov et al., 2018). The statistics for data collection and processing are included in Table 2.9.

Table 2.9: The statistics for data collection, processing and model refinement

	$OMCC_{OR}$	$OMCC_{IR}$	OMCC C1
Microscope	Titan Krios	Titan Krios	Titan Krios
Detector	K3	K3	K3
Voltage (kV)	300	300	300
Electron exposure (e^-/ Å^2)	50	50	50
Pixel size (Å/pix)	1.08	1.08	1.08
Defocus range (μm)	-0.7 to -3.0	-0.7 to -3.0	-0.7 to -3.0
Initial particle number	1,445,958	1,445,958	1,445,958
Final particle number	298,235	74,966	95,117
Symmetry applied	C13	C17	C1
FSC threshold	0.143	0.143	0.143
Map resolution	3.4	3.3	5.7
Map B factor	-93	-74	-304
EMDB code	EMD-12963	EMD-12962	EMD-13231
Model refinement			
Model vs data CC (mask)	0.72	0.75	N/A
Clash score	13.73	13.82	
MolProbity score	2.18	2.25	
Bonds length rmsd (Å)	0.007	0.005	
Bond angle rmsd (°)	1.136	1.033	
Poor rotamers (%)	0	0	
Ramachandran			
Favoured (%)	90.7	88.03	N/A
Allowed (%)	9.3	11.97	
PDB code	EMD-12963	EMD-12962	

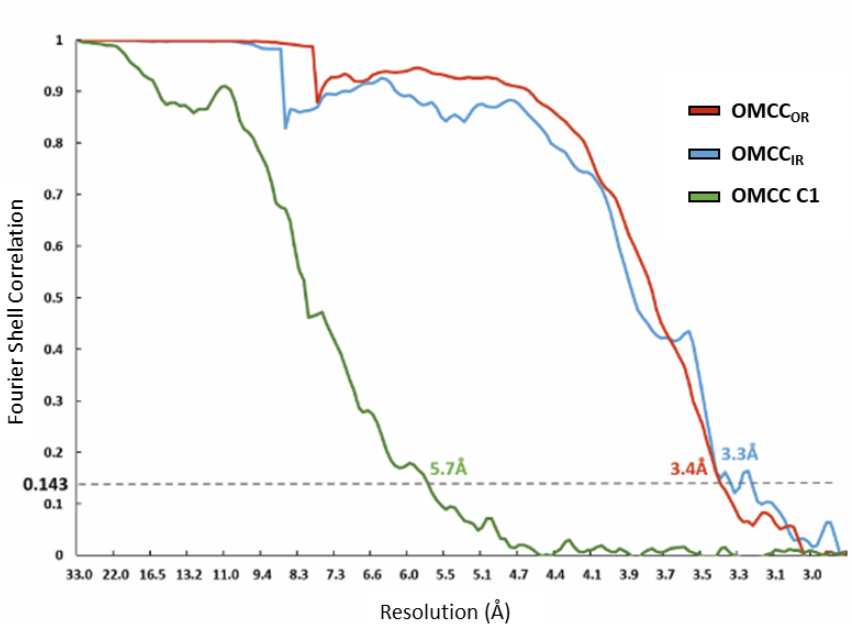

Figure 2.1: The resolution estimation of the F-OMCC maps. The overall resolution of the OMCC$_{OR}$, OMCC$_{IR}$ and, OMCC$_{C1}$ using the Fourier Shell Correlation (FSC) = 0.143 criteria.

2.4.3.4 Model building and refinement

The model building was carried out by collaborator Dr Aravindan Ilangovan at Queen Mary University of London. The model was built using Coot by initially identifying the asymmetric unit (ASU) in both the OMCC$_{OR}$ and OMCC$_{IR}$ maps (Emsley et al., 2010). This was then copied to occupy the other densities. The model refinement was performed in Phenix (Adams et al., 2010).

2.4.3.4 Structure analysis and visualisation

The visualisation of the cryo-EM maps and the atomic models was done using the following programmes: Chimera, ChimeraX and PyMol (Pettersen et al., 2004, Schrodinger, 2015, Goddard et al., 2018).

Chapter 3: The expression, optimisation of the F outer membrane core complex purification and cryo-EM dataset collection

3.1 Introduction

As introduced in Chapter 1, T4SS are remarkable machineries which are involved in a plethora of different functions. While structural differences exist amongst the T4SS, the general architecture is shared. Overall, they are formed from two membrane spanning subassemblies, an inner membrane complex (IMC) and an outer membrane core complex (OMCC), which are connected by a periplasmic channel (Figure 1.6). The OMCC are important subassemblies as they allow the passage of the substrate across the outer membrane. While in the expanded systems, the OMCC structures can have subunit complexity, in its minimal form, all OMCCs consists of the three proteins VirB10, VirB9 and VirB7 or their homologues (Figure 1.4). While there has been significant progress in terms of structural details for the other OMCCs, the complex from the F-T4SS has remained enigmatic.

3.1.1 The F outer membrane complex (F-OMCC)

The F-OMCC is formed from the three proteins TraB, TraK and TraV which share homology to VirB10, VirB9 and VirB7 respectively (Figure 3.1A). Evidence that the three proteins form a membrane spanning complex comes from a yeast two-hybrid study where it was observed that TraV and TraB bind to different segments of TraK suggesting that TraK acts as the bridging protein (Harris et al., 2001). A similar interaction has been characterised in other OMCCs like that belonging to *X. citri*, indicating remarkable conservation amongst the framework of OMCCs despite the apparatus being involved in different functions (Sgro et al., 2018). Since TraV is a lipoprotein, it was speculated that this membrane spanning complex was anchored to the outer membrane via TraV (Doran et al., 1994). The only structural insight into the F-OMCC comes from an *in situ* study of the whole conjugative apparatus using cryo-ET. Here, the complex was shown to adopt 13-fold symmetry and was connected to an IMC with 6-fold symmetry via a stalk. The symmetry of the F-OMCC was surprising as it hasn't been commonly seen in other conjugative T4SS OMCCs, and so far, has only been characterised in the complex belonging to the infectious Dot/Icm machinery. However, due to the low resolution of the complex, how the three Tra proteins arrange to form the F-OMCC was still unknown. This was further exacerbated by the lack of high confidence predicted models. As this work was

performed prior to the success of AlphaFold, due to lack of extensive sequence similarity to the homologous proteins from the other T4SS systems and the unique details of the individual proteins, structure prediction programmes were unable to provide a robust predicted model for the entire proteins. Moreover, while regions of TraK and TraB could be predicted with high confidence due to conserved folds using Phyre 2, TraV could not be predicted at all due to lack of a suitable template (Figure 3.2) (Kelley et al., 2015). This, therefore, supports the requirement of a high-resolution structure of the F-OMCC, which by cryo-electron microscopy (cryo-EM) can currently only be obtained using single particle analysis.

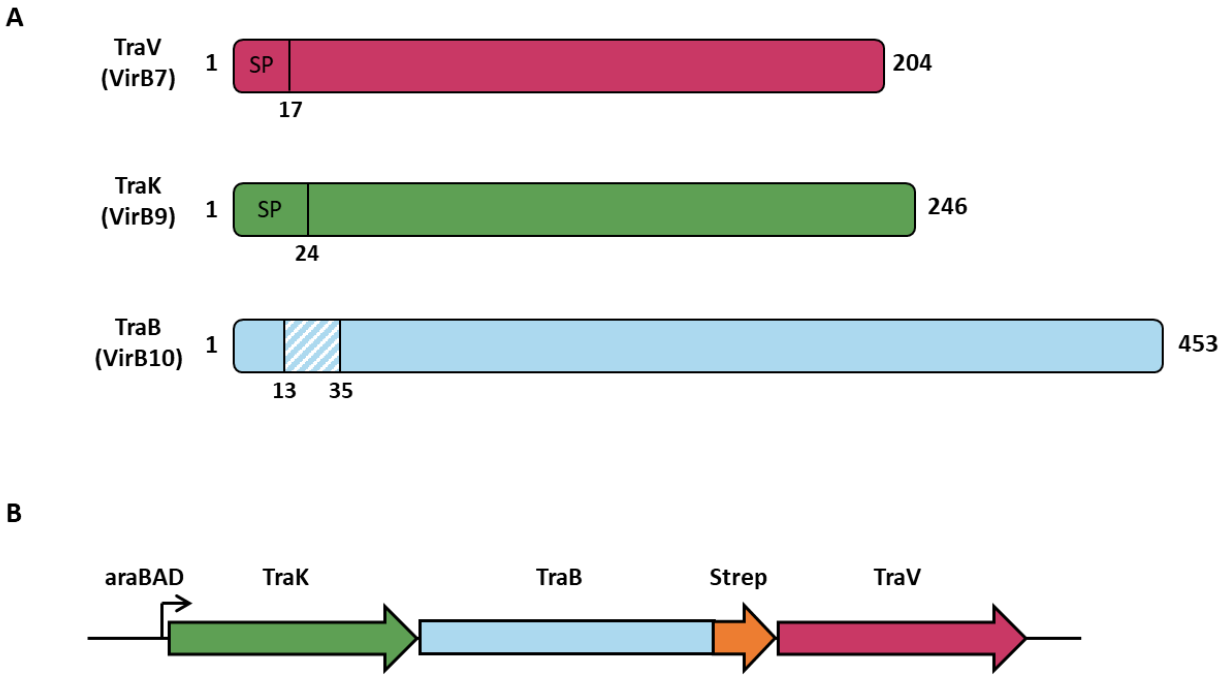

Figure 3.1: The F plasmid outer membrane core complex. (A) The three proteins which construct the F-OMCC along with their domains and boundaries: TraV (purple), TraB (blue), TraK (green). The cross hatching corresponds to the membrane spanning region. SP (Signal peptide) **(B)** The diagram of the operon structure of the plasmid used to express the F-OMCC under the arabinose promoter.

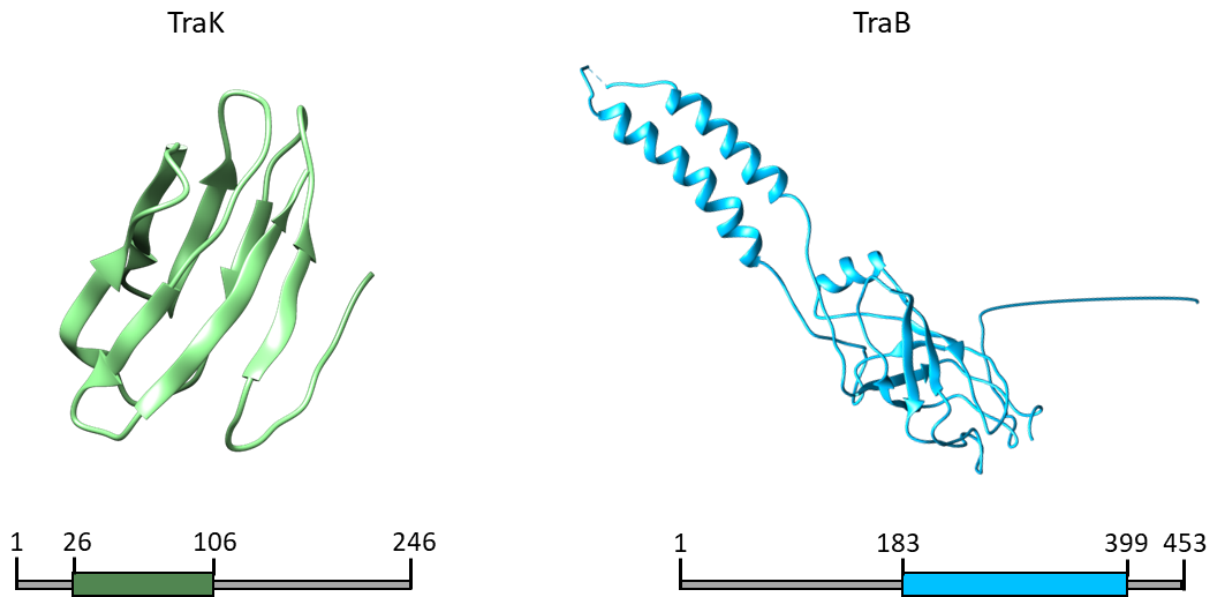

Figure 3.2: The Phyre 2 predicted model of TraK and TraB. The model predicted for TraK represents 32% of the protein and the region is denoted by the orange box. It was modelled with 95.8% confidence. The TraB predicted model represents 42% and was modelled with 99.9% confidence. The TraV predicted model is not included because a suitable template was not found to model the structure resulting in a low confidence score.

3.2 Aims of this study

The aim of this Chapter was to optimise the F-OMCC sample preparation so that it is suitable for structural studies by single particle cryo-EM. Firstly, a test was carried out to confirm expression of the complex followed by a detergent screen to assess the ability of different detergents to extract the F-OMCC from the membrane, to maximises protein concentration. Depending on the results of the detergent screens, the detergents which appeared to be the most effective at solubilising the complex was used for purification trials to probe their efficiency in stabilising the complex. The sample was visualised by negative stain to assess homogeneity and concentration before being optimised for cryo-EM vitrification.

3.3 Results

3.3.1 The expression and detergent screening of F-OMCC

The pBAD_TraK_TraB^{Strep}_TraV construct was engineered such that it encodes for a C-terminally Strep-II tagged TraB (Figure 3.1B). The expression of the complex was confirmed by western blot and using anti-Strep antibodies to probe for presence of TraB. The western blot approach was used as the basis for developing a detergent screen to assess the ability of different detergents to extract the complex from the bacterial membrane (Figure 3.3). The detergent screen was performed alongside my postgraduate student Jemma Betts. The solubilised protein was compared to the total protein, and it was found that the following detergents were the most effective at extracting the complex than others: Cymal-5, combination of DDM and digitonin, and combination of DDM and GDN (synthetic digitonin).

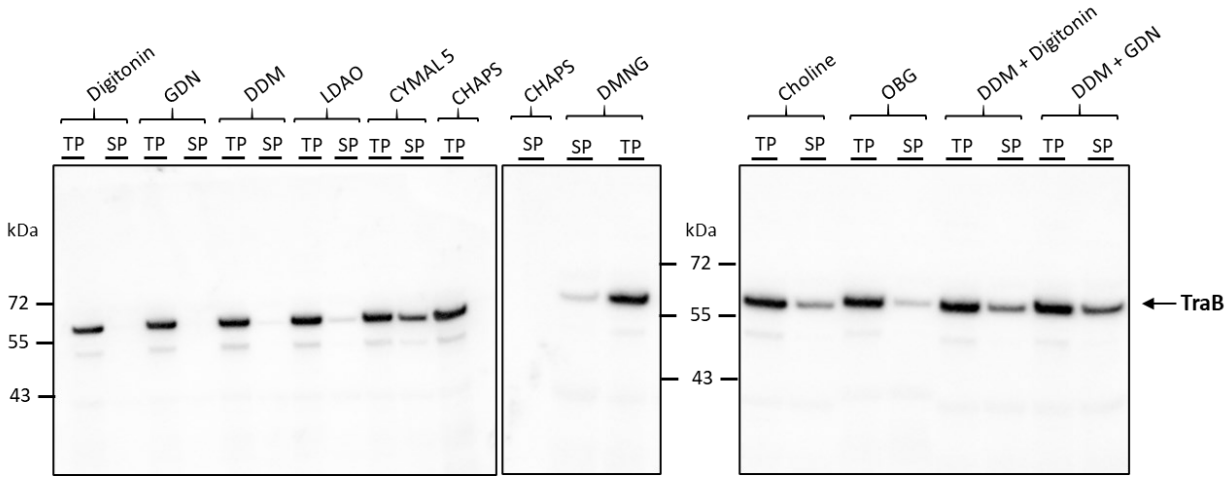

Figure 3.3: The detergent extraction of the F-OMCC. (A) A western blot assessing the ability of different detergents to solubilise F-OMCC from the membrane using TraB as proxy. TP (total protein), SP (soluble protein)

3.3.2 The preliminary cryo-EM structure of the F-OMCC purified with DDM and GDN

3.3.2.1 The purification of the F-OMCC indicates that the complex elutes as a shoulder peak to the void peak in the gel filtration

Based on the results of the detergent screen, the detergent combination of DDM and GDN was chosen. The decision to use synthetic digitonin (GDN) was made to avoid batch-to-batch variations that exists with naturally occurring digitonin and because GDN is more water soluble with a lower critical micelle

concentration (18 µM). The F-OMCC was extracted using the combination of detergents and then purified by two methods of chromatography. Firstly, the complex was subjected to affinity purification using a StrepTrap column (Figure 3.4A). A small elution peak was observed which upon analysis by SDS-PAGE appeared to contain the three proteins: TraB (49 kDa), TraK (24 kDa), and TraV (19 kDa) (Figure 3.4B). The fractions containing the proteins forming the F-OMCC were pooled, concentrated, and further purified by gel filtration. A sharp void peak was observed just after 8 ml with a shoulder peak (Figure 3.4C). The corresponding SDS-PAGE gel indicated that majority of the complex was present in the void, but a proportion was also present in the shoulder (Figure 3.4D).

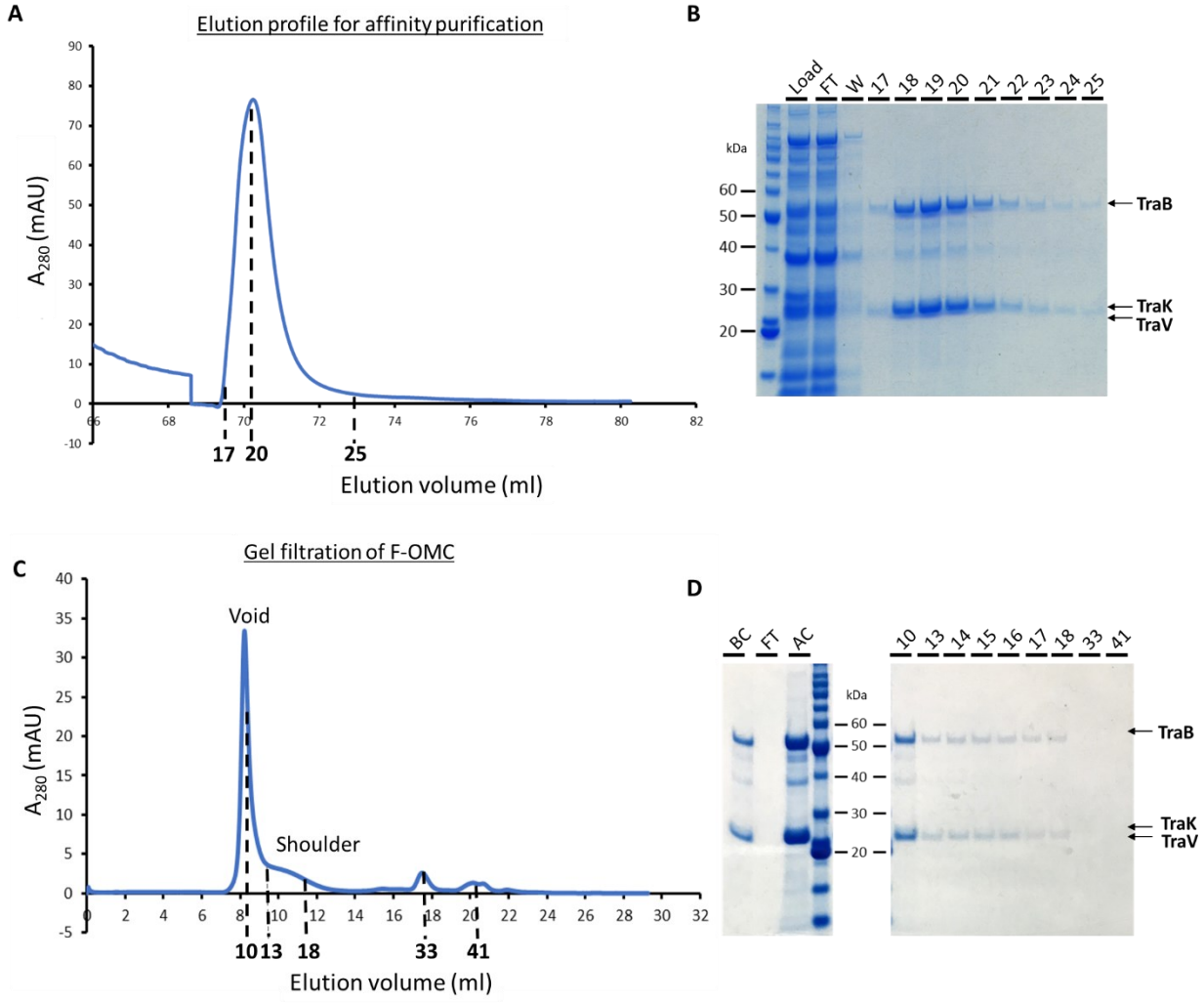

Figure 3.4: The purification of F-OMCC. (A) The StrepTrap elution profile of F-OMCC with the fractions marked. **(B)** The SDS-PAGE gel corresponding to the affinity purification. **(C)** The gel filtration of the F-OMCC which shows a large void peak with an accompanying shoulder. The fractions have been marked. **(D)** The SDS-PAGE gel corresponding to the gel filtration. FT (Flow-through), W (Wash), BC (Before concentrating), AC (After concentrating)

Upon visualisation of the complex from both peaks by negative stain, it showed that the F-OMCC in the void peak was aggregated (Figure 3.5). However, the particles in the shoulder were disperse and displayed clear symmetric rings.

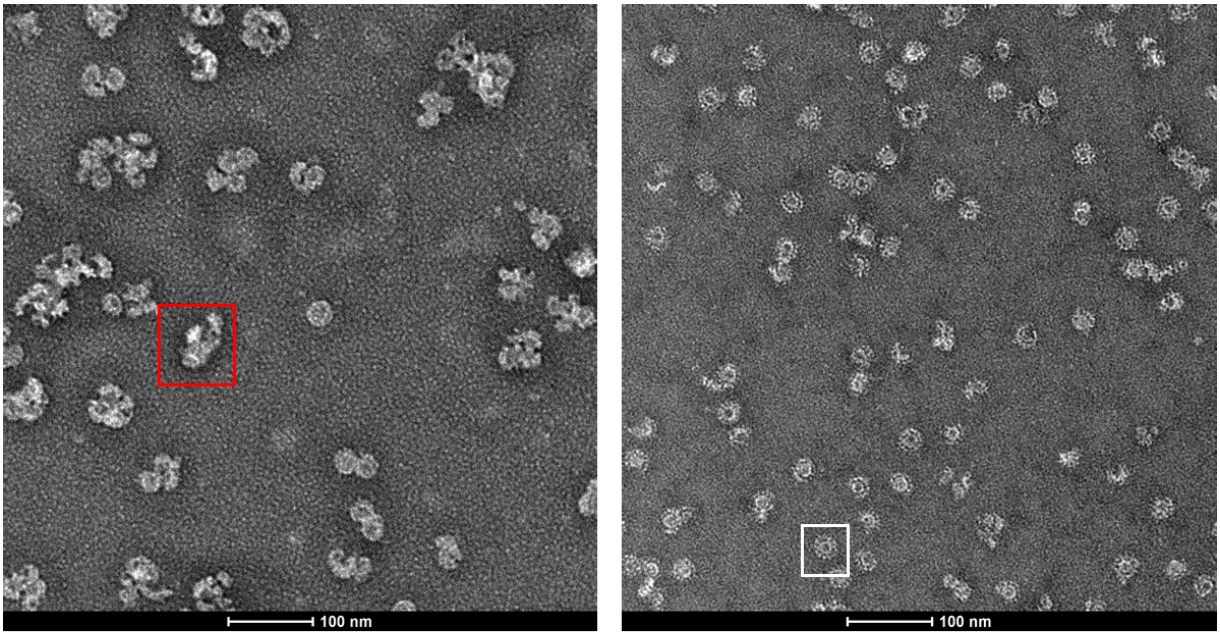

Figure 3.5: The negative stain electron microscopy of the F-OMCC from the gel filtration fractions. The complex in the void appears aggregated as exemplified by the red box. However, the F-OMCC in the shoulder appears more disperse and clear ring like structures can be seen, like the particle boxed in white.

3.3.2.2 Negative stain model of the F-OMCC

Negative electron microscopy was used as a sample quality control to ensure protein stability, concentration, homogeneity, and distribution. Initial screening found that particles from the shoulder peak fractions could be concentrated and then diluted as needed for grid preparation without causing protein degradation or complex dissociation or even, detrimental aggregation. This was valuable since the low absorbance of the shoulder peak meant that without the ability to concentrate the fractions, the particle number would be too little for cryo-EM. Since, the particles in the shoulder showed clear complexes, a small dataset containing 200 micrographs was collected on the FEI CM200 microscope operating at a voltage of 200 kV and equipped with a TVIPS camera. The dataset was processed using Relion 3.1 (Zivanov et al., 2018) as mentioned in section 2.4.3.1 where the 2D class averages clearly showed particles with 13-fold symmetry (Figure 3.6). Furthermore, various orientations of the complex

could be seen in the 2D class averages which highlighted that the complex did not suffer from preferential orientation. Approximately 25000 particles were subjected to C13 symmetry to obtain an initial model (Figure 3.6). The symmetry and dimensions were in line with that previously seen by cryo-ET highlighting that the three proteins which were expressed recombinantly are coming together to form the complex as seen in the *in vivo* analysis (Hu et al., 2019a).

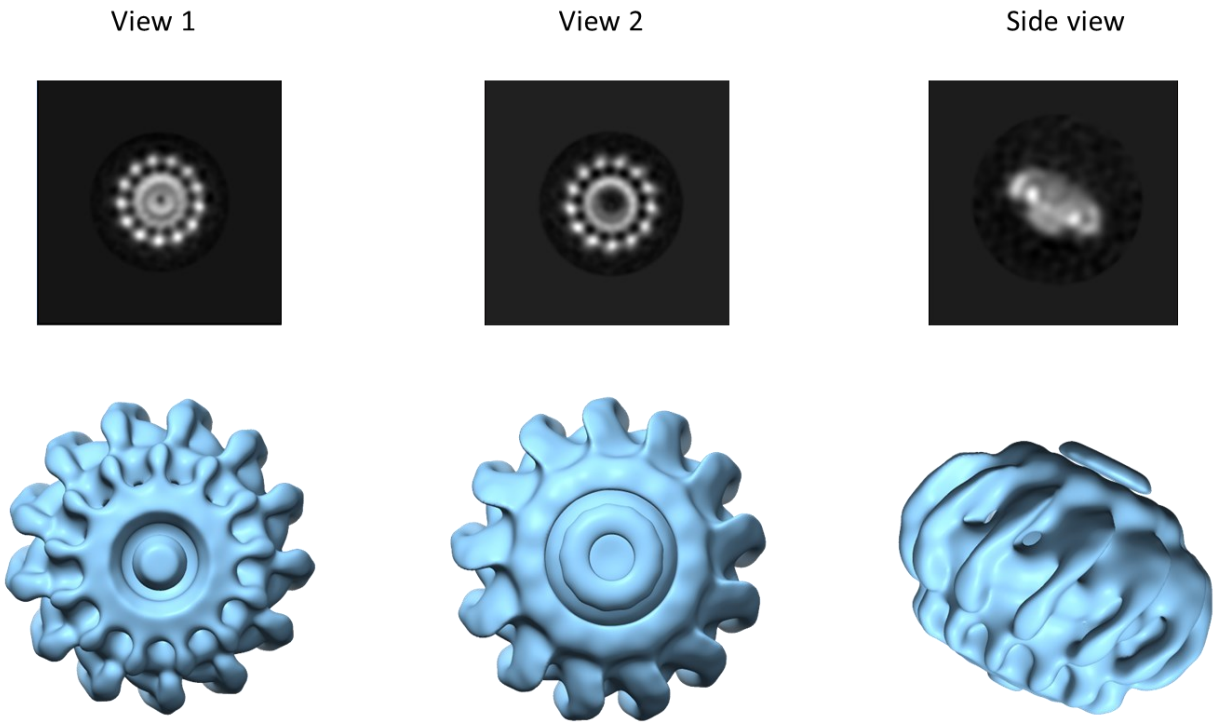

Figure 3.6: The negative stain initial model of F-OMCC. Three representative 2D class averages showing characteristic views along with the initial 3D model of the complex purified with DDM and GDN.

3.3.2.3 The F-OMCC appears to prefer carbon support to adsorb across the ice

As the particles appeared adequate under negative stain, the sample was taken to cryo-EM for high resolution structure determination. One of the key optimisations that was required for F-OMCC was to encourage adsorption across the ice (Figure 3.7A). Initially C-Flat grids were used, however, there was poor distribution across the holes. Moreover, the particles were found concentrated around the edges of the holes indicating preferential binding to the carbon support (Figure 3.7B). Alternatively, lacey grids with an ultrathin continuous carbon support were trialled which resulted in a good distribution of particles across the ice with a drastic improvement in

concentration (Figure 3.7B). Finally, various freezing conditions were tested on the Vitrobot with varying blot times to achieve optimal ice thickness.

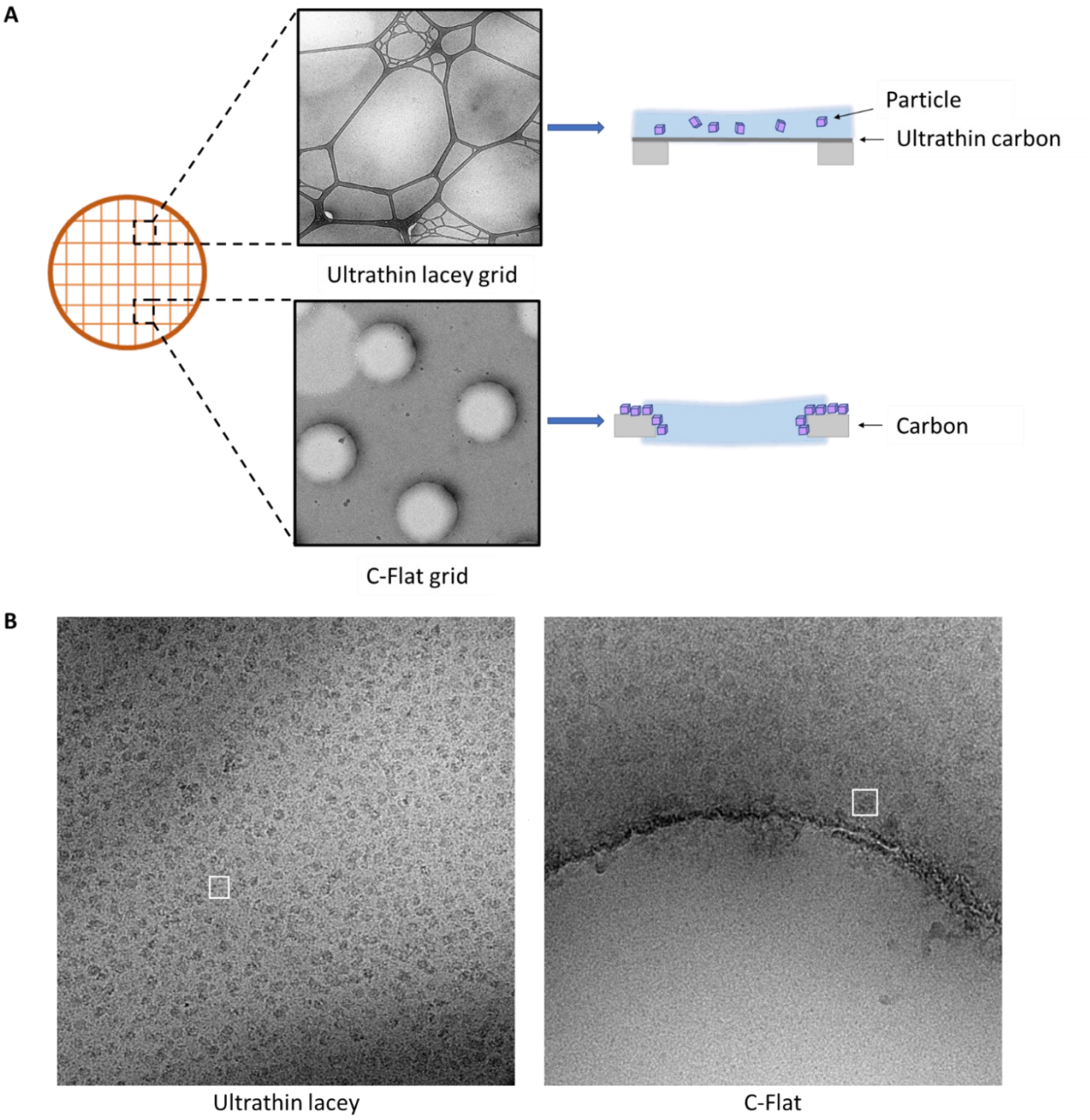

Figure 3.7: The F-OMCC appears to prefer carbon. (A) A schematic showing the difference between the Ultrathin lacey and C-Flat grid, and the distribution of particles when applied to the grids. **(B)** Representative micrographs which show that on the lacey grid, the particles (boxed in white) are well dispersed across the ice, whereas on the C-Flat grid, the particles are populated around the carbon edge.

3.3.2.4 Preliminary Cryo-EM structure of F-OMCC

Once the vitrification of the sample into ice on ultrathin lacey grids had been optimised, a dataset was collected at the eBIC facility at Diamond Light Source using the FEI Titan Krios transmission electron microscope operating at 300 kV and equipped with a K2 Summit detector (Gatan). The EPU software (FEI) allowed for automated collection resulting in the compilation of 4406 movies. Within Relion 3.1, the movies were firstly dose-weighted and then aligned using MotionCor2, and then the contrast transfer function (CTF) values were estimated using CTFFIND4 (Zivanov et al., 2018, Zheng et al., 2017, Rohou and Grigorieff, 2015). Using Relion, a total of 240,713 particles were auto-picked. At this point, image processing was done simultaneously on two different software packages: Relion and cryoSPARC (Punjani et al., 2017). After several rounds of 2D class averaging, the bad particles were filtered out and a set of approximately 64420 (RELION 3.1) and 58175 (cryoSPARC) particles belonging to the good classes were kept (Figure 3.8). Since the class averages appeared more defined in cryoSPARC, a 3D sharpened map was reconstructed from approximately 36000 particles using an *ab initio* model (Figure 3.9).

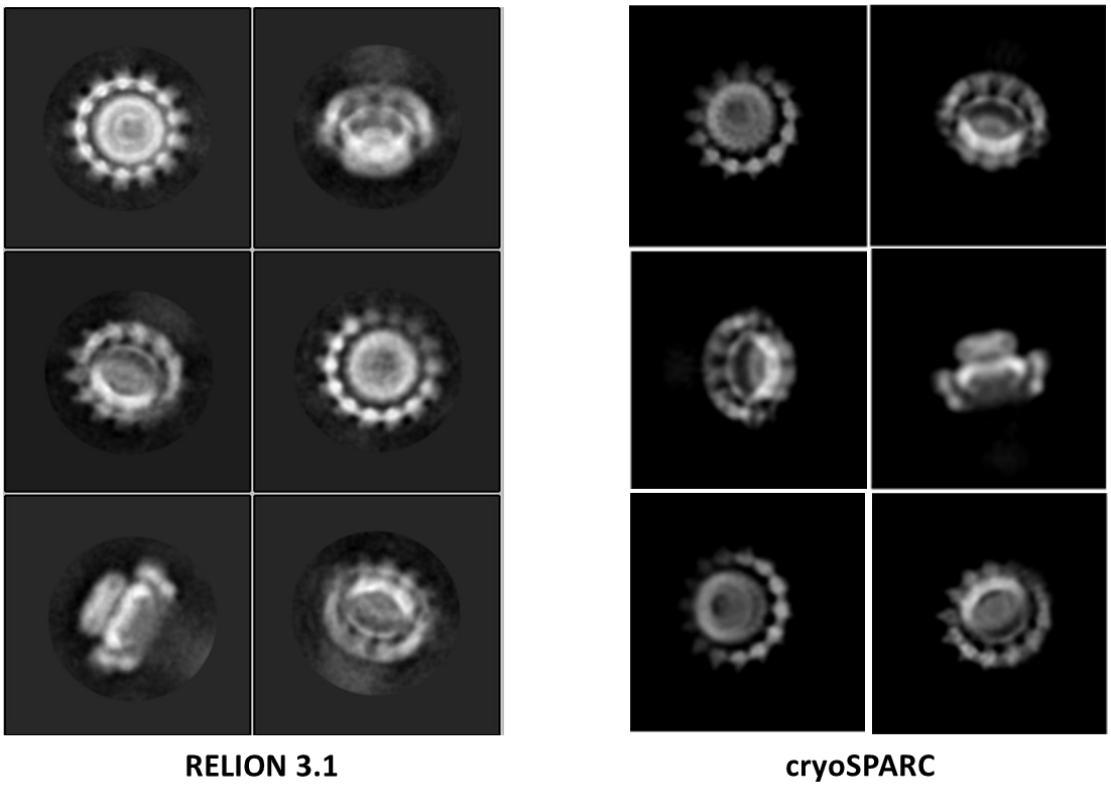

Figure 3.8: The comparison of the 2D class averages obtained from the two different processing software.

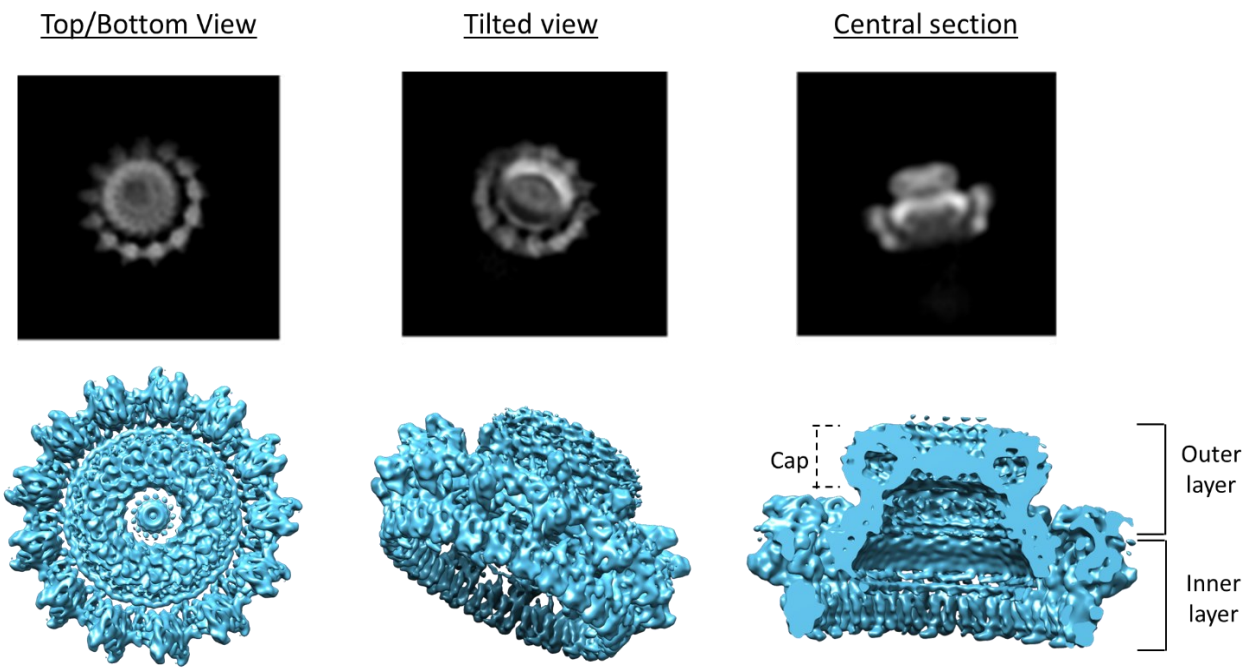

Figure 3.9: The preliminary Cryo-EM structure of F-OMCC. The representative views of the complex, purified with DDM and GDN, are shown with the representative 2D class averages. The F-OMCC shows two concentric rings with 13-fold symmetry. It appears that the complex is formed from two chambers: an outer and inner layer.

The structure can be described as a 'flying saucer' with 13 lateral pods. As previously indicated by the 2D class averages from the negative stain dataset, the complex appears to be formed from two rings. While the outer ring showed clear C13 symmetry, the inner ring could not be resolved adequately to give insight into its symmetric arrangement and appeared to introduce heterogeneity in the dataset. While the overall resolution of the structure was not high enough to build an atomic structure, it gave insight into the general architecture of the complex and some secondary structure definition could be observed. The F-OMCC has dimensions of approximately 260 Å in diameter and 165 Å in height, which can be divided into an outer and inner layer. The outer layer can be further separated into two segments: a cap and the main body. Unfortunately, details of the cap were occluded by a detergent shell. The structural details into how the proteins fit into this structure and interact was hindered due to the resolution. Overall, high resolution structure determination was limited by suboptimal ice thickness, inherent heterogeneity within the two rings and free detergent micelles. Together, this resulted in the loss of many particles during the processing.

3.3.3 Purification of F-OMCC with LDAO which resulted in high resolution cryo-EM data collection

3.3.3.1 Purification of F-OMCC with LDAO

One of the main drawbacks of the purification with DDM and GDN is the great presence of free detergent micelles. This not only causes electron scattering but also produced background noise and intense signal at the 2D class average stage which meant many particles had to be discarded. Moreover, DDM has been reported to form relatively large micelles (Mw 65–70 kDa) and a large detergent belt around the protein which was observed around the cap of the F-OMCC (Figure 3.9) (Slotboom et al., 2008, Chaptal et al., 2017). This could be because DDM has a large head group, which is also true for GDN, and together may form a large detergent shell around the transmembrane cap region of the F-OMCC.

Subsequently, efforts were made to improve the purification of F-OMCC by use of other detergents and changing buffer conditions to investigate if the complex could be eluted as a peak separate to the void during the gel filtration. Based on the detergent screen, Cymal-5 was identified to be effective in solubilising the F-OMCC, however during the purifications, there were issues with reproducibility when that detergent was used. While LDAO was not effective at solubilising the complex independently, the combination of LDAO and DDM was tried since it has been proven to be successful in solubilising the *X. citri* OMCC (Sgro et al., 2018). Furthermore, the detergent screen highlighted that although some detergents alone were ineffective at extracting the complex from the membrane, when combined, they proved to be successful, as indicated by DDM and GDN. For the purification using LDAO, DDM was only combined for the solubilisation step. The subsequent buffers used for the purification steps only contained LDAO, in an effort to buffer exchange and remove the DDM micelles.

The F-OMCC was initially purified using a StrepTrap column and then injected onto a gel filtration column for further purification (Figure 3.10). The same trend was observed as seen in Figure 3.4C with DDM and GDN, whereby there is a prominent void peak followed by a shoulder peak which upon SDS-PAGE analysis contained the three proteins TraB, TraK and TraV. The identity of the three proteins in the bands were confirmed by liquid chromatography with tandem mass spectrometry (LC-MS/MS) at University of St Andrew. Initial analysis by negative stain alluded to the absence of a large detergent belt and the 2D class averages looked well defined.

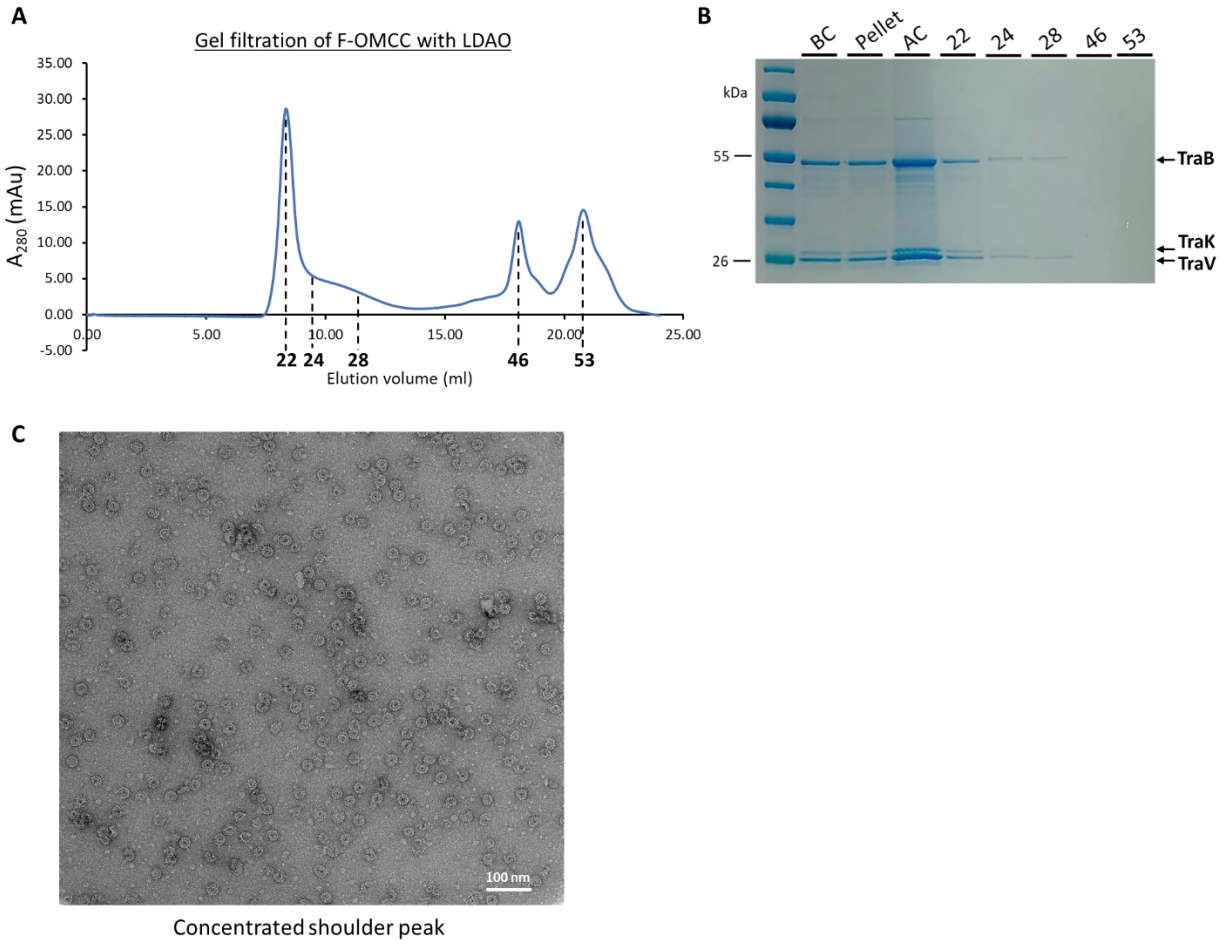

Figure 3.10: The purification of the complex with LDAO. (A) The gel filtration of the complex purified with LDAO, and the fractions have been indicated. **(B)** The SDS-PAGE gel corresponding to the gel filtration profile. (C) A representative micrograph of the sample from the shoulder peak. The scale bar is set at 100 nm. BC (Before concentrating), AC (After concentrating)

3.3.3.2 Collection of the high resolution Cryo-EM dataset of F-OMCC

The sample purified with LDAO was vitrified in ice using lacey grid with ultrathin continuous carbon film. To improve reproducibility in ice thickness across grid holes and between grid preparations, the Leica automatic plunge freezer EM GP2 was used to plunge freeze the complex due to technical issues with the in-house Vitrobot. The grids with optimal ice thickness and protein concentration were used for a high-resolution data acquisition at LonCEM (London Consortium EM, Francis Crick Institute) on the Titan Krios (FEI) operating at 300 kV and equipped with a K3 detector (Gatan). A total of 11588 movies were acquired

with 38 frames per image/movie with a pixel size of 1.08 Å. Full details of the data collection are provided in section 2.4.3.2 along with the full statistics being listed in Table 2.9.

3.3.3.3 Image processing of the high resolution Cryo-EM dataset of F-OMCC

The dataset was processed using cryoSPARC and the workflow is shown in Figure 3.11. The movies were firstly aligned and summed using patch motion correction, and then, the CTF was estimated. The resulting micrographs were screened for good Thon rings and ideal ice thickness and the remaining micrographs, approximately 11000, were further processed. Approximately 3500 particles were manually picked, and the resulting representative 2D class averages were used for automated particle picking of a small dataset to establish optimal auto-picking parameters to reduce picking of unwanted particles along and within the borders of the grids, and ensure all genuine particles were picked. Then, automated particle picking was carried out for the entire dataset leading to a collection of approximately 1,450,000 particles which yielded well-resolved 2D class averages displaying different orientations of the complex. Interestingly, upon closer visual inspection, the 2D class averages representing the top/bottom view indicated the presence of two different symmetries existing within the complex: the outer ring adopting 13-fold symmetry ($OMCC_{OR}$) and the central inner ring having 17-fold symmetry ($OMCC_{IR}$).

Due to the heterogeneity this introduces to the dataset, the two rings were processed and refined separately by masking the two regions and generating high resolution maps corresponding to the C13 and C17 rings. After rounds of 2D class averaging to filter bad particles, a set of approximately 298000 particles displayed defined $OMCC_{OR}$ features and upon further rounds of 2D classification, a stack of roughly 75000 particles showed clear $OMCC_{IR}$ features. For the $OMCC_{OR}$, an *ab initio* map was created and all the $OMCC_{OR}$ particles were used for 3D homogenous refinement with C13 symmetry imposed and a mask containing the outer ring section. Similarly, the $OMCC_{IR}$ particles were subjected to the same homogenous refinement but with C17 imposed and a mask encompassing the inner ring segment. Following application of a B factor to the two maps, the overall of resolution of the $OMCC_{OR}$ and $OMCC_{IR}$ was 3.3 Å and 3.4 Å respectively. Simultaneously, to confirm the two symmetries genuinely existed, a C1 map was generated from approximately 95000 particles which upon B factor application had a resolution of 5.4 Å (Figure 3.12).

The model building and refinement into the density maps was performed by collaborator Dr Aravindan Illangovan. Briefly, the model building commenced by identifying the asymmetric unit (ASU) in both the $OMCC_{OR}$ and $OMCC_{IR}$ maps.

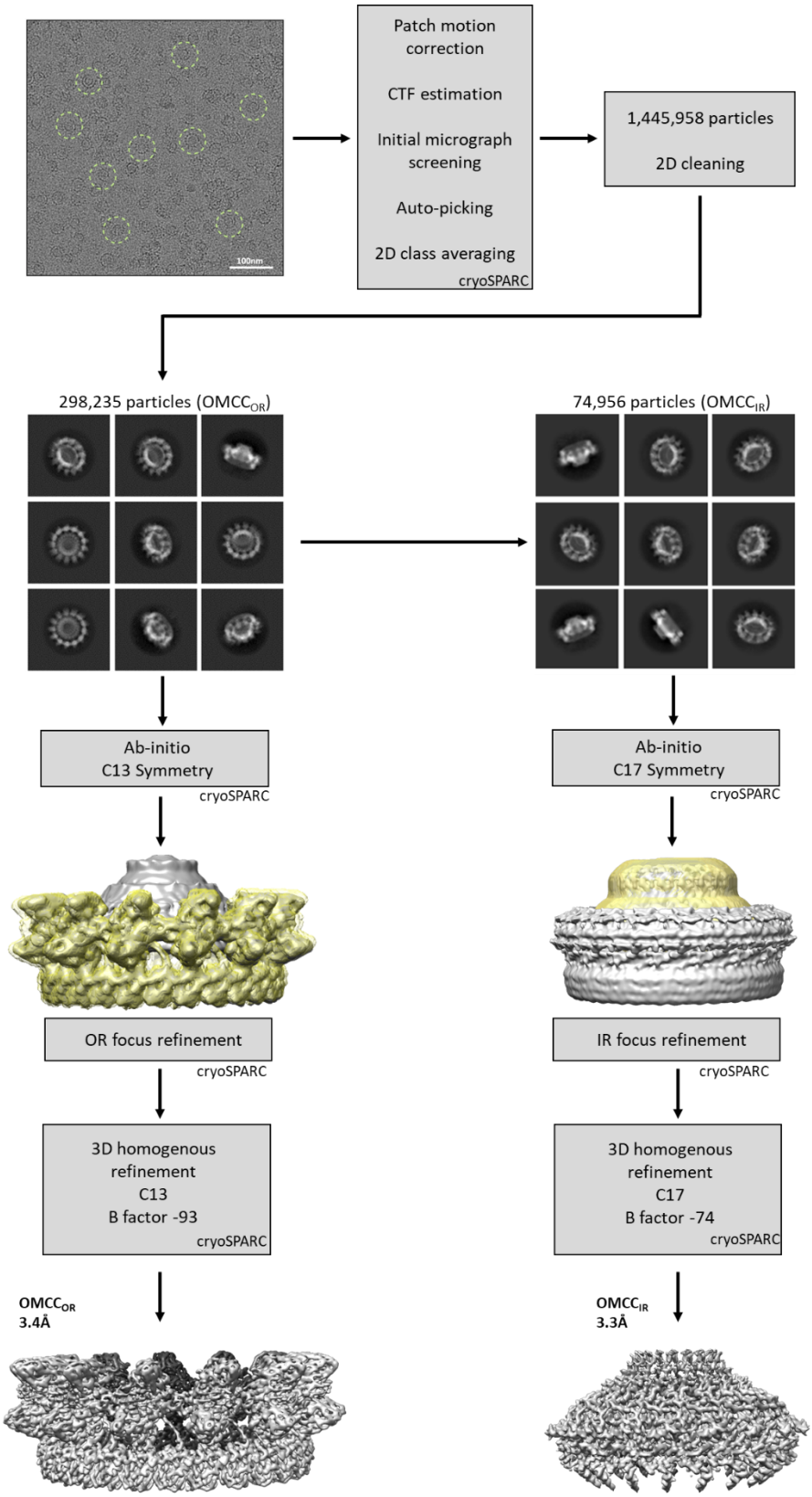

Figure 3.11: The Cryo-EM processing workflow that was used to generate the high-resolution maps for the OMCC$_{OR}$ and OMCC$_{IR}$. The mask is shown in yellow.

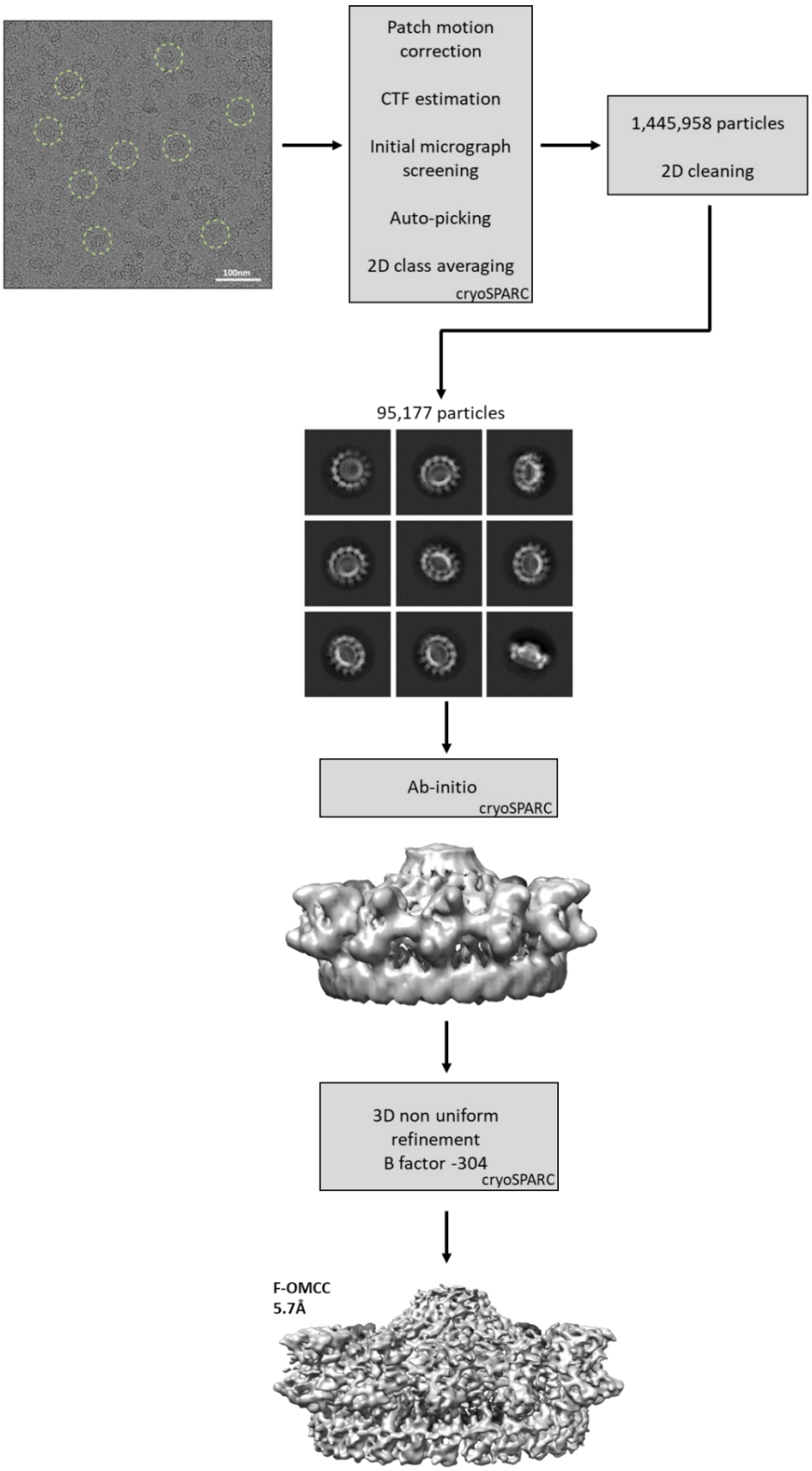

Figure 3.12: The Cryo-EM processing workflow that was used to generate the C1 map for symmetry analysis.

3.4 Discussion

This Chapter outlined the expression, purification trials and electron microscopy analyses of the F-OMCC. The complex was purified using two different detergent combinations, DDM and GDN, and LDAO with DDM supplemented to aid with solubilisation. At the gel filtration stage, the complex was observed to elute predominately in the void peak but was also present in the shoulder peak where it appeared dispersed. Using the sample from the shoulder fractions, a negative stain and cryo-EM model were constructed for the purification with DDM and GDN. For the sample in LDAO, a high resolution dataset was collected.

3.4.1 The preliminary models show a complex with 13-fold symmetry

The negative stain model and initial cryo-EM reconstruction of the complex purified with DDM and GDN gave insight into the structure of the conjugative F-OMCC which can be described as having a 'flying saucer' shape with 13-fold symmetry formed from two concentric rings (Figure 3.6 and 3.9). The symmetry and the overall shape observed is in agreement with the *in situ* structure (Hu et al., 2019a).

The cryo-EM reconstruction also showed a chambered architecture, splitting the structure into an outer and inner layer. This arrangement has been observed in all T4SS OMCCs solved to date, but the function it confers remains unknown. While it was not clear in the negative stain model, from the cryo-EM structure is appears that outer layer can be further divided into a cap region and the main body. It is plausible that the cap is one of the main ways in which the F-OMCC is associated with the outer membrane based on the observations from previously solved OMCCs, such as that from *X. citri* and more recently in the R388 structure (Sgro et al., 2018, Mace et al., 2022). This is further supported by the 3D cryo-EM structure which displays twisted cylindrical features at the cap which are characteristic of α helices (Figure 3.9). The cap is likely to be formed from TraB based on knowledge from previous structures where it is found to be occupied by VirB10 and its homologues. This is supported by the TraB predicted model which shows a two-bundle helix (Figure 3.2). Similarly, the bottom of the I-layer shows secondary structural features which based on previous structures of OMCCs, is likely to be occupied by the N-terminal domain of TraK. This is reinforced by the Phyre prediction of TraK which shows a conserved β-sandwich (Figure 3.2).

Nevertheless, the resolution of the dataset collected with DDM and GDN was limited by heterogeneity observed in the sample as represented by blurring in 2D class averages, suboptimal ice thickness and free

detergent micelles. This was improved in the purification of the complex with LDAO which resulted in the collection of a high resolution dataset, and with the structure detailed in Chapter 4.

3.4.2 The detergent LDAO appeared to be more suitable for cryo-EM studies than DDM and GDN

The cryo-EM 2D class averages obtained from the dataset with LDAO compared to the one obtained with DDM and GDN, appeared to show finer details (Figure 3.11). This was more clearly observed at the cap region of the F-OMCC in the 2D class averages representing the side view. For the DDM and GDN dataset, the cap seemed to be encompassed by a detergent shell thereby manifesting as a bulky density in both the 2D class average and in the 3D reconstruction where this detergent density could not be removed. On the contrary, the 2D class averages for the LDAO dataset displayed a finer cap region. A plausible explanation for this could be related to the detergent micelle size of the detergents. Both DDM and GDN have larger micelle size compared to LDAO with DDM micelle being 72 kDa and GDN being slightly higher (Timmins et al., 1988, Chae et al., 2012). The micelle size of LDAO is less than 20 kDa which means that the small diameter of the micelle covers a smaller proportion of the transmembrane region of the complex compared to the detergents with a larger micelle size (Timmins et al., 1988). A similar phenomenon was also observed in the electron microscopy of the N-type ATPase rotor ring from *Burkholderia pseudomallei* (Schulz et al., 2017). In addition, the 2D class averages in LDAO showed greater details. The density of LDAO is similar to amorphous ice compared to the other surfactants which would have improved contrast between the complex and the background allowing for better alignment at the image processing stage (Schulz et al., 2017). Altogether, the use of LDAO may have improved the cryo-EM sample preparation thereby aiding the collection of a larger dataset compared to the one collected with DDM and GDN and allowing for the clear observation that a symmetry mismatch exists in the complex.

3.4.3 The 2D class averages from the high resolution dataset show a symmetry mismatch existing in the F-OMCC

During image processing of the high resolution cryo-EM dataset, it became clear upon visual inspection that the two concentric rings adopt two different symmetries. The outer ring ($OMCC_{OR}$) shows 13-fold symmetry while the inner ring ($OMCC_{IR}$) adopts 17-fold symmetry. While this will be further explored in Chapter 4, briefly, the presence of this symmetry mismatch was surprising as there was no indication of it in the *in situ* analysis (Hu et al., 2019a). In addition, the presence of the two symmetries meant that to

optimise the resolution, the two rings had to be further processed separately at the 3D classification step in a similar manner to the processing required for the Dot/Icm OMCC (Durie et al., 2020, Sheedlo et al., 2021).

Chapter 4: The structure of the F-OMCC

4.1 Introduction

The F-OMCC is a subassembly in the F-T4SS and is found spanning the outer membrane. It is formed by three proteins: TraK, TraV and TraB. The complex is required for conjugation to proceed. Despite the importance of the F-OMCC, structural data has been sparse. Insight into the overall dimension was provided by a cryo-ET reconstruction which also showed the F-OMCC adopting 13-fold symmetry (Hu et al., 2019a). However, this did not further knowledge on how the three proteins form the substructure and information on the architecture remained elusive. To decipher these details, the structure of the F-OMCC was solved using single particle cryo-EM.

In Chapter 3, the experimental work which led to the 3D structure of the F-OMCC has been described. The three proteins which form the complex encoded by the IncFV plasmid pED208 were overexpressed recombinantly. The complex was extracted from the bacterial membrane and then, purified by affinity chromatography followed by gel filtration. The homogeneity and integrity of the sample was confirmed by negative stain EM. The complex was vitrified in ice using lacey grids with an ultrathin continuous carbon support. A cryo-EM dataset was collected and approximately 1.45 million particles produced 2D class averages with various orientations of the complex. Interestingly, upon close visual inspection, it became increasingly clear that the complex was formed from two rings with two different symmetries. Based on this symmetry mismatch, the two rings were masked from an initial low-resolution map, resulting in two high-resolution maps corresponding to $OMCC_{OR}$ and $OMCC_{IR}$.

4.2 The aim of this study

The aim of this Chapter was to describe the 3D structure of the F-OMCC that was obtained by single particle cryo-EM analysis.

4.3 Results

4.3.1 The overall architecture of the complex

The F-OMCC structure was solved at an overall resolution of 3.3 Å resolution. The complex is formed from two concentric rings which have two distinct symmetries. The outer ring, which spans the periphery of the complex, has 13-fold symmetry while the inner ring adopts 17-fold symmetry (Figure 4.1A). Together, they form a 2.1 MDa complex with overall dimensions of 115 Å in height and 268 Å in diameter. As suggested by preliminary work highlighted in section 3.3.2.4, the complex has a hollow centre with a chambered architecture. The conical inner chamber (I-chamber) is outlined by the I-layer and the outer chamber (O-chamber) is outlined by the O-layer (Figure 4.1B). Both are primarily defined by TraB which forms the interior walls of the complex. The O-chamber has an internal diameter of 110 Å which then tapers off to form a membrane embedded exit channel with a diameter of 50 Å. This makes the exit channel too narrow to accommodate the 87 Å external diameter of the pilus (Costa et al., 2016). Subsequently, it is likely that the structure represents the complex in the closed state.

Ultimately, the complex is formed from 69 polypeptide chains. The 17 chains of TraB span the inner ring of which 13 were observed to traverse down and form part of the outer ring. There are 26 polypeptide chains of TraK and TraV each. While the TraK remains at the outer ring, 17 out of the 26 TraV chains extend into the inner ring. Together, this creates an intricate connection between the two rings.

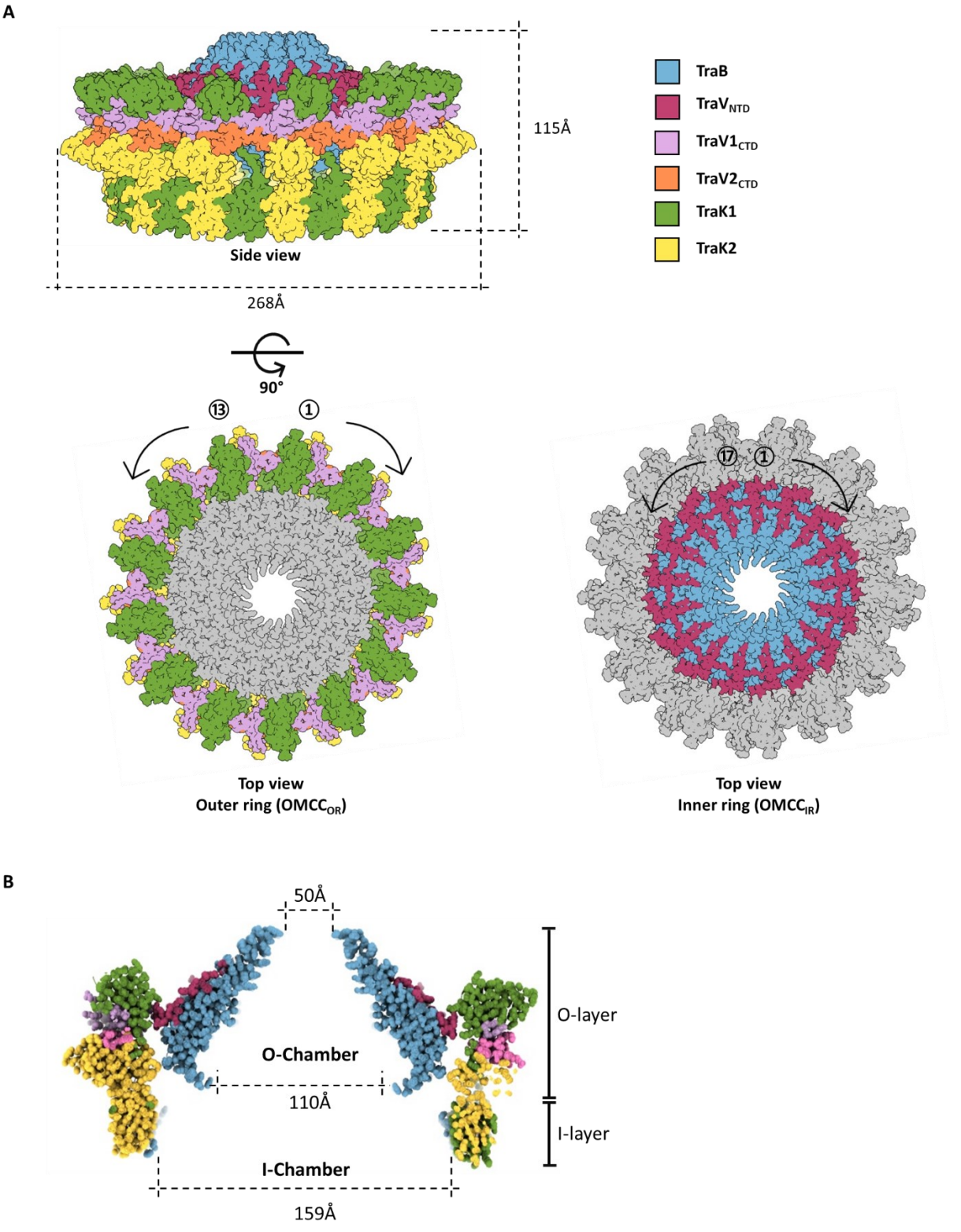

Figure 4.1: The structure of the F-OMCC. (A) Overview of the side and top view of the complex in surface representation with each protein coloured in their respective colours. The complex is made from two concentric rings with two distinct symmetries **(B)** The cut-away side view model of the F-OMCC showing structural features.

4.3.2 The F-OMCC asymmetric unit

Due to the symmetry mismatch, the F-OMCC has two asymmetric units. The asymmetric unit for the outer ring (ASU_{OR}) were built from the models of TraK1, TraK2, the C-terminal domain (CTD) of TraV1 ($TraV1_{CTD}$), $TraV2_{CTD}$ and $TraB_{CTD\ (177\text{-}186)}$ (Figure 4.2). Alternatively, the models of $TraB_{CTD}$ and $TraV_{NTD}$ fitted into the asymmetric unit for the inner ring (ASU_{IR}). The models were built by firstly identifying the single asymmetric unit in both the $OMCC_{OR}$ and $OMCC_{IR}$ maps and building into the density. With regards to the inner ring, the conserved antiparallel helices of $TraB_{CTD}$ were used as a guide and were manually positioned in the density. Following this, the remainder of the $TraB_{CTD}$ and $TraV_{NTD}$ was built manually by assigning bulky side chains first. With regards to the outer ring, the TraK model was built using the *X. citri* VirB9 structure as a template (Sgro et al., 2018). As with the TraV in the inner ring, the $TraV_{CTD}$ was built by assigning bulky side chains first and the rest of the polypeptide was then completed by identifying the connecting sequence between the bulky residues.

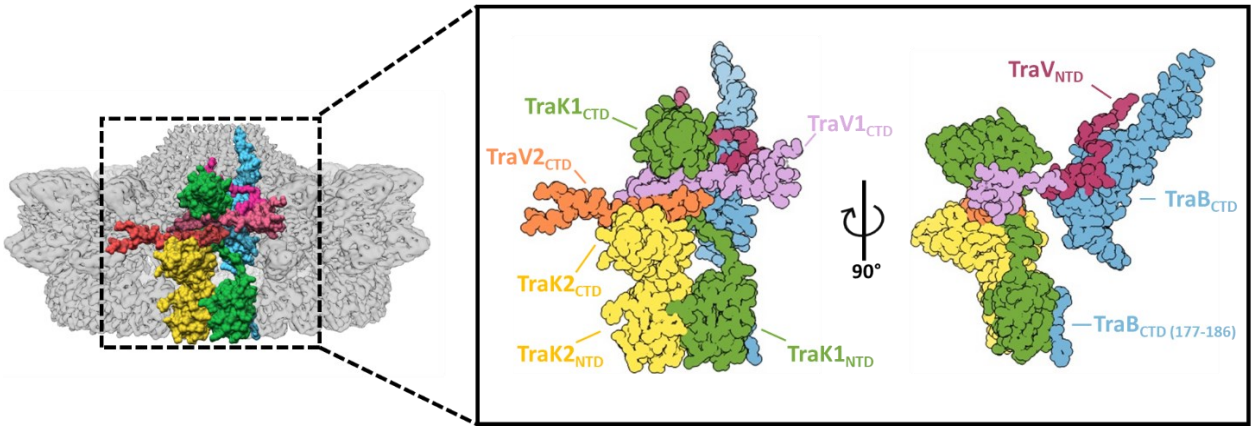

Figure 4.2: The asymmetric units (ASU). The F-OMCC shown in surface representation with the ASUs coloured. The F-OMCC ASUs consists of five proteins namely: TraB, TraK1, TraK2, TraV1 and TraV2.

4.3.2.1 The ASU_{IR}

The ASU_{IR} is largely built from $TraB_{CTD}$ (Figure 4.3A). The structure of $TraB_{CTD}$ is made from 17 β strands and 3 α helices, with 16 β strands forming a β barrel at the core. This β-barrel connects two antiparallel α helices (α1, α2) which forms the exit channel. No electron density was observed between the amino acid residues 328 to 358 for $TraB_{CTD}$ connecting the two α helices. A third α helix (α3) is present at the C-terminal end. This structure of $TraB_{CTD}$ is highly conserved amongst homologues of TraB. Thus, making it an important structural feature. The N-terminal domain of TraV, consisting of the amino acid residues 17

to 54, also forms a part of the ASU_{IR} and includes a single helix, α1. The $TraV_{NTD}$ lies laterally along the TraB in such a manner that an interaction is mediated by 6 of the 17 β strands of $TraB_{CTD}$.

The inner ring appears to be stabilised by a network of contacts between adjacent ASUs (Figure 4.3B). A single ASU_{IR} interacts with four other neighbouring units both in the anticlockwise and clockwise direction. The central reference ASU_{IR}, numbered 0, makes contact with the ASU_{IR} +1 and +2 in the clockwise direction and ASU_{IR} -1 and -2 in the opposite direction. The $OMCC_{IR}$ is secured by extensive inter ASU contacts consisting of hydrogen bonds, salt bridges and hydrophobic interactions. One interaction which is noteworthy is between the $TraV_{NTD}$ chains where they engage in the formation of inter di-sulphide bonds. One $TraV_{NTD}$ forms two di-sulphide linkages with its neighbouring ASU. Cys27 on ASU_{IR} 0 interacts with Cys35 on ASU_{IR} +1, while Cys35 on ASU_{IR} 0 participates in contact with Cys27 on ASU_{IR} -1. This extensive interchain di-sulphide bonds between adjacent $TraV_{NTD}$ chains creates a mesh over the $TraB_{CTD}$ and thereby further supports the inner ring structure. These cysteine residues are only conserved in TraV homologues from the F-T4SS family, indicating a feature unique to these machineries. Another important cysteine residue is Cys18 which is lipidated, as shown for TraV homologues across the T4SSs (Fernandez et al., 1996, McClain et al., 2020). This cysteine is situated as such that it is in an ideal position to aid insertion and stabilisation of the exit channel. One distinct stabilising interaction is between the main chain oxygen of Cys18 on TraV with Tyr366 of $TraB_{CTD}$, which occurs across the ASUs. For example, the $TraV_{NTD}$ located in ASU_{IR} 0 interacts with the $TraB_{CTD}$ in ASU_{IR} +2. Therefore, this highlights the interlinked network of contacts at the inner ring which contributes to the stability F-OMCC.

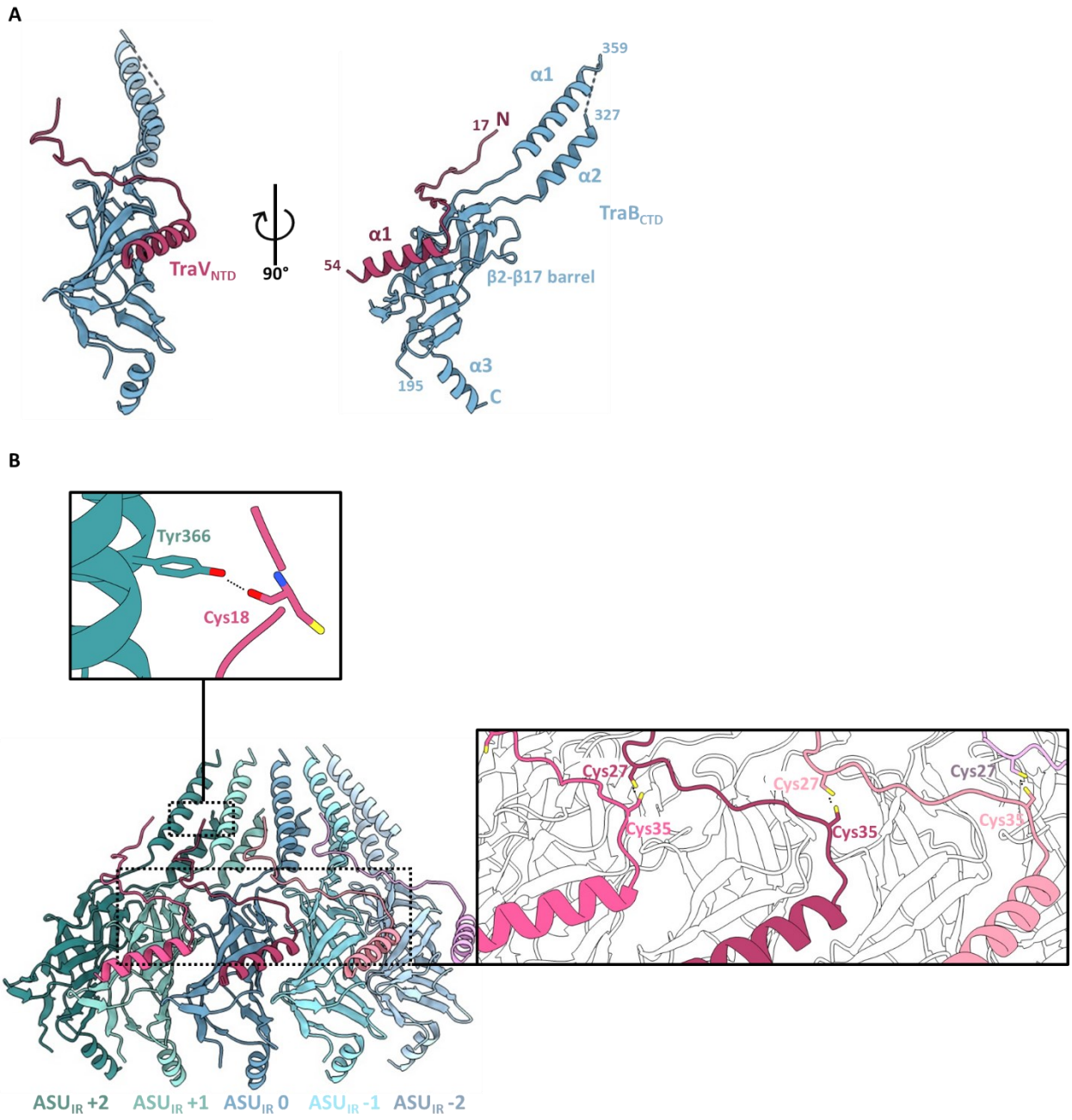

Figure 4.3: The asymmetric unit spanning the inner ring (ASU$_{IR}$). (A) The ribbon representation of the polypeptides forming the ASU$_{IR}$ with the secondary structural features labelled. The inner ring is formed by TraV$_{NTD}$ (magenta) and TraB$_{CTD}$ (blue). **(B)** Side view of five ASU$_{IR}$ highlighting the network of interactions that occur at the inner ring. This includes an interaction between TraB Tyr366 and TraV Cys18 (inset left) and contacts between the TraV$_{NTD}$ chains by formation of di-sulphide bonds between Cys27 and Cys35 (inset, right).

4.3.2.2 The ASU_{OR}

The ASU_{OR} consists of all five polypeptide chains, namely: TraK1, TraK2, $TraV1_{CTD}$, $TraV2_{CTD}$ and $TraB_{CTD\ (177-186)}$ (Figure 4.4A). A large proportion of the ASU_{OR} is structured by TraK which consists of two globular domains. The two TraK molecules adopt two different configurations within the ASU_{OR}. For TraK2, the NTD and CTD are connected by a short linker, thereby placing the two domains close-by. However, for TraK1, the NTD and CTD are further apart and bridged by a long linker. The $TraK_{NTD}$ is made from nine β strands (β1 – β9) divided between two sheets, adopting a β sandwich fold. The $TraK_{CTD}$ is comprised of six β strands (β10 – β15) and two α helices (α1 and α2). However, for $TraK1_{CTD}$, the α1 has been melted to form the long linker. The two CTDs of TraK are related by P2 symmetry. In the crevice formed between the two $TraK_{CTDs}$, two anti-parallel $TraV_{CTD}$ chains lie across supporting the structural arrangement of the two TraKs. The $TraV_{CTD}$ is made from four β strands. Therefore, $TraV1_{CTD}$ and $TraV2_{CTD}$ form two continuous β-sheets that creates a belt around the $OMCC_{OR}$ and stabilises it. A short linker sequence, formed by amino acid residues 69 to 80, assigned to TraV1 acts as one of the main connections between the outer and inner rings of the F-OMCC. However, no electron density was observed for two regions of TraV (residues 55 – 68 and 81 – 156). $TraB_{CTD}$ also forms a part of the ASU_{OR}, where a short stretch of sequence (residues 177 – 186) folds into a β-strand. Positioned by the $TraK_{NTD}$, this TraB fragment acts as one the connections between the two F-OMCC rings.

The outer ring is stabilised by a network of contacts between the ASUs (Figure 4.4B). Each ASU_{OR} contacts another ASU either side. The central ASU_{OR}, termed 0, interacts with the ASU_{OR} +1 and with ASU_{OR} -1. These contacts are reinforced by extensive interactions occurring between the units. Interactions between the TraV short linker (residues 69 - 80) and $TraK1_{CTD}$ was noted which occurs via salt bridge formation between Glu74 on TraV and Arg230 of TraK1, and by hydrogen bonding between TraV Ser71 and TraK1 Gln184. The stability of the $OMCC_{OR}$ is further enforced by hydrogen bonding and hydrophobic interaction between the $TraV_{CTD}$. This includes hydrogen bonding between Asn180 and Trp192 and hydrophobic contacts between Asn180 and Arg162 (Figure 4.4B).

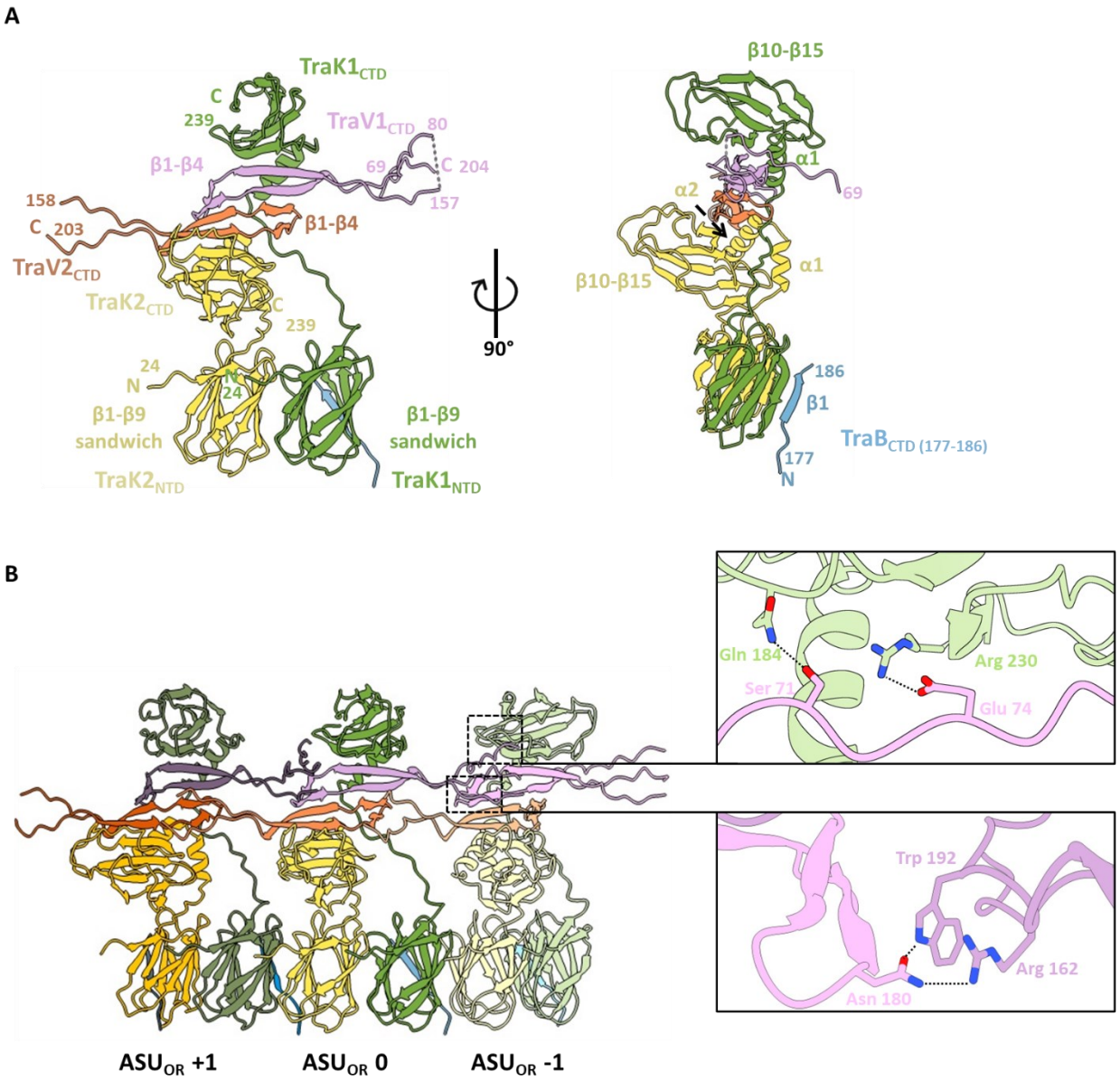

Figure 4.4: The asymmetric unit spanning the outer ring (ASU_{OR}). (A) The ribbon representation of the polypeptides that form the ASU_{OR}. The outer ring is formed from: TraK1 (green), TraK2 (yellow), $TraV1_{CTD}$ (purple), $TraV2_{CTD}$ (orange) and $TraB_{CTD\ (177-186)}$ (blue). The secondary structural features are labelled. **(B)** Three ASU_{OR} showing contacts between the units and particular focus on the protein-protein interactions. Interactions between the TraV short linker (residues 60 - 80) and $TraK1_{CTD}$ was observed which occurs via salt bridge formation between Glu74 and Arg230, but also by hydrogen bonding between Ser71 and Gln184 (inset top left). Interactions between the TraV chains further stabilise the outer ring which includes hydrogen bonding between Asn180 and Trp192 and hydrophobic contacts between Asn180 and Arg162 (inset bottom left).

4.3.3 The symmetry mismatch is accommodated by TraV and TraB which allows flexible contacts between the two F-OMCC rings

The C13:C17 mismatch within the F-OMCC is accommodated by TraB and TraV which traverses across both the outer and inner ring. While exact linkages could not be resolved due to regions of the polypeptides being untraceable, the structure does provide interesting insight into how the symmetry mismatch is enabled. TraB is the major constituent of $OMCC_{IR}$ where 17 chains structure the ring. From the 17 chains, 13 are seen to form a short β-strand (residues 177 – 186) in the inner wall of the outer ring formed by the $TraK_{NTD}$ (Figure 4.5A). Interestingly, it appears that the 13 TraB β-strands associates with the NTD from TraK1 rather than the TraK2 in the assembled $OMCC_{OR}$. This indicates that the TraB β-strand contacts one $TraK_{NTD}$ monomer preferentially or with a higher affinity than the second, and results in a periodic organisation. Nonetheless, the remaining four TraB polypeptides were not traceable in the outer ring. In the assembled F-OMCC, 26 copies of TraV form a part of the outer ring via their CTD by lying within the space created by the two $TraK_{CTD}$. From here, NTDs associated with TraV extend into the inner ring and bind along the TraB β-barrel domains. While majority of the linker which connects the TraV in the $OMCC_{IR}$ to that in the $OMCC_{OR}$ was not traceable, a short sequence (residues 69 - 80) was found to associate with $TraK1_{CTD}$. This $TraV_{69\text{-}80}$ sequence is positioned such that it is proximal to the $TraV_{NTD}$. In the structure, only 13 $TraV_{69\text{-}80}$ sequences were visualised and each of these extended from the $TraV1_{CTD}$, which shapes the top portion of the TraV belt at the outer ring (Figure 4.5B). Therefore, this suggests that 13 of the 17 TraV sites at the inner ring are from TraV1. Nevertheless, this resulted in 4 $TraV_{NTD}$ polypeptide chains in $OMCC_{IR}$ without an obvious connection to $OMCC_{OR}$. The 13 $TraV_{69\text{-}80}$ associated with TraV2 could not be detected in the structure and therefore it could be plausible that the remaining 4 $TraV_{NTD}$ is contributed by TraV2.

The TraB and TraV also facilitate flexible contacts between the two F-OMCC rings. To confirm the symmetry mismatch within the complex, a C1 map was generated without application of any symmetry. The unsymmetrised map showed clear C13 and C17 symmetry for the outer and inner rings respectively (Figure 4.6). When the atomic model of $OMCC_{IR}$ and $OMCC_{OR}$ was docked into the C1 map, most parts of the model fit in well apart from radial segments at the interface of the two rings where there was lack of density. This occurs at two locations in the complex, at the I- and O-layer. At the I-layer, the two F-OMCC rings are bridged by a proline rich linker (residues 187- 194) which connects the TraB β1 and β2 helices. In the C1 map, a continuous density can be seen corresponding to the proline rich linker for a few of the protomers, however, most of the linker connections is missing. This absence in density indicates that some

degree of flexibility is likely to exist between the outer and inner ring and appears to be mediated by this I-layer linker. Another region where signs of flexibility was observed is at the O-layer where a flexible sequence of 14 amino acids (residues 55 – 68) between TraV α1 and $TraV_{69-80}$ joins the $OMCC_{IR}$ to the $OMCC_{OR}$. The density corresponding to this sequence was absent in the C1 map indicating the presence of flexibility. This flexibility also extends to the TraV α1 helix. In the C1 map, some of the protomers appeared to exhibit varying degrees of electron density attributing to the α1, suggesting potential rotational movement between the two rings. Together, this highlights that flexibility exists within the complex and it is controlled at two independent locations.

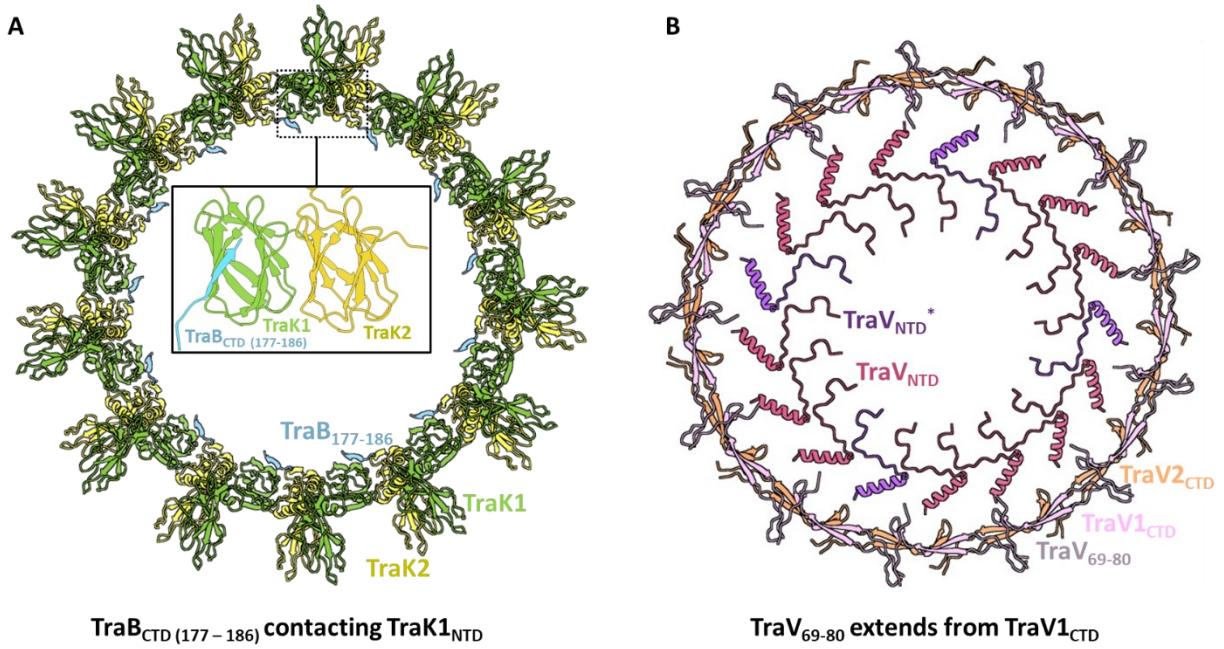

Figure 4.5: The symmetry mismatch is accommodated by TraB and TraV. (A) In the complete $OMCC_{OR}$, 13 $TraB_{177-186}$ (blue) contact the N-terminal domain (NTD) from TraK1 (green) as opposed to TraK2 (yellow). **(B)** A short TraV sequence ($TraV_{69-80}$) was found associated with $TraV1_{CTD}$ (light purple) which is positioned proximal to $TraV_{NTD}$ and acts as the connection between the two rings. Therefore, 13 of the 17 $TraV_{NTD}$ (magenta) are formed from TraV1 which leave 4 polypeptides without an obvious link to the outer ring ($TraV_{NTD}$*, dark purple).

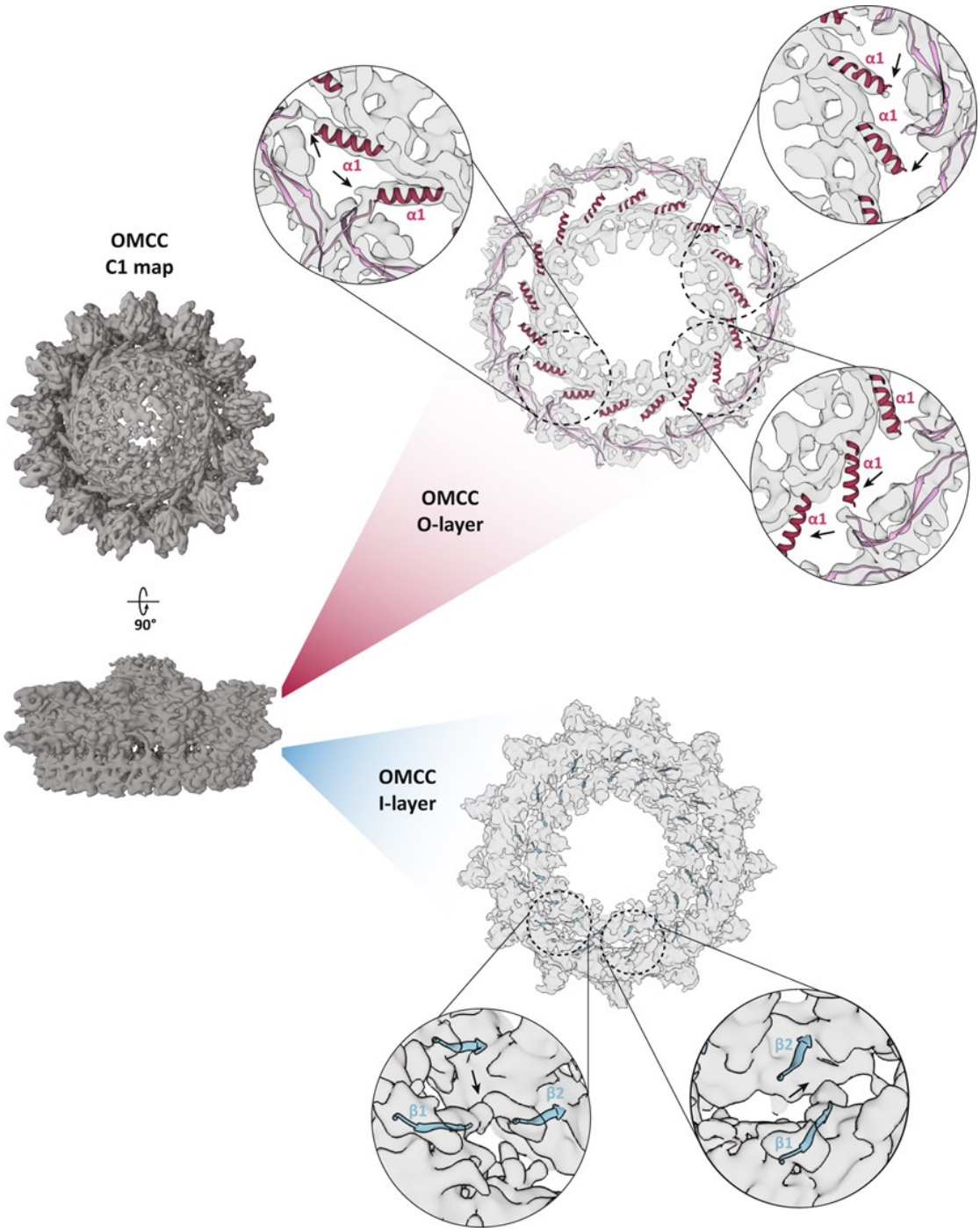

Figure 4.6: TraB and TraV facilitate flexible contacts between the F-OMCC rings. The C1 map generated confirms the C13:C17 symmetry. Flexibility was observed at two prominent areas of the complex which includes the O- and I-layer. The horizontal slice through the C1 map at the O-layer indicates the flexible contacts observed between the $TraV1_{69\text{-}80}$ linker and TraV α1 helix at the outer and inner rings respectively which is highlighted by the associated insets. The horizontal slice through the C1 map at the I-layer indicates the flexible connections between TraB β1 and β2 located the outer and inner rings respectively which is further shown in the insets. The arrows denote varying levels of electron density. This figure has been reproduced from (Amin et al., 2021).

4.4 Discussion

The structure of the conjugative outer membrane core complex encoded by the InFV plasmid pED208 was solved. The atomic model of the complete F-OMCC is made up from TraB, TraK and TraV which are homologues of VirB10, VirB9 and VirB7 respectively. A symmetry mismatch existed within the complex, and it provides interesting insight into how the complex may be involved in conjugation.

4.4.1 Comparison of the F-OMCC with the other OMCCs

The C13:C17 symmetry mismatch between the rings of the complex was surprising. This was not alluded to in the cryo-ET structure which is likely to be due to the low resolution of the map along with the application of C13 symmetry during the reconstruction (Hu et al., 2019a). Nevertheless, the existence of variable symmetries within a OMCC is not uncommon. It has been observed in the core complexes of the T4SSs involved in protein translocation, namely the *L. pneumophila* Dot/Icm core complex with C13:C18 symmetries and the *H. pylori* Cag OMCC which has C14:C17 symmetry (Figure 1.5) (Chung et al., 2019, Durie et al., 2020, Sheedlo et al., 2020, Sheedlo et al., 2021). At the time this study was carried out, the F-OMCC was the first conjugative T4SS OMCC to adopt two different symmetries. However, since then, the core complex encoded by the conjugative R388 plasmid, was found to also contain a symmetry mismatch (Figure 1.5) (Mace et al., 2022). Although, the number of solved T4SS core complexes that show variable symmetries is growing increasingly higher, there are still some which show a homogenous symmetry such as the equivalent complex in the *X. citri* minimised VirB/D4 machinery (Sgro et al., 2018). This highlights the great diversity that exists within the core complexes of T4SSs. The reasoning for the existence of a symmetry mismatch and what it attributes functionally is still unclear. However, it has been suggested that it create somewhat of a ratcheting effect which aids the transport of the substrate in a unidirectional manner through the central pore (Sheedlo et al., 2022). Furthermore, this phenomenon is not restricted to the T4SSs, but has also been reported in several other bacterial secretion systems including T2SS, T3SS, T6SS, the flagellar motor and the F_1F_0-ATPase (Worrall et al., 2016, Dix et al., 2018, Chernyatina and Low, 2019, Sobti et al., 2020, Takekawa et al., 2021). Therefore, the existence of symmetry mismatch within these diverse machineries implies that it must be of significance.

At the structural level, the arrangement of TraB and TraK is reminiscent of their counterparts from other T4SS (Figure 4.7). TraB forms the interior wall and spans across the entirety of the complex. At the inner

ring, the $TraB_{CTD}$ moulds the exit channel and at the outer ring, a short β strand (residues 177 – 186) engages in contact with a $TraK_{NTD}$ monomer. This organisation of TraB is conserved amongst its counterparts and has been seen in all T4SS core complexes solved to date (Figure 4.7A). In addition, the fold of TraB at the exit channel appears to be conserved with all the solved TraB counterparts also displaying a β barrel structure with two antiparallel α helices (Skurray et al., 1974, Chandran et al., 2009, Sgro et al., 2018, Chung et al., 2019, Durie et al., 2020, Sheedlo et al., 2020, Sheedlo et al., 2021, Mace et al., 2022). This highlights a structural conservation amongst the T4SS core complexes.

The presence of density corresponding to TraB at the inner wall of the OMCC I-layer contacting TraK was not unexpected since it has been previously alluded to and appears to be conserved. Also, an interaction between TraB and TraK has been reported in previously solved T4SS core complexes. However, there seems to be two prevailing ways in which this contact is mediated by TraB with the fold the protein adopts at this location differing. For F-OMCC, along with the Dot/Icm T4SS OMCC, TraB and its Dot/Icm homologue DotG, interact with the cognate $TraK_{NTD}$/$DotH_{NTD}$ via a β-strand (Durie et al., 2020, Sheedlo et al., 2021). However, in the core complex structures from *X. citri* and the R388 plasmid, this interaction occurs via an α helix of VirB10/TrwE (Sgro et al., 2018, Mace et al., 2022). Whether this detail bears any significance is yet to be unravelled. Also, the interaction between TraB and TraK was previously observed by yeast two-hybrid studies which demonstrated that only the TraB segment incorporating the β1 residues is capable of associating with the β4 of TraK (Harris et al., 2001).

There is also structural conservation within TraK, with the NTD globular structure of TraK being highly conserved. It shares the highest degree of structural homology with its homologues from pKM101 ($TraO_{CTD}$) and *X. citri* ($VirB9_{NTD}$) (Figure 4.7B) (Chandran et al., 2009, Sgro et al., 2018). Despite similarities with other T4SSs, F-OMCC also displayed unique features. The presence of two TraK molecules per ASU_{OR} which adopts two different arrangements is novel. Furthermore, the structure of TraV seems to be unique and does not bear any structural homology with any of the solved OMCC proteins. Thus, supporting the inability to predict the TraV structure in Chapter 3.

Together, the arrangement of proteins within the F-OMCC contributes to a remarkable stoichiometry that is unique. TraB, TraK and TraV are related with a stoichiometry of 1:1.5:1.5 respectively with a total protein number of 17:26:26 shaping the entire 2.1 MDa complex. While this can be considered to be unusual, the stoichiometry can be justified by the symmetry mismatch between the $OMCC_{IR}$ and $OMCC_{OR}$. At the inner ring, 1:1 ratio of TraB and TraV exists per ASU to form the C17 ring, whereas, at the outer ring, a ratio of 1:2:2 or TraB, TraK and TraV is present per ASU to shape the C13 ring.

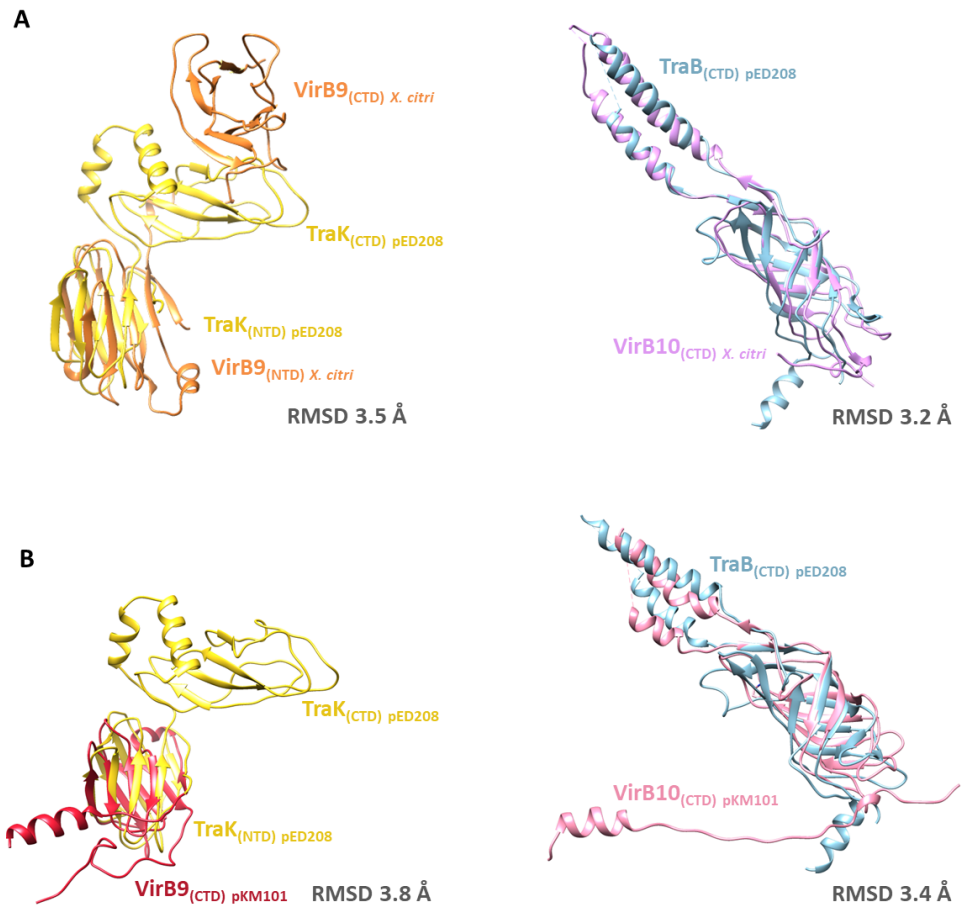

Figure 4.7: Comparing TraB and TraK with homologues from pKM101 and *X. citri*. (A) Comparison of TraK (yellow) with *X. citri* VirB9 (orange), and TraB (blue) with *X. citri* VirB10 (purple). **(B)** Comparison of TraK (yellow) with VirB9 encoded by pKM101 (red), and TraB (blue) with pKM101 VirB10 (pink).

4.4.2 The F-OMCC is a dynamic complex

The symmetry mismatch within the F-OMCC is accommodated by TraB and TraV which span across both the outer and inner rings. Out of the 17 TraB, 13 extend down to the $OMCC_{OR}$ where it makes contact with $TraK1_{NTD}$. This leaves 4 $TraB_{CTD}$ proteins at the inner ring without an obvious connection to the outer ring. It could be plausible that these 4 TraB interacts with $TraK2_{NTD}$ but occurs through a weaker association and could explain why it could not be captured by the cryo-EM structure. This notion is supported by the work carried out by Liu and colleagues where stronger densities for $TraB_{176\text{-}186}$ were observed with TraK1 than with TraK2 (Liu et al., 2022). This, together with the fact that the interaction between TraB β1 and

TraK β4 is mediated solely by the main chain interactions, could suggest that these weak contacts may facilitate movement of the proteins. This is reinforced by signs of flexibility at the O-layer, where all but a few TraB protomers lacked a continuous density corresponding to the proline rich linker that connects the β1 and β2 strands. This indicates that the proteins experience variable levels of flexibility with the lack of density possibly being attributed to movement at the O-layer, specifically translational movement along the $OMCC_{OR}$, or by a "sliding" mechanism that alters the protein register to allow the rotational movement of the F-OMCC.

It is perhaps not surprising that one of the key proteins that mediates the symmetry mismatch and governs the flexibility within the complex is TraB since its counterpart in the VirB/D system, VirB10, has been reported to be a dynamic modulator of pilus assembly and substrate secretion. VirB10 and its homologues span across both the outer and inner membrane and is central to the T4SS machinery. For VirB10, it has been shown that it acts as a sensor protein to coordinate the conformational change from pilus biogenesis to substrate secretion (Jakubowski et al., 2009, Darbari et al., 2020). Therefore, it is plausible that TraB may also be involved in a similar role, by relaying signals that result in structural rearrangement causing the opening of the exit channel formed by the TraB α helices, potentially due to movement of the loops connecting TraB α1 and α2, and thus, allowing conjugation to proceed.

Another bridging protein between the two rings is TraV. A flexible stretch of sequence (residues 55 – 68) connects the TraV α1 in the inner ring with the $TraV1_{CTD}$ linker in the outer ring. This radial connection provides flexibility at the I-layer. Furthermore, the TraV protomers also exhibit variable degrees of electron density corresponding to the α1 helix, which is indicative of rotational movement. TraV is a particularly interesting protein as it plays a central role in scaffolding and stabilising the complex by creating a large mesh over the TraB helices at the inner ring and forming a continuous belt across the $TraK_{CTD}$ at the outer ring. Vital to the formation of the mesh, and therefore the integrity of the F-OMCC, is the network of di-sulphide bonds formed between Cys27 and Cys35. These cysteine residues are important for the efficient localisation of TraV to the outer membrane (Harris and Silverman, 2002). TraV has also been postulated to be involved in forming an interaction platform with other Tra/Trb proteins to the F-OMCC, namely the proteins TraH, TraF, TraU,TraW, TrbC and TrbI (Arutyunov et al., 2010, Koch et al., 2020). In accordance with this, a recent model for the F-T4SS assembly and dynamics, has proposed that TraF, TraH, and TraW interact with the F-OMCC to orchestrate structural rearrangements (Kishida et al., 2022). Therefore, it could be plausible that the presence of these additional proteins may aid the

stabilisation of the regions of TraV that could not be assigned a sequence, and further supports the notion that the F-OMCC is a dynamic complex.

Overall, the flexibility observed within the F-OMCC appears to be controlled at two independent locations (O-layer and I-layer) at the specific regions where the two rings are connected. This suggests that the site of symmetry mismatch provides regions of mobility between $OMCC_{OR}$ and $OMCC_{IR}$. Furthermore, given that the structure represents the complex in the closed state, the flexibility provides insight into how the complex may undergo structural rearrangements during pilus extension and retraction, and ultimately into how DNA transfer occurs during conjugation.

4.4.3 Future perspective

The F-OMCC seems to represent a hybrid of the complexes from the minimised and expanded T4SS. Like the equivalent complex from the minimised T4SS, the F-OMCC is composed from three proteins, but reminiscent of the expanded systems, it shows an intricate network of interactions which occurs across two distinct symmetries in an elaborate manner. Given that compared to the expanded T4SS core complexes, the F-OMCC is compositionally simpler, it makes it a suitable tool for future investigations into the biological basis and significance of symmetry mismatch. Furthermore, the flexibility observed within the complex provides exciting new insight into the dynamics of pilus extension and retraction, and how conjugation proceeds. Thereby, this work paves the way for future investigations into deciphering the mechanistic details of how DNA transfer occurs. One such avenue that would be interesting to explore is if the F-OMCC interacts with other Tra/Trb proteins to mediate the pilus biomechanics, as this will give a more insight into the detailed steps that occur for conjugation to proceed. From a broader medical perspective, the F-OMCC structure could set the stage for the development of therapeutics such as small inhibitor compounds to hinder the proliferation of antibiotic resistance.

Chapter 5: Investigation into the interactions of F-OMCC and other Tra/Trb proteins

5.1 Introduction

Along with the three proteins that form the F-OMCC, another interaction group of proteins has been identified by yeast two-hybrid studies (Harris and Silverman, 2004). The group consists of the periplasmic proteins TraF, -H, -U, -W, TrbC and the inner membrane protein TrbI and hereinafter will be referred to as the periplasmic complex (Figure 5.1). Collectively these proteins act in the assembly and retraction of the pilus and are unique to the F and F-like T4SS. There have been several lines of evidence that suggest that members of this second interaction group may engage with the F-OMCC. The first evidence for a potential interaction came from a yeast two-hybrid assay which detected an interaction between TraH and TraB (Harris and Silverman, 2004). This alludes to a potential interaction occurring between the periplasmic complex and the F-OMCC mediated by TraH. Furthermore, a recent bacterial two-hybrid study demonstrated a network of interactions occurring between various proteins in the periplasmic complex with the F-OMCC proteins in such a manner that all three core complex proteins were participants (Figure 5.1). This is plausible given the central role of the core complex within the conjugative apparatus and as a scaffold for the pilus to extend through. This is supported by the high-resolution structure described in Chapter 4. There were regions in TraV that could not be assigned a sequence due to inherent flexibility. These regions may be stabilised in the presence of the potential interactions that have been hypothesised. If the members of the periplasmic complex did indeed interact with the F-OMCC, it could provide exciting insight into how the pilus action is coordinated with the core structural machinery during conjugation.

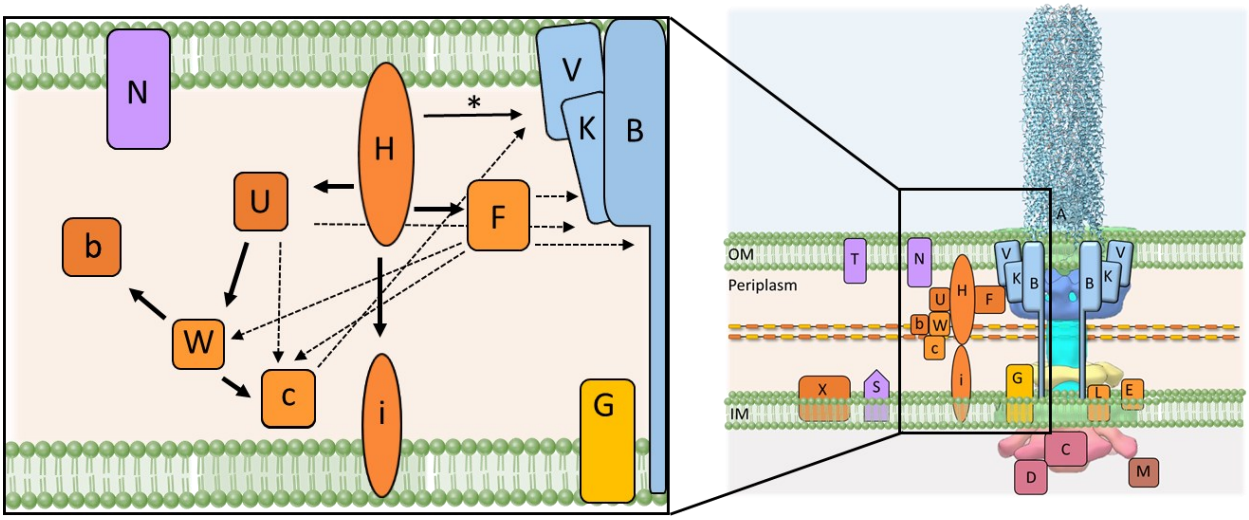

Figure 5.1: A schematic highlighting the interactions observed between the *E. coli* F plasmid encoded Tra/Trb proteins. The thick arrows denote the interactions that were identified from a yeast two-hybrid study (Harris and Silverman, 2004). This group of proteins are referred to as the periplasmic complex. The asterisk represents an interaction that was seen with low positive hits. The dashed arrows indicate the interactions that were noted from a bacteria two-hybrid assay (Koch et al., 2020). The proteins with a capital letter have the prefix Tra whereas those with lower case letters have the prefix Trb.

5.1.1 TraW and TrbC

Two proteins that have been suggested to interact with the F-OMCC are TraW and TrbC. TraW plays a role in pilus extension as demonstrated by Anthony et al using the M13K07 transducing phage assay where mutations in TraW prevented pilus outgrowth but did not affect pilus tip formation (Anthony et al., 1999). A previous study using electron microscopy showed that F *traW546* amber mutant resulted in abrogated pilus formation (Maneewannakul et al., 1992). This highlights the importance of TraW within T4SS function and conjugation. Resembling TraW, TrbC localises to the periplasm and is involved in pilus function (Maneewannakul et al., 1991). In some F plasmid systems, these two proteins are expressed as a single polypeptide with TrbC being fused to the N-terminus of TraW (Lawley et al., 2003). The two proteins were found to form a dimeric complex in solution, and this interaction was functionally relevant as it was required for conjugation to occur (Shala-Lawrence et al., 2018). Together, this indicates the two proteins have cooperative function. TraW and TrbC were suggested to interact with the F-OMCC by bacterial two-hybrid assay which showed that TrbC interacts with TraK and TraV. The TraV and TrbC interaction was further shown *in vitro* as they co-purified together (Koch et al., 2020). A similar interaction was characterised between the TraW and F-OMCC from the *N. gonorrhea* T4SS (Koch et al., 2020). As the

two proteins was independently expressed in previous studies, it will be interesting to explore how this interaction occurs in the context of the TraW-TrbC dimer and the F-OMCC, and if a larger interacting group of proteins can be isolated.

5.1.2 TraH and TraF

TraH and TraF are involved in pilus extension (Anthony et al., 1999). While TraF is found in the periplasm, TraH is thought to be peripherally associated with the outer membrane (Arutyunov et al., 2010). Among the proteins that form the periplasmic complex, TraH is particularly notable due to the results from the yeast two-hybrid studies suggesting a high number of interacting partners (Harris and Silverman, 2004). As a result, TraH was positioned at the centre of the interaction network that forms the periplasmic complex and was directly linked to TraF, TraU and TrbI (Figure 5.1). The C-terminal region of TraH contains a coiled coil domain which has been speculated to aid oligomerisation and facilitate interaction with TraF (Lawley et al., 2003). Yeast two-hybrid analyses also showed that the interaction with TraW and TrbI was localized to the C-terminal portion of TraH whereas the N-terminal segment of the protein was associated with TrbI (Harris and Silverman, 2004). Through bacterial two-hybrid analyses, TraF was found to interact with TrbC, TraW and members of the F-OMCC, specifically TraK and TraB (Koch et al., 2020). A putative interaction between TraH and TraB has been observed by yeast two-hybrid experiments (Harris and Silverman, 2004). Furthermore, a recent model based on deletion mutant analyses has proposed that TraH, TraF and TraW binds to the F-OMCC and thereby inducing a conformational transition to a T4SS adept for pilus biogenesis upon target cell contact (Kishida et al., 2022). Nevertheless, this interaction between TraH/TraF with the core complex remains to be confirmed and studied by *in vitro* techniques such as pull-down and co-purification.

5.1.3 TraN

TraN does not form part of the periplasmic complex but may interact with the core complex. TraN is an integral outer membrane protein and is thought to have 20 transmembrane domains (Klimke et al., 2005). It is a key protein in Mps which allows the donor and recipient cells to mate more successfully by protecting against shear forces and addition of harsh chemicals (Achtman et al., 1978, Manning et al., 1981). The mating aggregates are formed as a result of TraN in the donor cell recognising the cognate outer membrane protein (Omp) receptor in the recipient cell in a plasmid specific manner (Skurray et al.,

1974, Klimke and Frost, 1998). In the prototypical F plasmid, the receptor is OmpA and the specificity was mapped to the region of TraN formed from amino acid residue 164 to 333 (Skurray et al., 1974, Manoil and Rosenbusch, 1982, Klimke and Frost, 1998). The determinant of specificity was further narrowed down to the tip region of TraN, with structurally close tips recognising the same outer membrane receptor in the recipient cell (Low et al., 2022). The C-terminus of TraN is cysteine rich and is important for TraN function (Klimke et al., 2005).

In terms of interactions, along with the cognate Omp receptor, TraN has been suggested to interact with the inner membrane protein TraG, and potentially TraU. However, the latter two interactions are yet to be confirmed experimentally. Aside from these interactions, a model presented by the Christie lab proposed that TraN, together with TrbC and TraU, is recruited to the F-OMCC to form an active pilus assembly factory that controls the dynamic extension and retraction of the filament (Kishida et al., 2022). Given the cellular location of TraN, such an interaction with the F-OMCC may be possible, however, this remains to be verified.

5.2 Aim of this study

This Chapter aimed to investigate the interaction of the F-OMCC with other Tra/TrbI proteins namely TraW, TrbC, TraH and TraN. Pull-down assays were employed to explore if the proteins interact coupled with SDS-PAGE and western blot. Simultaneously to this, efforts were made to isolate the periplasmic complex using affinity purification and gel filtration.

5.3 Results

5.3.1 The TraW and TrbC interaction with F-OMCC could not be confirmed confidently

To investigate the interaction of TraW and TrbC with F-OMCC, three constructs were engineered: pASK_TraW^{His}, pASK_TrbC^{His} and pASK_TraW^{His}_TrbC. The hexahistidine tags were added at the C-terminus due to the presence of a signal peptide at the N-terminus (Teufel et al., 2022) (Figure 5.2). The pASK_TraW^{His}_TrbC was designed such that the TraW would be tagged. The rationale behind this is that it has been demonstrated that TrbC exists in a mixture of oligomeric states in solution, however, in complex with TraW, the protein is stabilised in a TraW-TrbC dimer (Shala-Lawrence et al., 2018). Therefore, by tagging TraW, this dimer of TraW-TrbC would be purified and tested for interaction with the F-OMCC. The plasmids expressing the individual proteins were constructed to allow for pull-down assays of the individual proteins with the F-OMCC.

As the two plasmids are compatible, the pASK_TraW^{His}_TrbC and the pBAD_TraK_TraB^{Strep}_TraV were co-transformed into the same TOP10 cells, and then, plated onto agar plates enriched with both spectinomycin and carbenicillin (Figure 5.3). The pull-down was performed in this manner to ensure the proteins were within the same bacterial membranes. Therefore, if an interaction occurs this should not be disrupted by the detergent shell during extraction from the membranes. The cells were grown, lysed and membranes solubilised as described in section 2.3.6. Two independent pull-downs were performed using either the StrepTrap column where the F-OMCC is used as the bait or by HisTrap column where the TraW is immobilised on the column (Figure 5.3). At this point, it is necessary to highlight that at the time this work was carried out, the ability to confirm presence of TrbC was limited. Hence, only the involvement of TraW could be accurately investigated when using the pASK_TraW^{His}_TrbC plasmid and therefore, will be commented on.

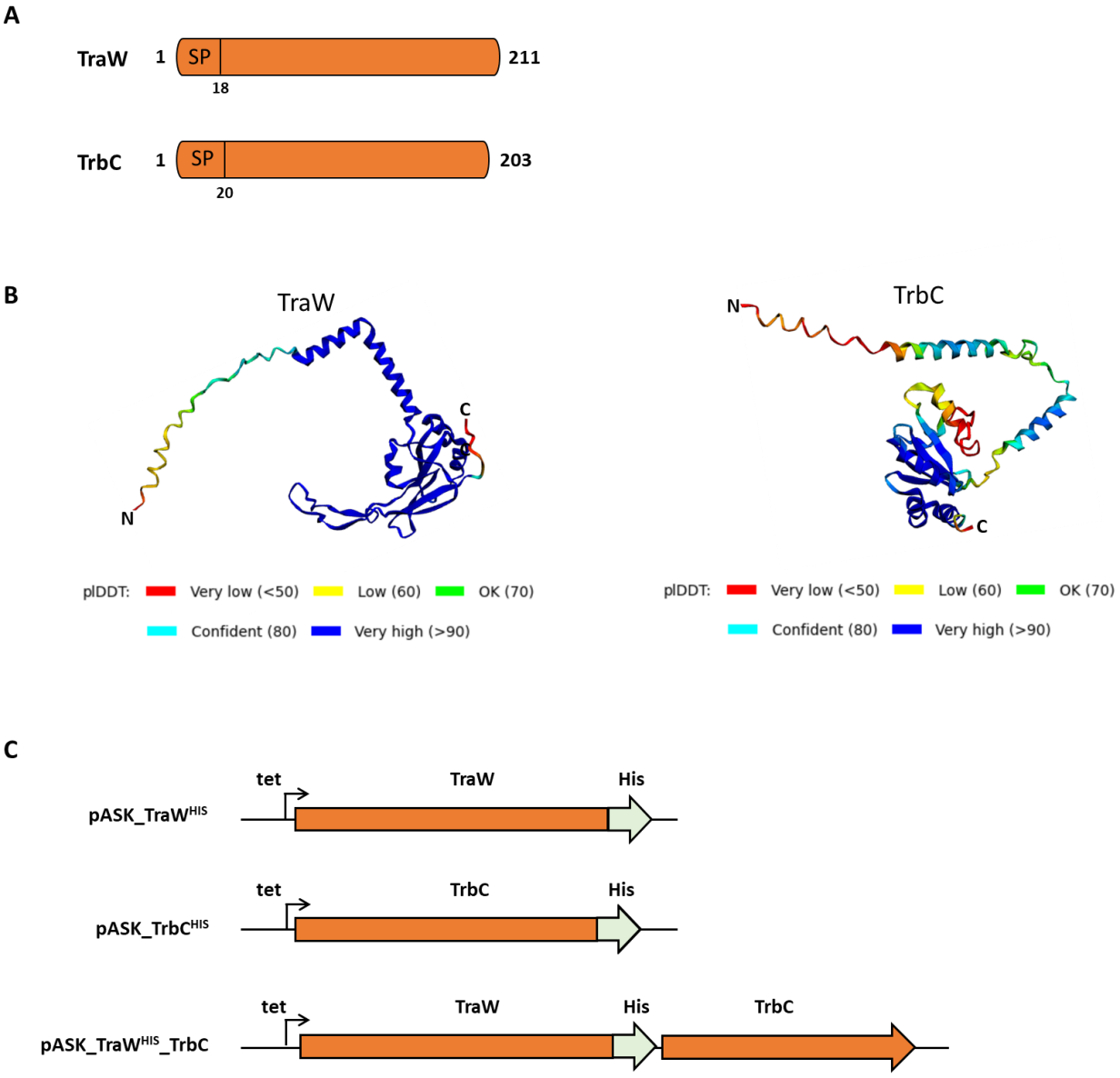

Figure 5.2: The bioinformatic analyses of TraW and TrbC. (A) The predicted domains of TraW and TrbC which indicate a signal peptide (SP) at the N-terminus. **(B)** The AlphaFold2 predicted structures of the two proteins coloured according to the pIDDT confidence. **(C)** The diagram of the operon structure of the plasmids used to express the two proteins under the anhydrotetracycline (tet) promoter.

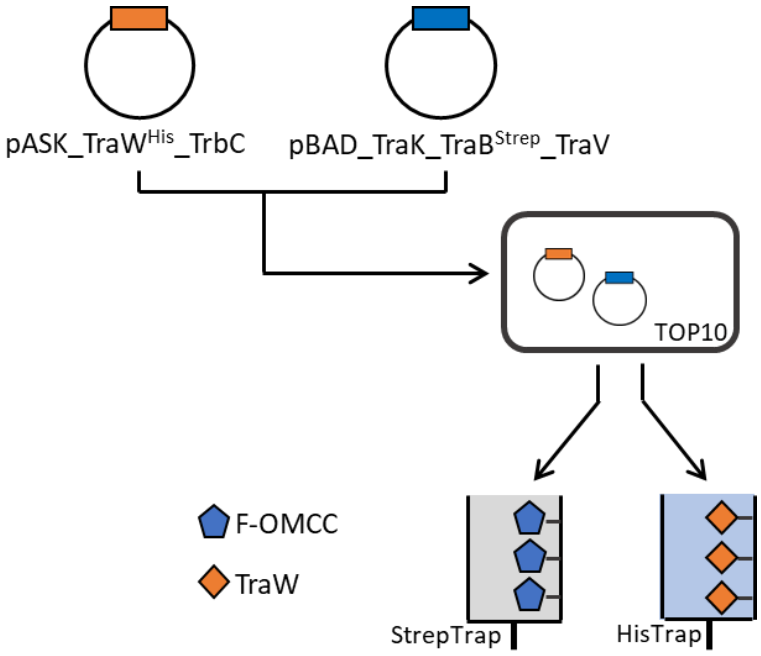

Figure 5.3: An overall schematic of the two independent pulldowns performed for TraW/TrbC and F-OMCC. TOP10 competent cells were transformed with both plasmids.

Regarding the StrepTrap assisted pulldown, an elution peak was observed upon addition of desthiobiotin (Figure 5.4A). The fractions were visualised by SDS-PAGE gel which showed the characteristic three bands corresponding to the F-OMCC (Figure 5.4B). The estimated molecular weight of TraW is 24 kDa which is similar to the size of TraK and could be masked by the band. As a result, based on the SDS-PAGE gel alone, it was difficult to decipher whether TraW had co-eluted with the F-OMCC. The His-tag assisted pull-down showed a prominent peak with an accompanying shoulder peak in the elution upon addition of approximately 400 mM imidazole (Figure 5.4C). The fractions corresponding to the peak were analysed by SDS-PAGE (Figure 5.4D). The gel showed immense banding, indicating that the hexahistidine tag was not effective in purifying the protein. To investigate if the F-OMCC and TraW were co-eluting in the elution peaks, western blots were carried out using anti-Strep and anti-His antibodies (Figure 5.5). The anti-His western blot, confirmed the presence of TraW in the elution peak of the HisTrap pull-down by the presence of a band at above 26 kDa. It appears TraW was predominately being eluted in the shoulder of the elution peak corresponding to fraction 23. The western blots showed that TraW did not co-elute with the F-OMCC as indicated by the absence of a band for the elution fraction 9 of the StrepTrap pulldown. It should be noted that the faint bands observed is due to a technical error during the process. Instead,

TraW appears to have just passed through the column as demonstrated by the prominent band in the StrepTrap flowthrough (FT) fraction and the rest was subsequently washed off.

On the contrary, the anti-Strep western blot indicated that the F-OMCC was co-eluting with TraW as shown by the TraB band in the HisTrap elution fractions 19 and 23. As a control, the ability of F-OMCC to bind to the HisTrap was probed. This showed that independently, F-OMCC does not bind to the Ni Sepharose and is lost in the flowthrough (Figure 5.6). Thus, if F-OMCC is eluting in the HisTrap fraction it is probably occurring through association with another protein.

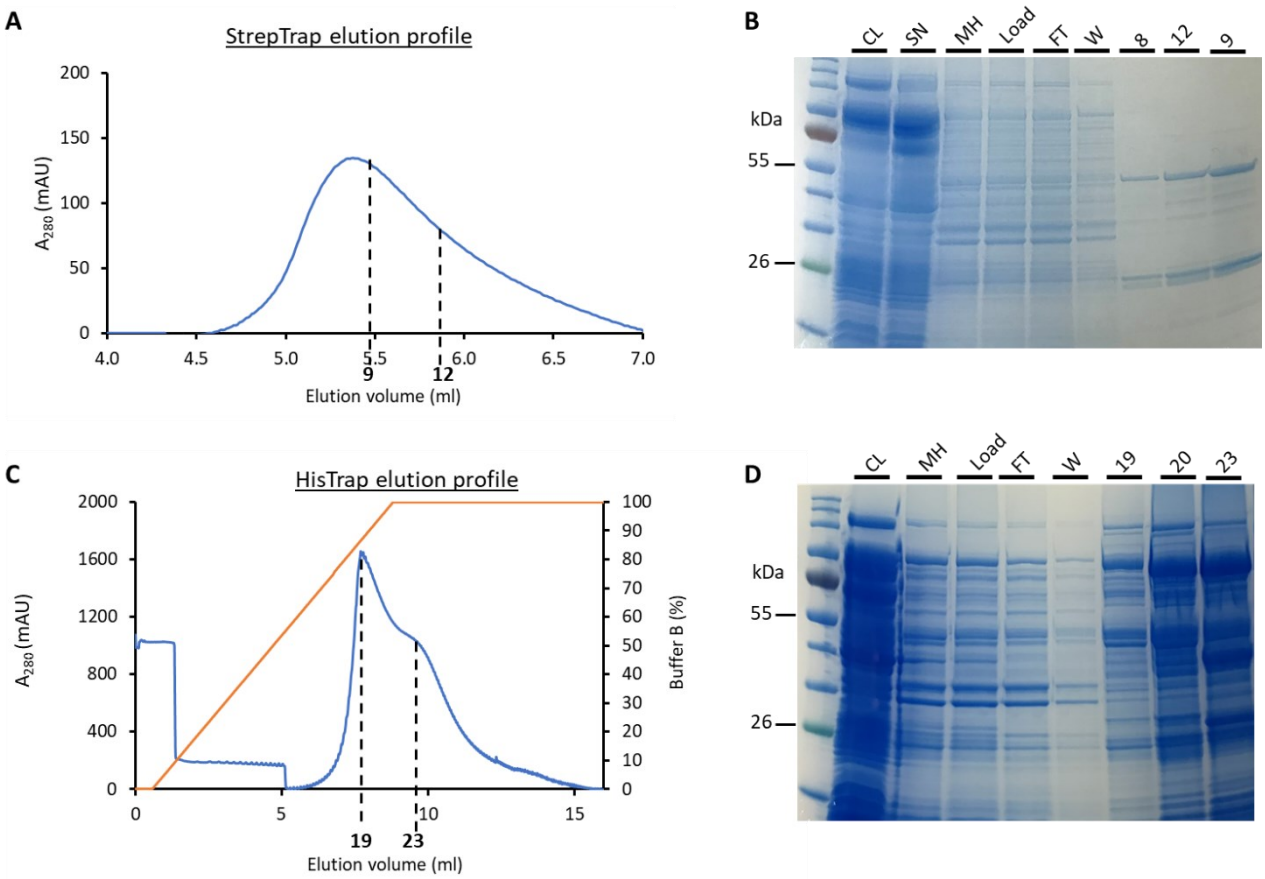

Figure 5.4: The pulldown of TraW/TrbC with the F-OMCC. (A) The StrepTrap elution profile for the pull-down with the fractions marked. **(B)** The SDS-PAGE gel for the StrepTrap pull-down. **(C)** The HisTrap elution profile for the pull-down with the fractions marked. The 100% buffer B represents 500 mM imidazole. **(D)** The SDS-PAGE gel for the HisTrap pull-down. CL (Clear lysate), SN (Supernatant), MH (Membrane homogenised), FT (Flow-through), W (Wash)

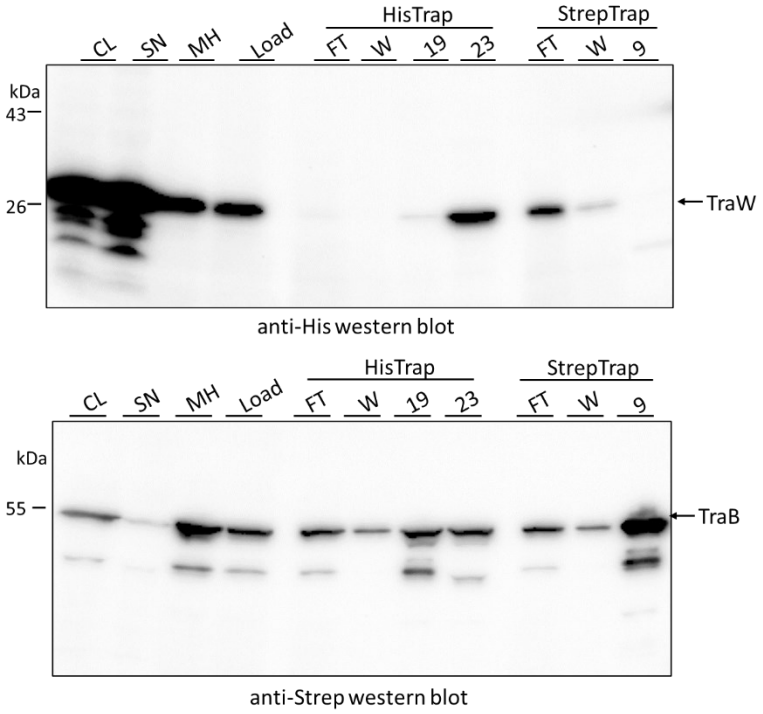

Figure 5.5: The western blot for the pull-down. The anti-His western blot displaying the presence of TraW in the HisTrap elutions. TraW, however, was not co-purifying in the StrepTrap fractions. The anti-Strep western blot showing the presence of TraB and therefore the F-OMCC. TraB was co-eluting in the HisTrap elution fractions. CL (Clear lysate), SN (Supernatant), MH (Membrane homogenised), FT (Flow-through), W (Wash)

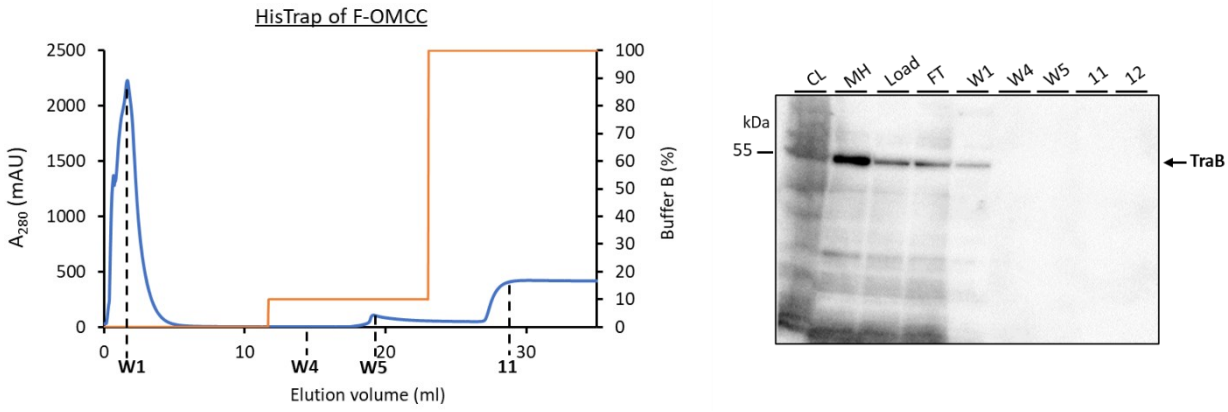

Figure 5.6: Investigating whether the F-OMCC binds to the HisTrap column. This control shows that the F-OMCC does not bind to the F-OMCC and is lost in the flowthrough (FT) and wash (W) steps. The western blot was performed with anti-Strep antibodies. CL (Clear lysate), SN (Supernatant), MH (Membrane homogenised)

The results of the pull-down indicated that a potential interaction was occurring between TraW and F-OMCC but could only be isolated by HisTrap and was lost when a StrepTrap was performed. A possible explanation could be that the interaction between TraW and the F-OMCC is not 1:1 and there might be fewer TraW to F-OMCC per interacting complex. Therefore, the TraW may need to be concentrated on the HisTrap column to enrich this interaction before isolating it. However, the HisTrap elution fractions were comparable to the load fraction on the SDS-PAGE gel, highlighting the lack of purification from contaminating proteins. Consequently, to purify the sample and to confirm that an interaction was occurring between TraW and F-OMCC, a double pull-down was carried out.

The double pull-down was set up as described in section 2.3.6 and such that the proteins would elute directly from the HisTrap column into the StrepTrap column (Figure 5.7). The HisTrap profile showed a small peak when 500 mM imidazole was added in a one-step elution (Figure 5.8A). This peak is the flowthrough since it was due to the protein which passed through the attached StrepTrap column without binding. The StrepTrap elution profile contained a small trailing peak (Figure 5.8B). The fractions from this double pulldown were analysed by both SDS-PAGE and western blotting. The SDS-PAGE gel showed the three F-OMCC bands in the StrepTrap elution fractions (Figure 5.8C). The anti-His and anti-Strep blots showed that the elution fractions contained the F-OMCC but not TraW (Figure 5.8D). The majority of TraW appears to have eluted in the flowthrough fraction 35, meaning that it elutes from the HisTrap column but did not bind the StrepTrap column. Surprisingly, a proportion of F-OMCC appears to have also eluted in the flowthrough indicated by the faint band in fraction 35 in the anti-Strep blot. While this may suggest an interaction, the SDS-PAGE analysis of fraction 35 indicates the presence of contaminating proteins by the immense banding. Therefore, this makes it difficult to decipher whether an interaction is occurring or if the two components were coincidentally present in fraction 35 along with a range of other contaminating proteins.

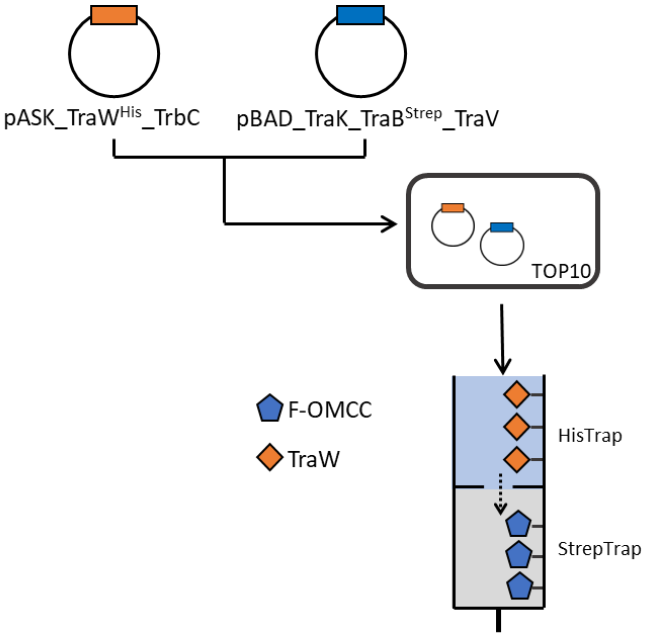

Figure 5.7: An overall schematic of the double pulldown performed for TraW/TrbC and F-OMCC. TOP10 competent cells were transformed with both plasmids.

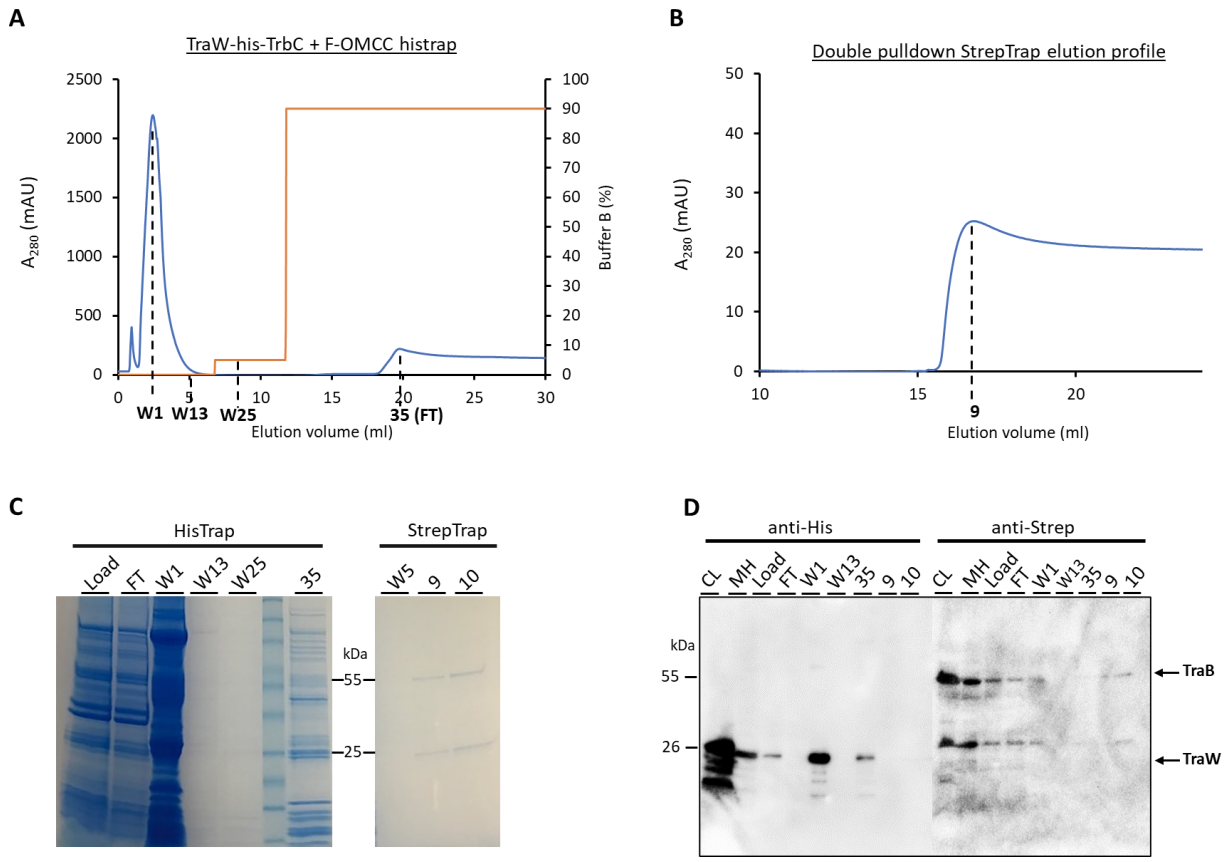

Figure 5.8: The double pull-down of TraW and TrbC with F-OMCC. (A) The HisTrap profile with the fractions marked. The Fraction 35 is actually the flow-through of the column since the StrepTrap column was attached directly below. **(B)** The StrepTrap elution profile with the fractions marked. **(C)** The SDS-PAGE gel for the double pull-down. **(D)** The western blot for the double pull-down showing that TraW and the F-OMCC did not co-elute. CL (Clear lysate), MH (Membrane homogenised), FT (Flow-through), W (Wash)

Currently, it is too premature to conclude whether TraW is or isn't interacting with the F-OMCC given the mixed results. Nevertheless, the results described here is preliminary and further work is required to firstly check for TrbC expression in this plasmid and to probe whether throughout the pull-down assays, the TrbC was forming a complex with TraW.

5.3.2 TraH, TraF and the periplasmic complex

Another group of proteins that have been demonstrated to interact are those that are involved in pilus assembly and retraction. Central to this group of interactions is TraH.

5.3.2.1 TraH may interact with F-OMCC

As TraH has been suggested to localise to the outer membrane, based on its location it was investigated whether TraH may interact with F-OMCC. An interaction between TraH and TraB has been observed by yeast two-hybrid studies (Harris and Silverman, 2004). Therefore, to establish if such an interaction was occurring and could be isolated *in vitro*, a construct was engineered alongside my postgraduate student Hayley Jackson. Here, TraH was cloned into the pBAD_TraK_TraB^{Strep}_TraV plasmid downstream of the *traV* gene, forming the construct pBAD_TraK_TraB^{Strep}_TraV_TraH^{His} (Figure 5.9). The TraH was tagged with a hexahistidine tag at the C-terminus due to SignalP-6.0 predicting a N-terminal signal peptide (Figure 5.9) (Teufel et al., 2022). The pulldown was carried out alongside my undergraduate student Matthew Tang using TraH as the bait and performing the experiment on a HisTrap column as described in section 2.3.6. The elution profile showed a small peak at approximately 100 mM imidazole concentration (Figure 5.10A). The fractions were analysed by western blotting (Figure 5.10B). The estimated molecular weight of TraH is 48 kDa. Both TraH and the F-OMCC, as indicated by TraB, were expressed. The presence of the faint bands in the elution fractions 16 and 19 in both the anti-His and anti-Strep western blots may imply that TraH interacts with the F-OMCC. Nevertheless, due to the low intensity of the bands, it cannot unequivocally be concluded that an interaction is occurring. Furthermore, there appears to be issues with the expressed TraH. A significant proportion of the protein is being lost in the wash and coupled with the low intensity of the elution peak; it indicates that TraH is binding weakly to the column. Therefore, the tagging of TraH required optimising.

A similar approach was taken to investigate whether TraF interacts with F-OMCC as this had been alluded to by bacterial two hybrid studies (Harris and Silverman, 2004). Together with Hayley Jackson, TraF with a C-terminal hexahistidine tag was cloned into the plasmid expressing the F-OMCC, yielding a construct termed pBAD_TraK_TraB^{Strep}_TraV_TraF^{His}. The HisTrap assisted pulldown was performed as described in section 2.3.6 alongside my undergraduate student Ambre Bexter. The elution profile and the accompanying western blot demonstrated issues with TraF expression and a lack of binding to the HisTrap column. Therefore, insight into TraF interactions with F-OMCC was limited by the expression and accessibility of the hexahistidine tag.

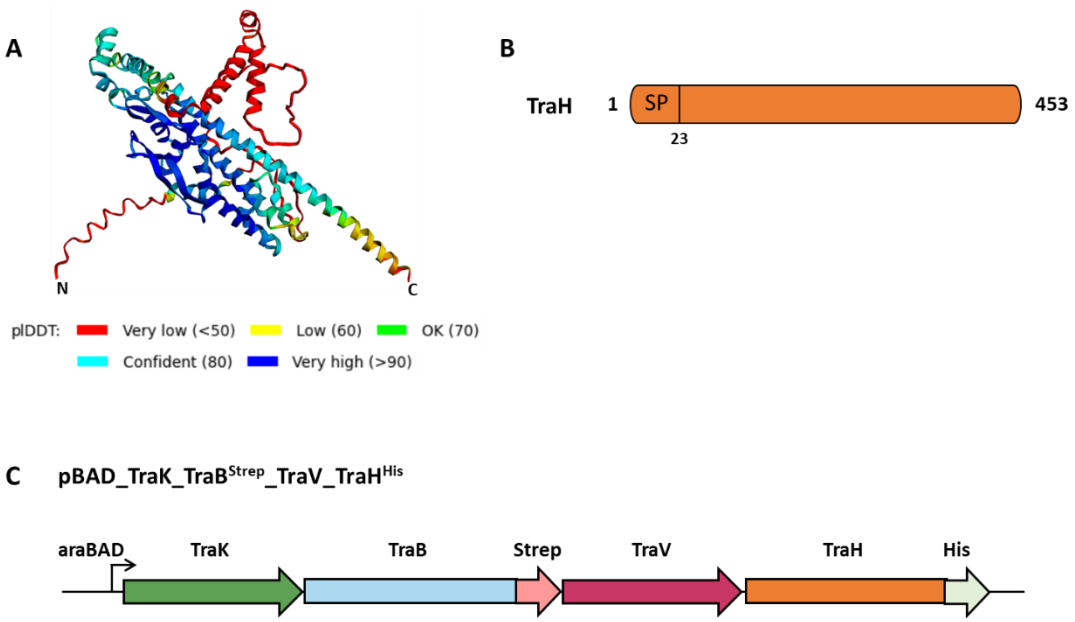

Figure 5.9: The TraH bioinformatic analyses. (A) The AlphaFold2 predicted model of TraH. **(B)** The domain prediction of TraH which shows that there is a N-terminal signal peptide. **(C)** The diagram of the operon structure of the plasmid used to express TraH along with the F-OMCC.

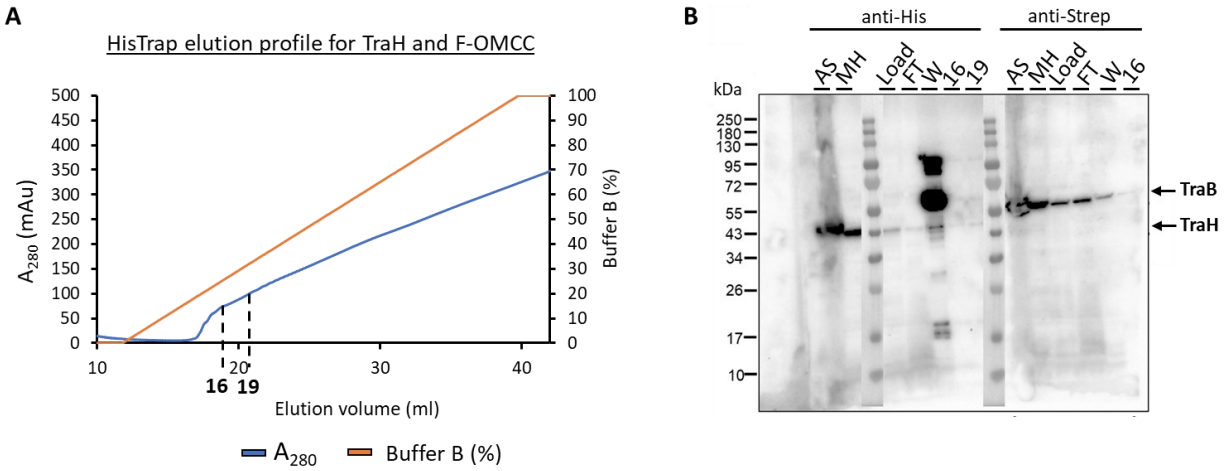

Figure 5.10: The TraH and F-OMCC pull-down. (A) The HisTrap elution of TraH and F-OMCC with the fractions marked. The 100% (v/v) buffer B correspond to 500 mM imidazole. **(B)** The western blot, done alongside Matthew Tang, for the pull-down where a faint band for TraH and TraB is observed in the elution fraction. AS (After sonication), MH (Membrane homogenised), FT (Flow-through), W (Wash)

5.3.2.2 The periplasmic complex

As TraH is involved in a myriad of interactions with other Tra/TrbI proteins to the form the periplasmic complex, it was speculated whether the expression of TraH could be improved in the presence of these proteins. One such significant interaction of TraH is with the inner membrane protein TrbI, whereby in a *trbI* background, a 40-fold reduction was observed in the level of TraH (Harris and Silverman, 2004). In the *Neisseria gonorrhoeae* T4SS, TraG was required for TraH to localise to the outer membrane (Koch et al., 2020). Furthermore, the interaction group described by Harris and Silverman may be indicative of a periplasmic complex that forms prior to interaction with F-OMCC (Harris and Silverman, 2004) . Therefore, to address this, a construct was engineered containing the *trbI* to *traG* stretch of the Tra operon. Efforts to clone this into a pASK vector which would enable co-transformation with the F-OMCC was hindered by issues with cloning. However, with assistance from Ambre Bexter, a postgraduate student in the laboratory, that stretch of the operon was cloned into a pBAD vector. This plasmid encodes the following genes: TrbI, TraW, TraU, TrbC, TraN, TraF, TrbB, TraH and TraG. To better the specificity of the His-tag during the purification, the hexahistidine tag at the C-terminus of TraH was extended to 10 histidines. A GSGSGG linker was introduced between the protein and the tag to improve accessibility, and to avoid issues with binding of the protein to the HisTrap column as seen in Figure 5.10.

The protein was expressed, and a purification was carried out as described in section 2.3.8 to test for improvement in TraH levels and binding to the affinity column, before pull-down experiments with F-OMCC were performed. The sample was first purified by HisTrap and a prominent elution peak was observed. The sample was further purified by gel filtration using a Superdex 200 GL 10/300 column. The profile showed multiple peaks (Figure 5.11A). The SDS-PAGE gel showed that the peaks contained various proteins, however, it was difficult to accurately identify the TraH band (Figure 5.11B). The anti-His western blot indicated that TraH was present in all the gel filtration peaks (Figure 5.11C). The full-length protein predominately eluted earlier in the run. The peaks formed towards the end of the run appear to mainly contain N-terminal degraded fragments of TraH. The sample within the peaks were visualised by negative stain electron microscopy. The sample from fraction 23 appeared interesting and showed ring like structures with an approximate diameter of 90 Å (Figure 5.12). Nevertheless, the purification of TraH and/or accompanying periplasmic complex was not conclusive which hindered interaction studies with F-OMCC. The purification was not improved when the His-tag was replaced by the more specific Strep-II tag

by my postgraduate student Ambre Bexter. Therefore, the expression and isolation of TraH requires optimising.

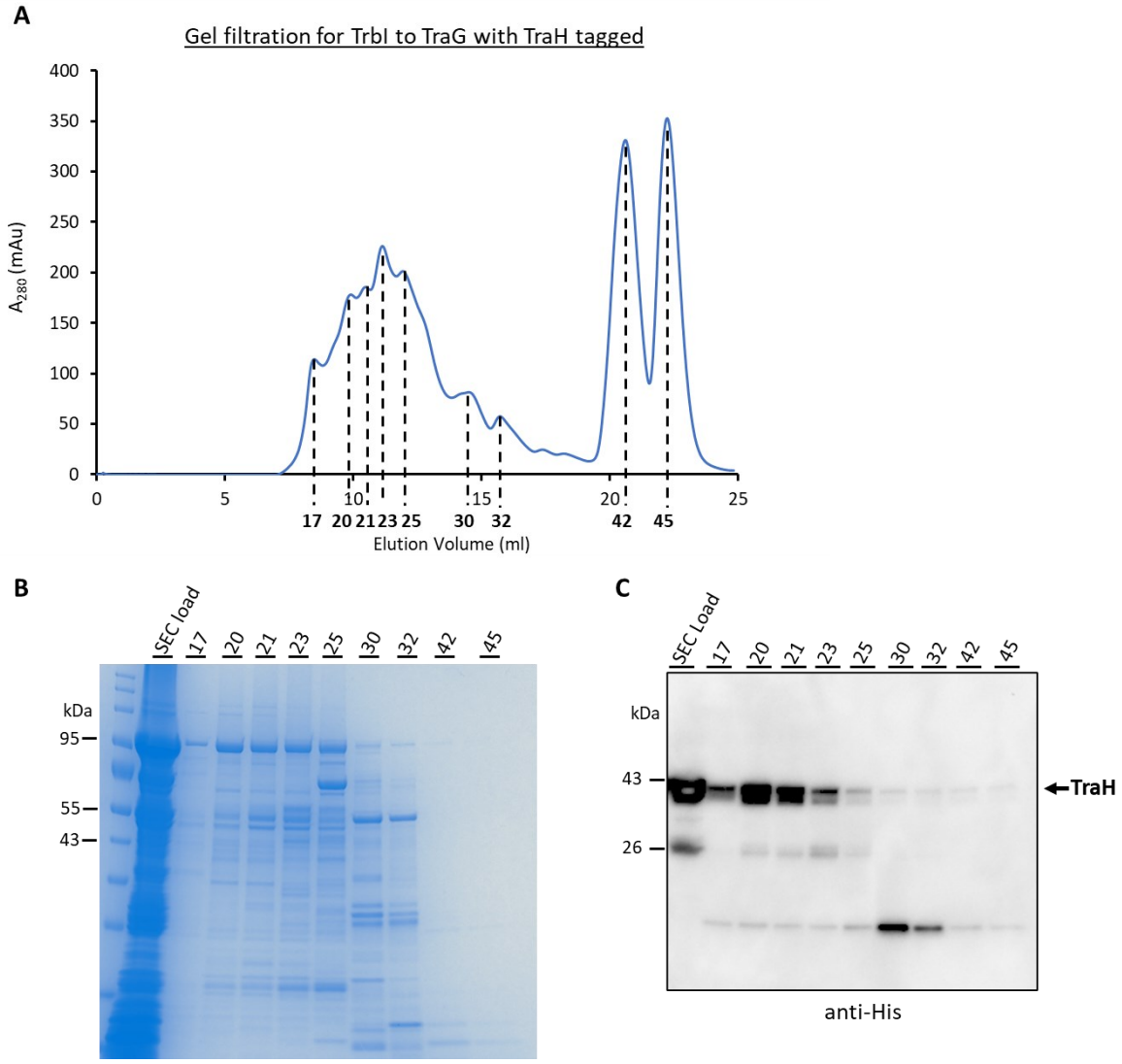

Figure 5.11: The purification of the periplasmic complex. (A) The gel filtration profile with a Superdex200 10/300 column. The fractions have been marked. **(B)** The corresponding SDS-PAGE gel. **(C)** The anti-His western blot for the purification displaying that TraH is present throughout the gel filtration run.

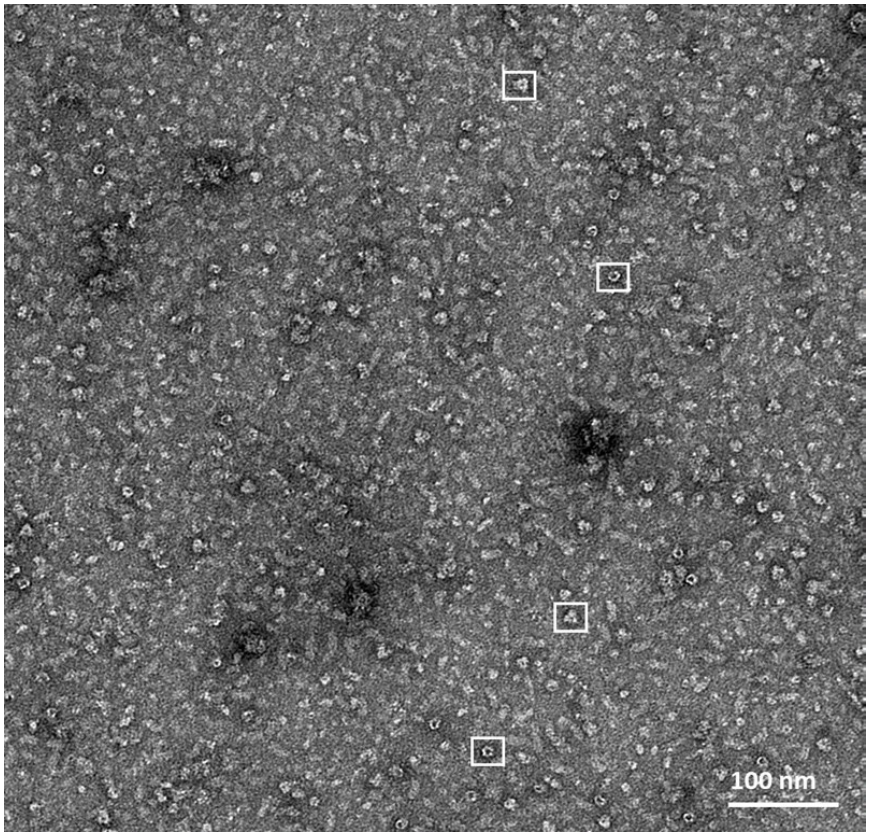

Figure 5.12: The representative negative stain micrograph of fraction 23 from the gel filtration. Ring like and triangular particles were commonly observed (white box). The scale bar represents 100 nm.

5.3.3 Investigating the TraN interaction with F-OMCC

To test for an interaction between TraN and F-OMCC, an expression construct was engineered. To aid the cloning and determine where the affinity tag should be placed, AlphaFold2 was employed to generate a predicted model of TraN and SignalP6.0 was used to predict the presence of a N-terminal signal peptide (Figure 5.13A) (Jumper et al., 2021, Teufel et al., 2022). A pASK vector was built containing the *trbI* to *traG* stretch of the Tra operon with the TraN tagged with 10 histidines at the C-terminus and a flexible GSGSGG linker separating the protein from the tag (Figure 5.13B).

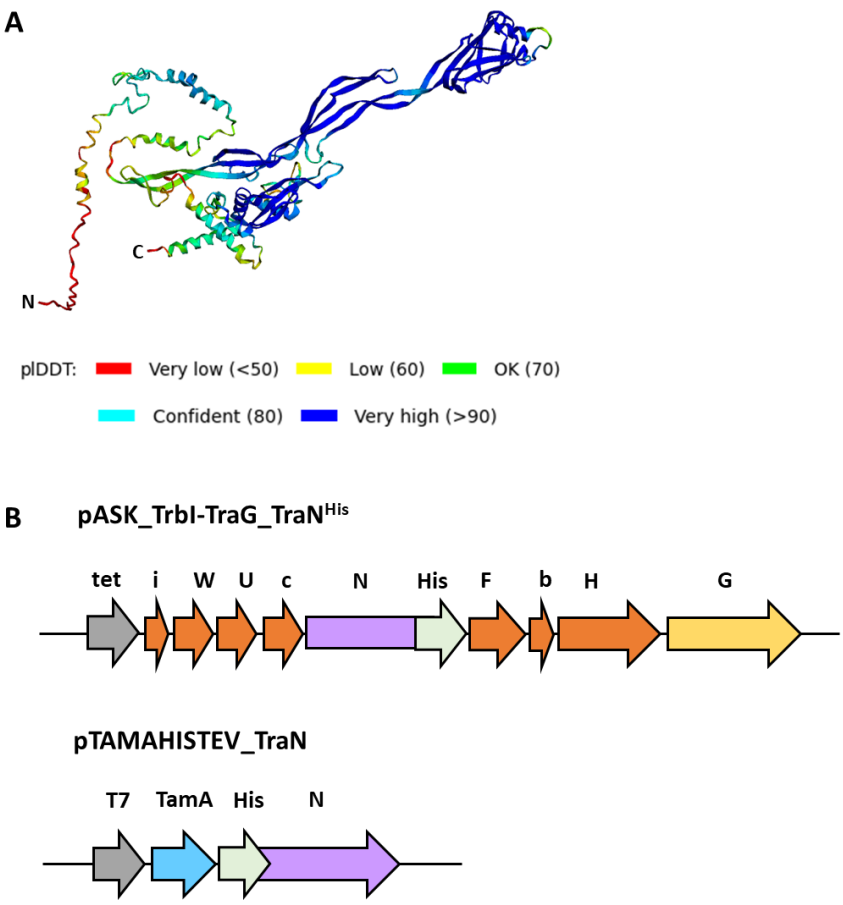

Figure 5.13: The TraN bioinformatic analyses. (A) The AlphaFold2 predicted structure of TraN that has been coloured by pIDDT confidence. **(B)** The diagram of the operon structure of the plasmids used to express TraN in this study. The uppercase letters represent genes which encode for proteins that have the prefix Tra, whereas the lowercase letters denote those with the prefix Trb.

The pASK_TrbI-TraG_TraN^{His} plasmid and the pBAD_TraK_TraB^{Strep}_TraV were co-transformed into the same TOP10 cells and expressed as described in section 2.3.7 (Figure 5.14). The pull-down was performed using the StrepTrap column with F-OMCC as the bait (Figure 5.14 and 5.15A). The elution peak was analysed by SDS-PAGE and western blot. The SDS-PAGE gel showed the characteristic three bands corresponding to the F-OMCC and a faint band was visible just below the 72 kDa molecular ladder marker (Figure 5.15B) This is likely to be TraN as it has an estimated molecular weight of 66 kDa and a band around the same molecular weight was observed in the anti-His western blot (Figure 5.15C). The TraN and F-OMCC appeared to co-elute together as both proteins were present in the elution fractions, implying that they may be interacting. The same result was observed when the pull-down was repeated with a HisTrap

column and using TraN as bait. Together, the two independent pull-down experiments are suggestive of an interaction occurring. However, the low concentration of TraN in the eluent, as highlighted by the faint band in the SDS-PAGE gel, prevented further structural analyses.

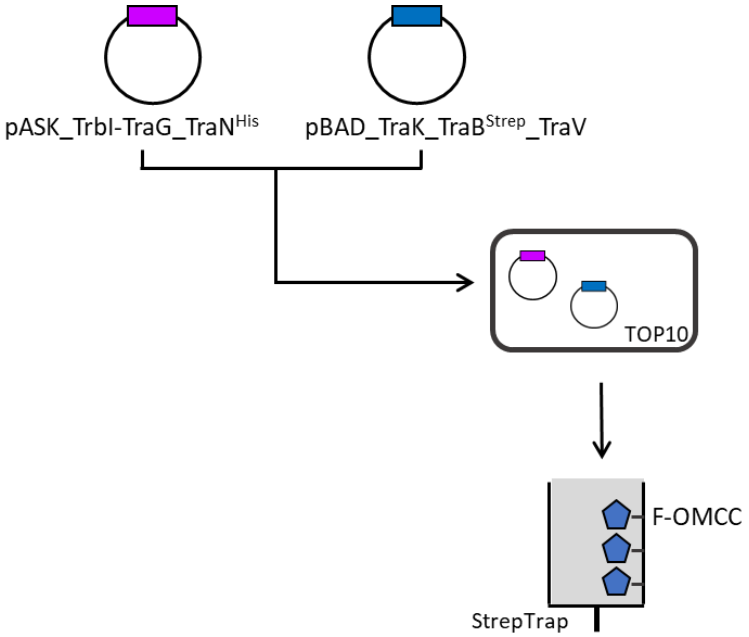

Figure 5.14: An overall schematic of the StrepTrap assisted pulldown of TraN with F-OMCC in the background of other Tra/Trb proteins. TOP10 competent cells were transformed with both plasmids. The blue hexagons the represent the F-OMCC.

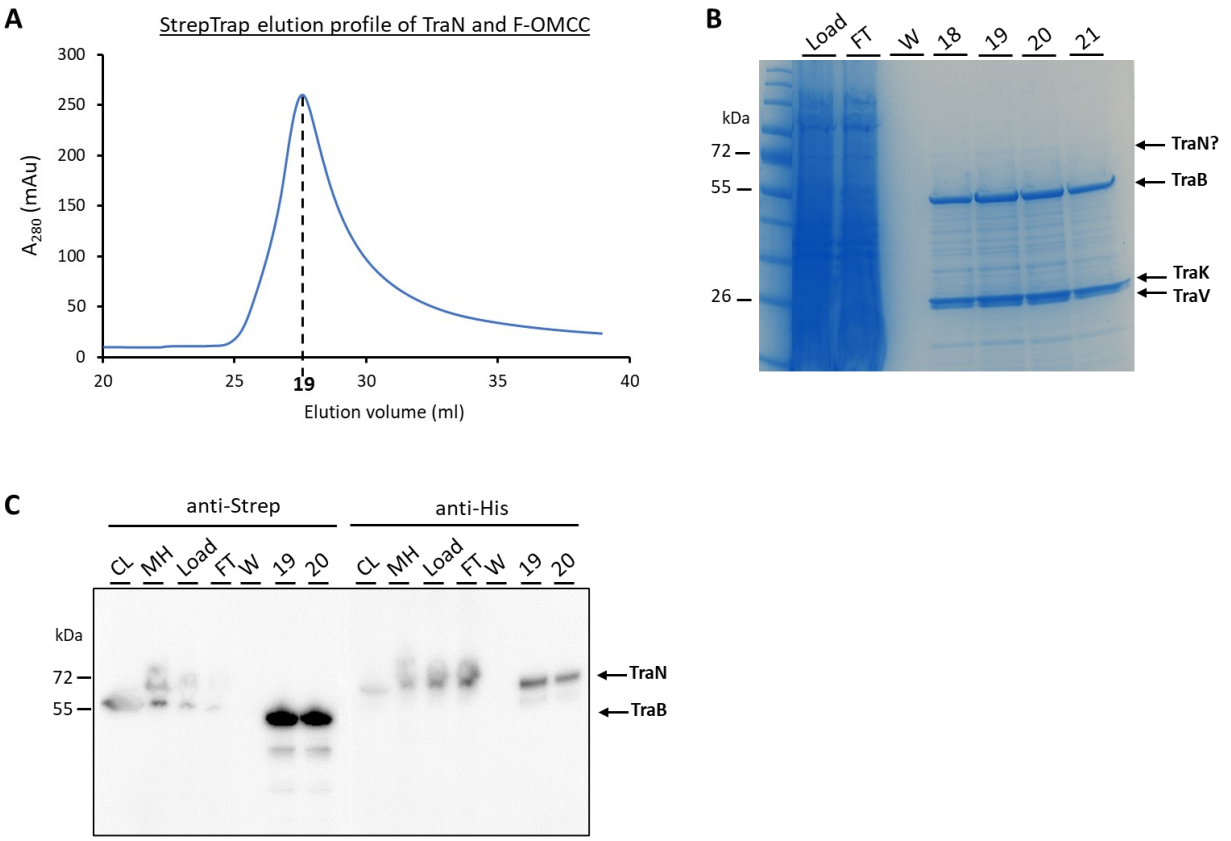

Figure 5.15: The TraN and F-OMCC pull-down in the background of other Tra/Trb proteins. (A) The StrepTrap elution profile of TraN and F-OMCC in the background of proteins present within the TrbI to TraG stretch of the Tra operon. **(B)** The SDS-PAGE gel for the pull-down showing the three F-OMCC bands and a potential TraN band. **(C)** An anti-His and anti-Strep western blot showing TraN and F-OMCC co-eluting. CL (Clear lysate), MH (Membrane homogenised), FT (Flow-through), W (Wash)

To increase the concentration, TraN was cloned into the pTAMAHISTEV vector which encodes for a TamA signal peptide followed by six histidines; thereby facilitating the N-terminal tagging of TraN and ensuring that TraN is directed to the outer membrane (Moynie et al., 2019). This approach was adopted as it has been demonstrated to be effective in expressing the *K. pneumoniae* pKpQIL TraN and enabled successful pull-down with OmpK36 (Low et al., 2022). The pTAMAHISTEV_TraN plasmid was transformed into BL21 competent cells and the plasmid expressing the F-OMCC was transformed into TOP10 cells, as both plasmids could not be co-transformed due to incompatibility. At the stage of lysis, the cell culture resulting from both plasmids were combined and processed together as mentioned in section 2.3.7 (Figure 5.16). To probe for an interaction, a pull-down was performed using a StrepTrap column together with my undergraduate student Michael Garger (Figure 5.17A). The elution peak was analysed by SDS-PAGE gel,

however, a band around the 72 kDa molecular band could not be seen (Figure 5.17B). It should be noted that the profile appeared staggered due to technical issues, however, an overall peak can still be observed. To confirm the lack of TraN in the eluent, a western blot was performed (Figure 5.17C). The anti-His blot showed that TraN was lost in the flowthrough and not present in the eluent. This alludes to the absence of an interaction occurring between the F-OMCC and TraN.

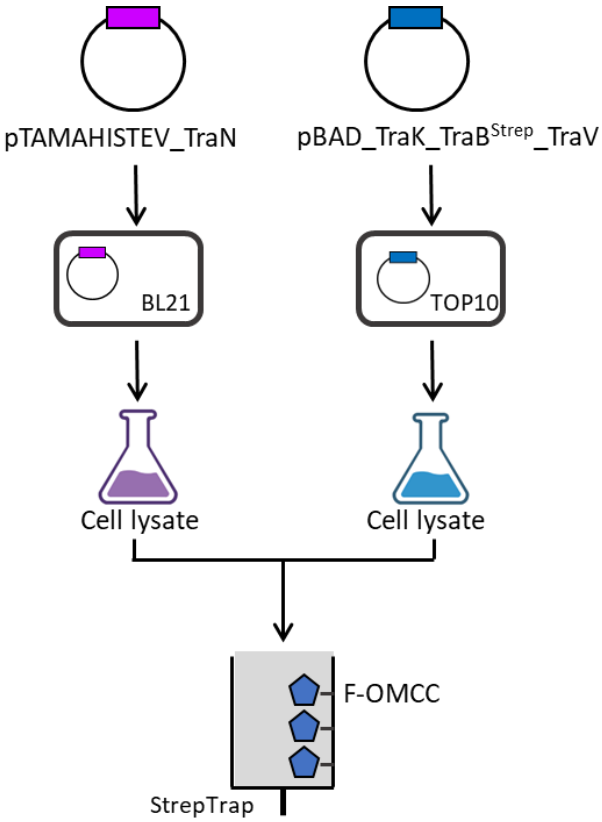

Figure 5.16: An overall schematic of the pulldown of TraN with F-OMCC. BL21 was transformed with the plasmid encoding *traN*, while TOP10 cells were transformed with the plasmid encoding the F-OMCC. The proteins were expressed independently. The cell lysates were combined, and a pull-down was carried out using a StrepTrap column. The blue hexagons the represent F-OMCC.

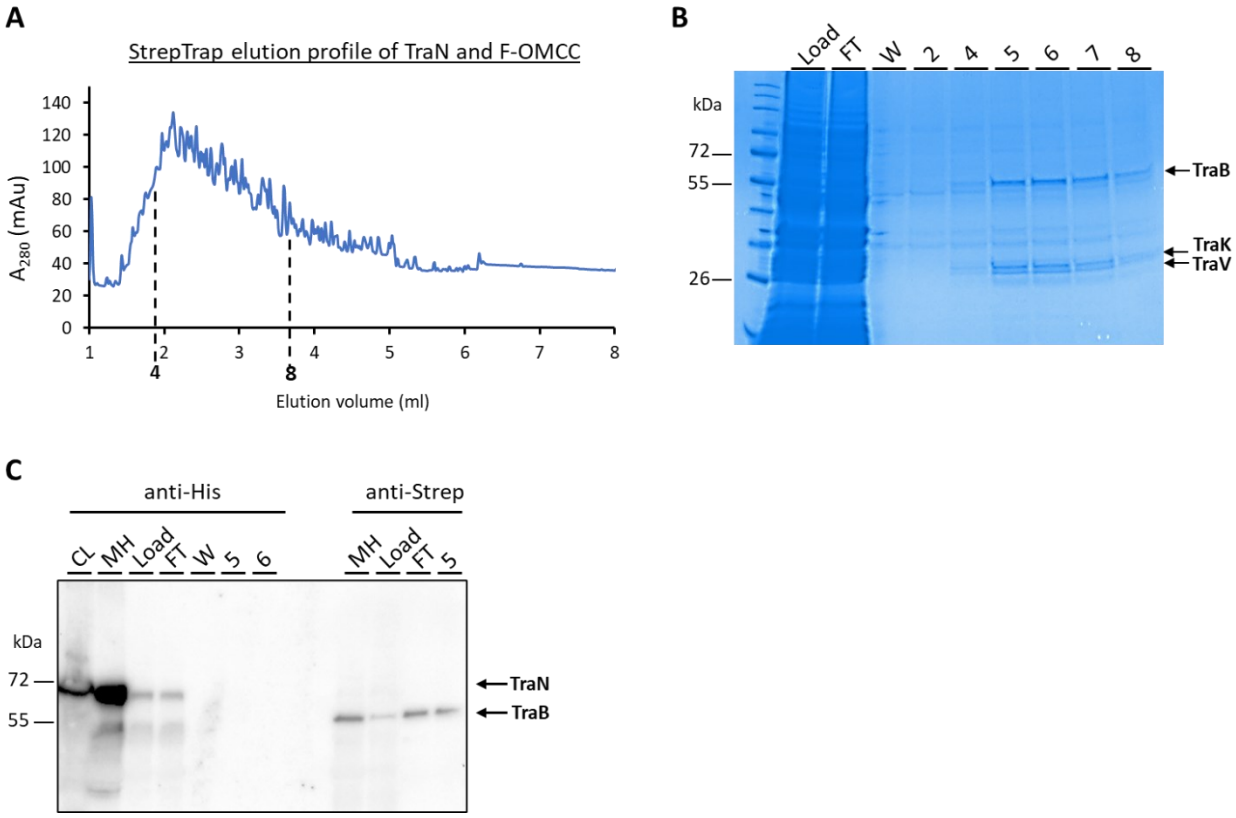

Figure 5.17: The pull-down of TraN with F-OMCC. (A) The elution profile for the StrepTrap pull-down of TraN and F-OMCC. The fractions have been marked. **(B)** The SDS-PAGE gel for the pull-down. **(C)** The corresponding western blot showing that TraN and F-OMCC did not co-elute. CL (Clear lysate), MH (Membrane homogenised), FT (Flow-through), W (Wash)

5.4 Discussion

5.4.1 An interaction between TraW and F-OMCC could not be unequivocally concluded

The results of the individual pull-downs and the double pull-down make it difficult to understand whether a genuine interaction is forming between TraW and F-OMCC. The two components were co-eluting in the HisTrap elution fractions, but this could be questioned considering the impurity of the eluent as highlighted by the corresponding SDS-PAGE gel (Figure 5.4D). While a direct interaction has not been recorded between TraW and F-OMCC from the F-T4SS, an interaction was observed between the TraW and TraV from *N. gonorrhea* T4SS (Koch et al., 2020). This interaction was suggested by a bacterial two-hybrid study and then further confirmed by pull-down. Moreover, it has been suggested in a recent model that TraW together with TraH and TraF bind to the F-OMCC to orchestrate a structural change in the F-OMCC that is needed for pilus production and DNA transfer, but this remains to be proven (Kishida et al., 2022).

An aspect which remains to be explored is whether an interaction between TraW and F-OMCC is facilitated via TrbC. While this might contradict the model proposed by *Kishida et al., 2022*, the association between TrbC and TraV has previously been recorded (Koch et al., 2020). Thus, the presence of TrbC might be required to isolate a definitive interaction. This supports the requirement for probing whether TrbC is present in the pull-down assays and if it is forming a dimer as expected with TraW.

5.4.2 A possible interaction between TraH and F-OMCC may be occurring

Faint bands in the elution corresponding to TraB and TraH were observed in the anti-His and anti-Strep western blots respectively in the preliminary pull-down that was performed between TraH and F-OMCC (Figure 5.10). While this could allude to an interaction occurring between the two components, it is too premature to make any conclusions given the low concentration of both TraH and F-OMCC present in the elution fraction. Furthermore, the pull-down should be repeated but using TraB as bait to probe if the two components were still co-eluting. Nevertheless, the possibility of an interaction occurring is supported by evidence from a T4SS systems where TraH and a component of the OMCC are expressed as a single polypeptide. Recently, a novel T4SS that has been shown to support pathogenesis of *E. coli* associated with Crohn's disease, was found to encode for an outer membrane core complex protein denoted TraBH (Elhenawy et al., 2021). TraBH appears to be a fusion of the N-terminus and C-terminus of TraB and TraH

respectively and is an essential component of the core machinery. So far, this has not been observed in any other F or F-like T4SS but does provide compelling evidence that an interaction between the two components is plausible.

Another aspect which still remains to be explored is whether an undeniable interaction between TraH and the F-OMCC may have been observed in the presence of the other Tra/Trb proteins. As mentioned in section 5.1.2, TraH has many interacting partners, which together are proposed to form a second interaction group referred to as the periplasmic complex in this study. Therefore, TraH may require the presence of the other proteins to effectively interact with the F-OMCC.

5.4.3 A potential interaction may be occurring between TraN and the F-OMCC in the background of other Tra/Trb proteins.

Insight into a potential interaction between the F-OMCC and TraN was observed by both StrepTrap and HisTrap assisted pull-down. The results from the StrepTrap pulldown are particularly convincing given that purifications via a Strep-II tag are more specific compared to the His-tag which can be promiscuous. A faint band was observed in the SDS-PAGE gel which is likely to be TraN as supported by the western blot. This is in line with a proposed model whereby TraN is recruited to the F-OMCC, in the presence of the other F specific factors (Kishida et al., 2022). Furthermore, an interaction between TraN and TraV had been suggested but was not followed up (Lawley et al., 2003). A gel filtration would have been interesting to validate if this potential interaction is occurring. This was limited by the low concentration of TraN which would have been diluted if subjected to further purification.

The pTAMAHISTEV cloning strategy did improve TraN expression, but a possible interaction with the F-OMCC was lost. A possible explanation for this may be that for an interaction to occur, the other Tra/Trb proteins are required. This has been implied in the model suggested by Kishida and colleagues where TraW, TraH and TraF, which were termed Class I proteins, localise to the F-OMCC first. This then creates a platform for the recruitment of TraN, TraU and TrbC to regulate pilus assembly (Kishida et al., 2022). Consequently, it may be the case that the Class I proteins are required in order for TraN to interact with the F-OMCC explaining why a possible interaction was observed when the pull-down was performed with the pASK_TrbI-TraG_TraN^{His} plasmid and not when solely TraN was expressed using the pTAMAHISTEV_TraN plasmid. Another possible reasoning for the lack of interaction could be the experimental set-up. A potential technical issue may be that the TraN and F-OMCC were expressed in two

different cell cultures and hence, encompassed within different membranes. This may mean that when detergent solubilisation is carried out, an interaction is hindered by the detergent micelles shielding TraN and F-OMCC independently. This is particularly likely if the interaction was to occur through the membrane associated regions of the proteins. The technique employed for this pull-down was reminiscent of that used to successfully isolate the *K. pneumoniae* pKpQIL encoded TraN with the cognate OmpK36 which showed that combining detergent solubilised protein components can still result in the successful formation of a complex. (Low et al., 2022). Nonetheless, it is noteworthy that this specific interaction occurred via the extracellularly exposed segments of TraN. Thus, it is unlikely to be bound by detergent micelles (Low et al., 2022). Therefore, perhaps the detergent shielding of TraN and F-OMCC could have resulted in the lack of a possible interaction.

If an interaction is indeed occurring between TraN and the F-OMCC, it raises the question of how TraN partakes in both the regulation of the pilus by localising to the F-OMCC but and in Mps where it interacts with cognate Omp receptor. A scenario that has been implied is that TraN's function in pilus assembly may be antecedent to its role in Mps (Kishida et al., 2022). However, the direct role of TraN in pilus biogenesis and dynamics still remains enigmatic especially given that conjugation still occurs in a *ΔtraN* background albeit at a low level and coupled with the fact that the direct involvement of TraN in a pilus function is a recent proposition.

5.4.4 Future perspectives

The work described in this Chapter provides insight into potential interactions that may be occurring between other Tra/Trb proteins and the F-OMCC. Preliminary experiments have been discussed which upon further work could enable the isolation of a large complex of interacting proteins. Future work could include improving expression and purification of certain proteins, such as TraH. This can be done by exploring other expression vectors or changing the position of the affinity tag such that it still enables effective localisation of the protein. Furthermore, raising antibodies against specific Trb/Tra proteins which are presumably involved in various interaction networks with other will be valuable to ensure their presence throughout the experiments. For example, antibodies against TrbC will be useful for exploring whether the protein is being expressed by the plasmid pASK_TraW^{His}_TrbC and, if it is present in the pull-down. Overall, exploring interactions within the F-T4SS and particularly those occurring with the F-OMCC is important as it will improve understanding of how these proteins are synergistically involved in conjugation.

Chapter 6: The trifunctional, inner membrane protein TraG

6.1 Introduction

TraG is one of the largest proteins encoded in the *tra* region with a molecular weight of 102.5 kDa (Frost et al., 1994). The protein consists of an N-terminal inner membrane region, which is thought to potentially traverse the membrane three to five times by hydropathy plot analysis, and a large periplasmic CTD which can be proteolytically cleaved (Figure 6.1A and B) (Firth and Skurray, 1992, Frost et al., 1994). The NTD is needed for pilus assembly and is considered to be the VirB6 counterpart in the F-T4SS (Audette et al., 2007, Frost et al., 1994, Moore et al., 1981). The periplasmic CTD is homologous to VirB8 and is involved in both Mps, and Eex (Manning et al., 1981, Audette et al., 2007). Therefore, TraG is a remarkable protein which is a fusion of VirB6 and VirB8.

6.1.1 The trifunctional role of TraG

TraG has been shown to be important for three various steps of conjugation. Firstly, it is involved in Mps through action with TraN. Secondly, TraG has also been implicated in entry exclusion alongside TraS to prevent redundant conjugative transfer between two F^+ cells (Audette et al., 2007). Lastly, it has been suggested to play a role in pilus assembly which may be due to the protein being a fusion of the inner membrane components VirB6 and VirB8 (Frost et al., 1994, Bragagnolo et al., 2020). Therefore, TraG is likely to form part of the core conjugative apparatus.

For entry exclusion, TraG in the donor cell was found to recognise TraS in the inner membrane of the recipient cell in a plasmid-specific manner (Audette et al., 2007). There have been several propositions to explain how two inner membrane proteins may interact given the vast cellular space that would need to be traversed. A common scenario that has been proposed is that TraG may be translocated to the recipient cell to then interact with TraS which subsequently prevents the transfer of DNA between equivalent donors (Marrero and Waldor, 2005, Audette et al., 2007). However, given that TraG is likely to form a part of the core conjugative apparatus, like VirB6 and VirB8, and thereby involved in a network of interactions, it is difficult to envision how such a crucial protein embedded in the membrane could be transported across to the recipient cell.

The CTD of TraG was also found to be important in Mps along with TraN which enables the donor cell to mate more successfully in liquid media and also withstand shear forces or the addition of harsh agents such as SDS (Manning et al., 1981). It was previously hypothesised that the CTD is proteolytically cleaved to produce a periplasmic 458 amino acid product (TraG*) which is then involved in Mps (Firth and Skurray, 1992, Klimke and Frost, 1998). However, it was later shown that in order for Mps to occur, full-length TraG is required and that TraG* alone is not sufficient (Audette et al., 2007). Some have even speculated that TraG may interact directly with TraN but, direct evidence for this has not yet been provided (Klimke and Frost, 1998).

6.1.2 The current understanding regarding the oligomeric state of TraG

There has been uncertainty regarding the oligomeric state of TraG. This is further complicated by TraG being a fusion of VirB6 and VirB8 which have both been shown to have different symmetrical arrangements. Several VirB8 structures have been solved from numerous species which have shown varying oligomeric states in solution. The VirB8 from *A. tumefaciens*, *Brucella suis* and *Rickettsia typhi* were found to be dimers (Terradot et al., 2005, Bailey et al., 2006, Gillespie et al., 2015). The VirB8 counterpart from pKM101, TraE was found to form a hexamer when isolated, and when in complex with the VirB6 homolog TraD. More recently, the structure from R388, showed VirB8 forming a hexamer of homotrimers (Mace et al., 2022). This indicates the great diversity that appears to exist in the organisation of VirB8 and homologs in these systems. On the other hand, VirB6 was shown to assemble as a pentamer and the individual subunits were found to interact extensively (Mace et al., 2022). Therefore, it is difficult to predict the oligomerisation of TraG and this can be addressed using structural analyses.

In section 5.3.2.2, when a pull-down of the periplasmic complex with tagged TraH was attempted, ring-like particles were clearly seen in the negative stain micrographs. It was suspected whether these could be TraG since similar structures were seen for the hexameric TraE from pKM101 (Casu et al., 2018). However, this would require isolation of the protein to confirm that the ring like particles are indeed TraG.

6.2 The aim of this Chapter

The goal for the work described in this Chapter was to obtain a TraG sample which would be amenable for single particle cryo-EM to enable high resolution structure determination but will also enable pull-

down assays to investigate whether TraG interacts with TraN. Firstly, a construct was engineered to express full-length TraG. The TraG was then solubilised from the bacterial membrane and purified by affinity purification and gel filtration. The elution profile from the gel filtration gave insight into the oligomeric state of TraG. Alongside this, pull-down assays was performed to investigate if TraG and TraN interact since both proteins are involved in Mps.

6.3 Results

6.3.1 Preliminary work suggests that TraG forms high oligomeric structures

A pBAD construct was cloned whereby TraG was tagged with a Strep-II tag at the C-terminus since the N-terminus has been shown to contain transmembrane domains and is structured which was supported by the AlphaFold2 predicted structure (Figure 6.1A,B) (Jumper et al., 2021, Hallgren et al., 2022). To increase accessibility of the tag and avoid steric hindrance, a flexible GSGSGG linker was introduced between the protein and the tag (Figure 6.1C).

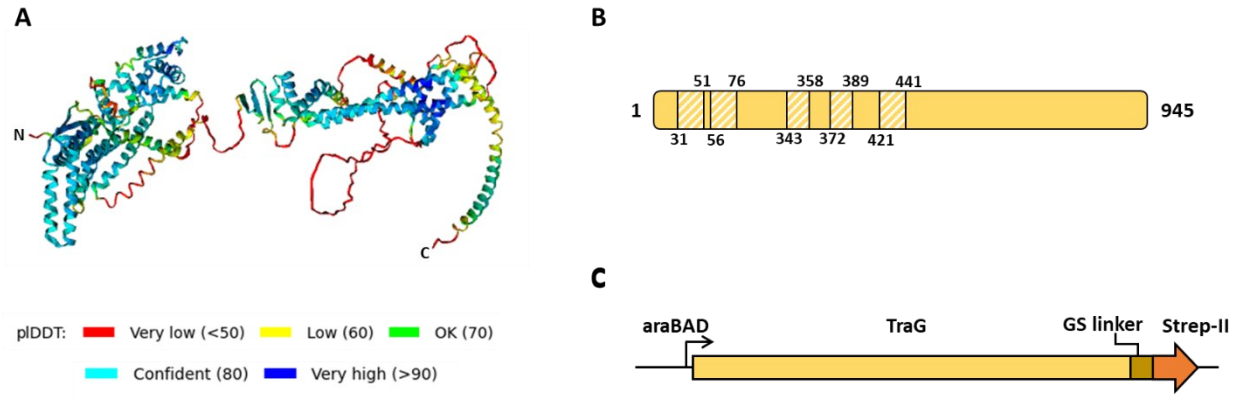

Figure 6.1: The bioinformatic analyses of TraG. (A) The AlphaFold predicted structure of TraG that is coloured according to the plDDT confidence (Jumper et al., 2021). **(B)** TraG with the predicted transmembrane domains displayed as striped boxes. The domains were predicted using DeepTMHMM (Hallgren et al., 2022). **(C)** The diagram of the operon structure of the plasmid used to express TraG with a C-terminal Strep-II tag and under the arabinose promoter. The protein and tag were separated by a GSGSGG linker.

The TraG was expressed as mentioned in section 2.3.9 and a trial purification was performed using DDM. The protein was first purified by affinity purification using a StrepTrap column and eluted with desthiobiotin. In the elution profile, a small peak was observed which upon analysis by SDS-PAGE appeared to contain a protein at approximately 100 kDa indicated by the prominent band. This is likely to be TraG which has a predicted molecular weight of 102 kDa (Figure 6.2A, B). This band was confirmed to be TraG by western blot using anti-Strep antibodies (Figure 6.2C). The blot also showed that the other contaminating bands at a lower molecular weight are TraG degradation from the N-terminal region. The fractions containing TraG were pooled, concentrated, and then injected onto a Superdex 200 increase 10/300 column for further purification by gel filtration (Figure 6.3A). The profile showed several peaks

which were investigated by SDS-PAGE (Figure 6.3B). The peak at around 10 ml was shown to contain TraG by the band at approximately 100 kDa (fraction 21). However, it appears that this TraG peak is associated with the void peak which is expected at just after 8 ml. This is likely due to the size of the protein being large and at the top-end of the protein fractionation power of the Superdex200 column which is 10 kDa to 600 kDa (Cytiva handbook). Using the standard curve calculated for this column in the lab, the protein migrates with a molecular weight of roughly 570 kDa which would correspond to a high oligomeric state of TraG, potentially a hexamer. The fraction 21 corresponding to the TraG peak was concentrated and visualised by negative stain EM (Figure 6.3C). The micrograph showed small white specks of protein amongst which ring-like structures could be seen.

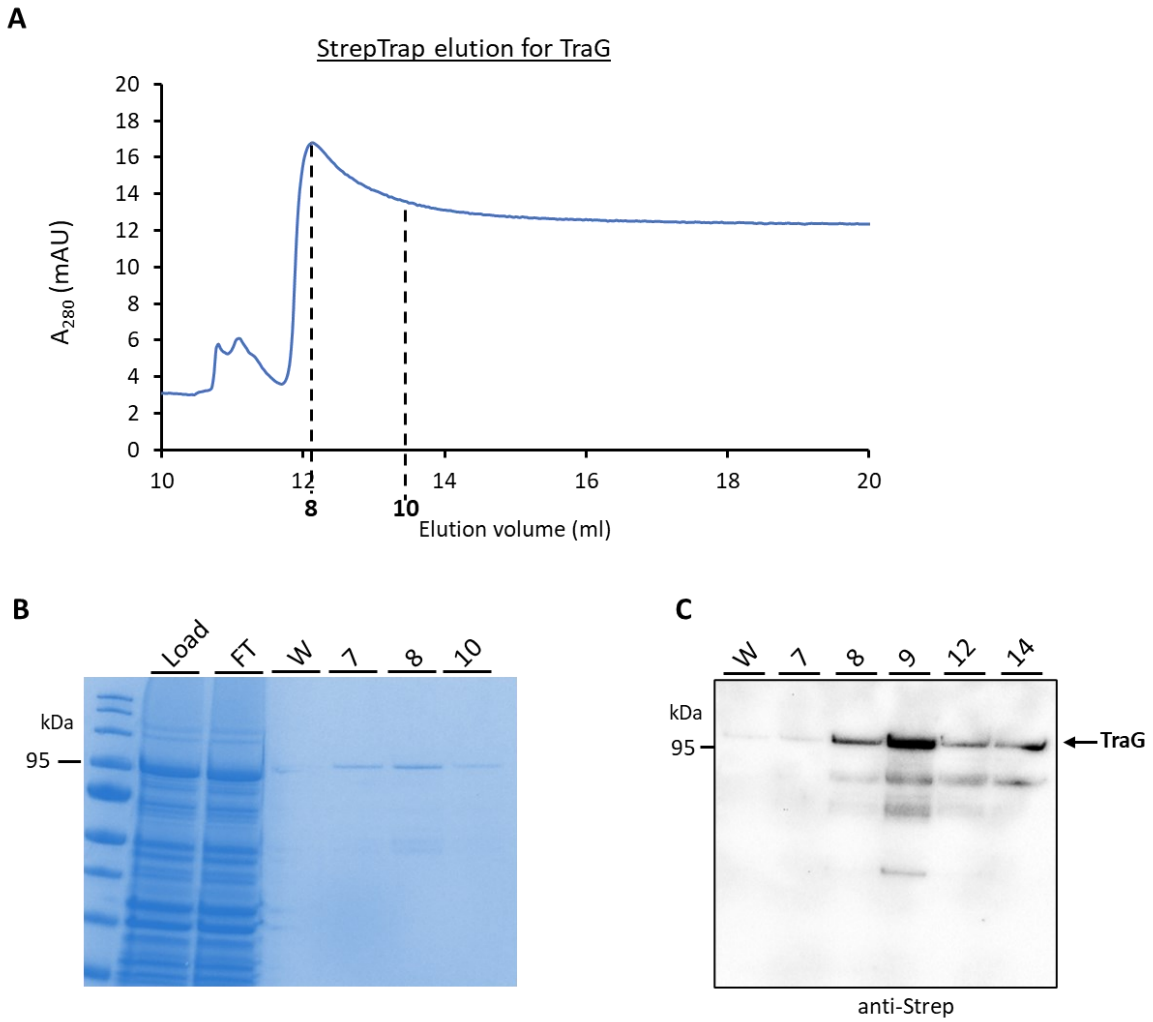

Figure 6.2 The affinity purification of TraG with DDM. (A) The elution profile of the StrepTrap purification showing a peak with the fractions marked. **(B)** The corresponding SDS-PAGE gel which shows that the elution peak contains TraG as indicated by the band at approximately 100 kDa. **(C)** An anti-strep western blot showing that the band at 100 kDa in the elution peak is due to TraG. FT (Flow-through), W (Wash)

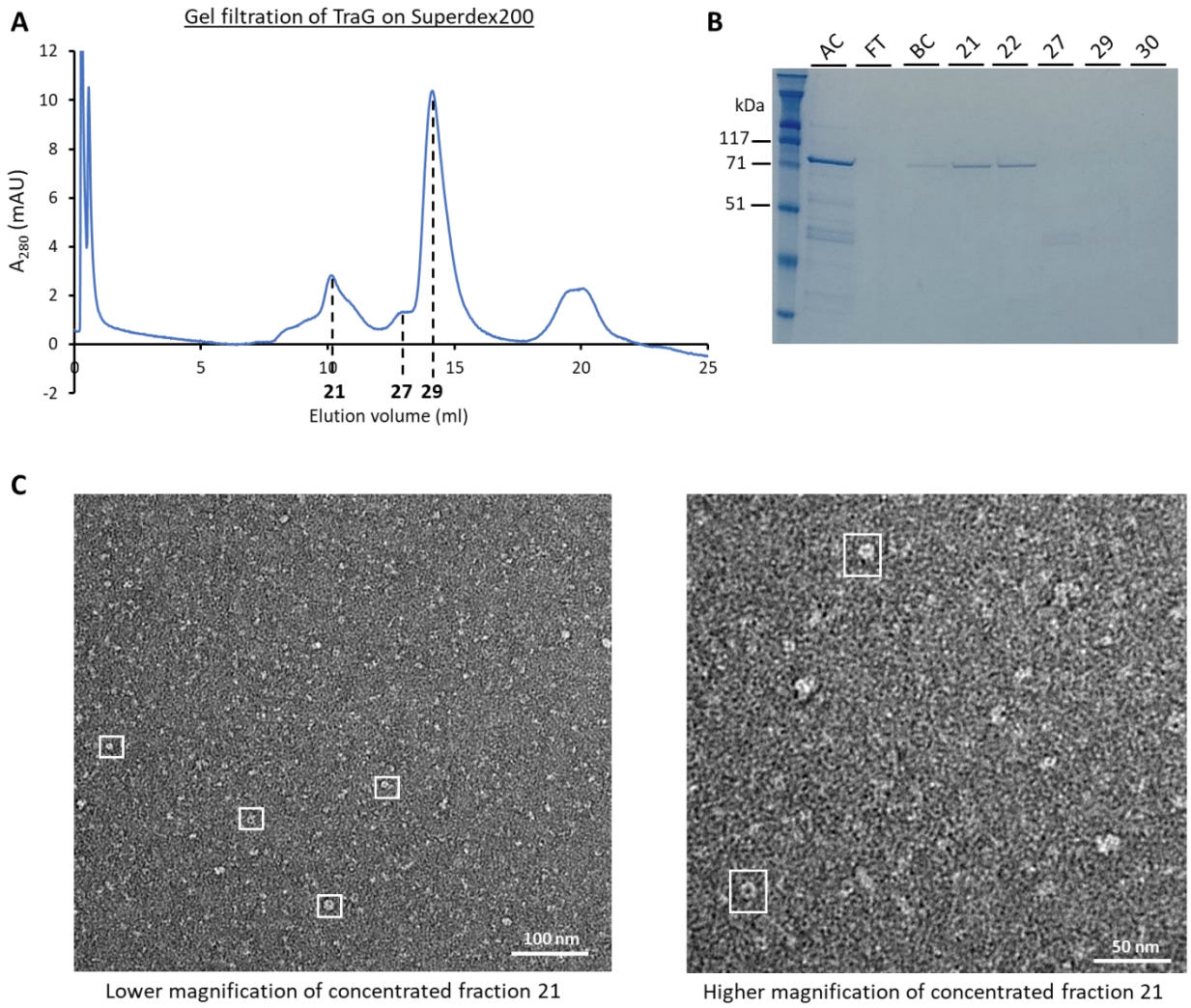

Figure 6.3 The gel filtration of TraG with DDM. (A) The gel filtration profile of TraG using the Superdex200 column with fractions marked. **(B)** The corresponding SDS-PAGE gel **(C)** The lower magnification and higher magnification micrographs of the concentrated sample from fraction 21 in the gel filtration, with scale bars representing 100 nm and 50 nm respectively. Ring like structure were observed like those boxed in white. AC (After concentrating), FT (Flow-through), BC (Before concentrating)

It was apparent from the DDM purification and the negative stain micrograph that concentration of TraG needed to be improved for further analysis, such as cryo-EM. The purification was repeated with the detergent OGNG which was found to be better at extracting the TraG homologue from the pKM101 T4SS, TraE (VirB8), from the bacterial membrane (Casu et al., 2018). The elution profile from the affinity purification and the corresponding SDS-PAGE gel showed a remarkable improvement in concentration of TraG, as illustrated by the increase in absorbance values and band intensity respectively (Figure 6.4A, B).

The fractions corresponding to TraG were pooled, concentrated, and further purified by gel filtration. To allow for better separation of the TraG peak from the void, a Superose 6 increase 10/300 column was used (Figure 6.4C). Unfortunately, the run was limited by issues with the UV lamp. Nevertheless, the profile shows numerous peaks with the most prominent peak (peak 1) eluting at approximately 11 ml and subsequent peaks eluting at approximately 14 ml (peak 2), 15 ml (peak 3) and 16.5 ml (peak 4). SDS-PAGE analysis showed that peak 1, 2 and 3 contained TraG and peak 4 may also consist of TraG but due to low concentration could not be detected by Coomassie staining (Figure 6.4D). This indicates that TraG exists in a mixture of oligomeric states in solution. Preliminary insight into the molecular weight of the TraG in the various peaks was probed using a standard curve for the Superose 6 column in the lab. Based on this, the TraG in peak 1 appears to migrate with a molecular weight of roughly 1 MDa. The protein in peak 2,3 and 4 migrates with size of roughly 570 kDa, 400 kDa and 150 kDa respectively which would suggest TraG migrating close to a hexamer, tetramer and dimer/monomer.

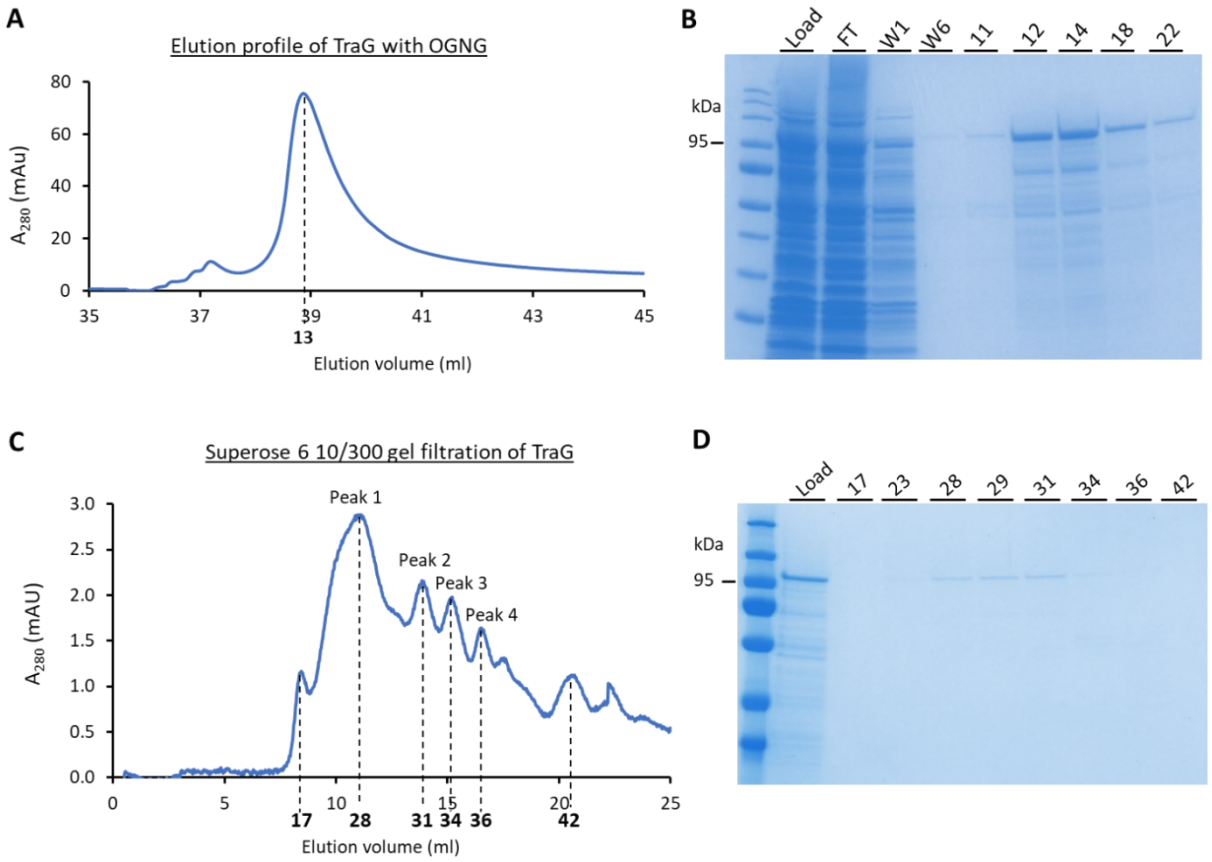

Figure 6.4: The purification of TraG with OGNG. (A) The affinity purification of TraG with the fractions marked. **(B)** The SDS-PAGE gel corresponding to the affinity purification. **(C)** The gel filtration profile of TraG using the Superose 6 10/300 column. The fractions have been indicated with the prominent peaks numbered. **(D)** The SDS-PAGE gel corresponding to the gel filtration run. FT (Flowthrough), W (wash)

As the TraG in peak 2 was migrating close to a hexamer, an oligomeric order which has been observed in other T4SSs, the sample within this peak was concentrated and visualised by negative stain EM (Figure 6.5). Small particles were observed which when magnified, appeared to be ring-like structures which were homogenous, disperse and abundant. The average diameter of the particles is 0.9 nm. A larger but rarer structure was also observed which has a diameter of 18 nm.

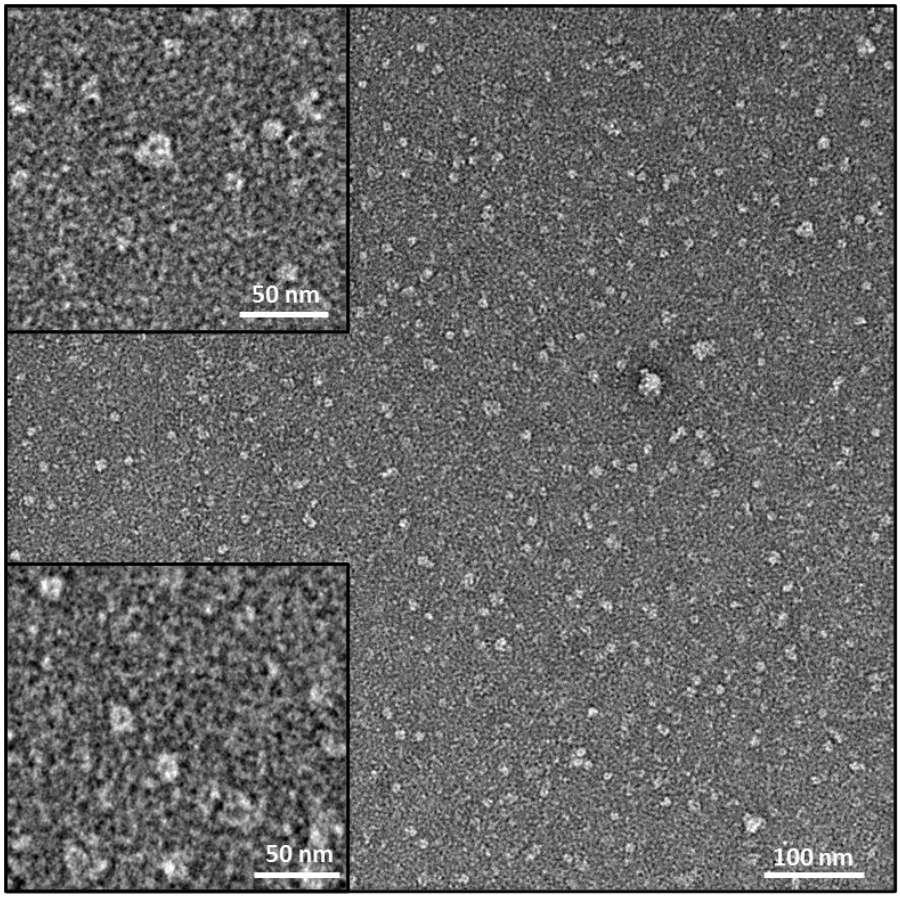

Figure 6.5: Negative stain micrograph of fraction 31 from peak 2 from the gel filtration run. A representative micrograph of concentrated fraction 31 showing small particles. The magnified section on the top left highlights a large particle that was present. The magnified section on the bottom left indicates the small ring like particles that were commonly seen. The scale bar represents 100 nm and 50 nm for the magnified images.

6.3.2 Cross-linking of TraG

To investigate whether the larger structure was indeed TraG that was unstable and thereby dissociating into the plentiful smaller ring like particles, cross-linking was carried out. This technique proved to be effective with TraE from pKM101 T4SS. This should also improve the gel filtration profile by producing a prominent peak corresponding to the cross-linked TraG. A cross-linking assay was carried using disuccinimidyl suberate (DSS) which showed high molecular weight banding on the SDS-PAGE gel (Figure 6.6A). The reaction was carried out as indicated in (Casu et al., 2018). For the purification, 2 mM DSS was used since the TraG was still predominately monomeric when 1.6 mM of DSS was used (Figure 6.6A). The DSS was combined with TraG following affinity purification and concentration. After 1 hr at room temperature with DSS, precipitation was observed which was pelleted. The sample was further purified by gel filtration using the Superose6 increase 10/300 column (Figure 6.6B). The profile showed a large void peak at just after 8 ml. The corresponding SDS-PAGE gel shows an intense staining at the wells of the gel for the void fractions 21 and 22 which could be due to protein being aggregated and unable to pass through the gel matrix (Figure 6.6C). A similar staining is also observed in the well labelled pellet which is a sample from the precipitation following cross-linking which is likely to be aggregated TraG. Nevertheless, compared to the AC sample, it appears that most of the monomeric TraG in the load was cross-linked, suggesting that the 2 mM of DSS is effective. The sample from the void peak was visualised by negative stain EM, however, the sample appeared aggregated. Therefore, the cross-linking of TraG still requires optimising.

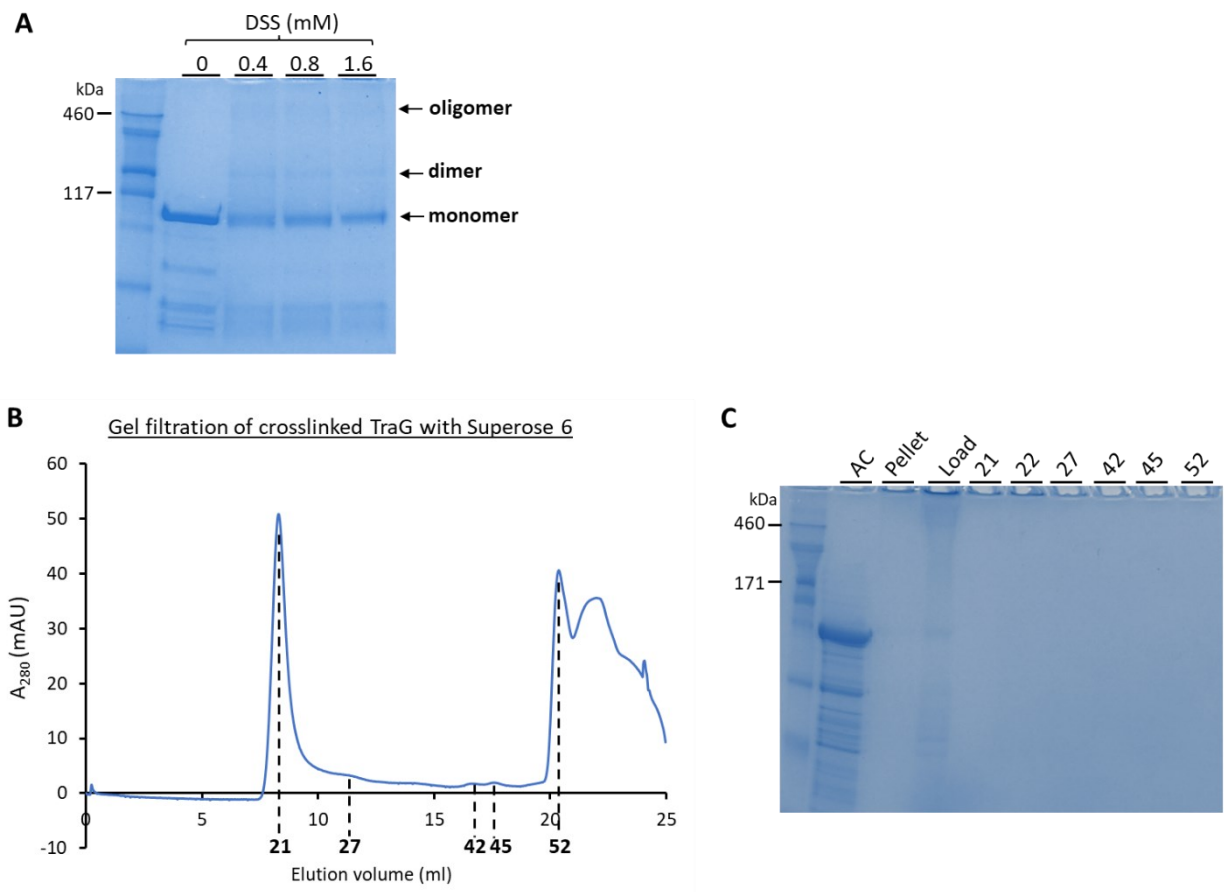

Figure 6.6: The cross-linking of TraG with DSS. (A) A cross-linking assay to determine optimal DSS concentration to effectively cross-link TraG. Various high molecular weight bands were observed following cross-linking indicating oligomers of TraG. **(B)** The gel filtration profile of cross-linked TraG using the Superose 6 10/300 column and with the fractions indicated. **(C)** The SDS-PAGE gel for the gel filtration profile. AC (After concentrated)

6.3.3 Investigating whether TraG interacts with TraN

Since both TraG and TraN have been proposed to be involved in the early steps of conjugation by facilitating the mating pair stabilisation, a pull-down assay was carried out to investigate whether the two proteins interact. For this, the pTAMAHISTEV_TraN construct was used which was shown in section 5.3.3 to enable expression of His-tagged TraN. Since both the TraG and TraN construct have the same antibiotic resistance cassette, it prevents co-transformation. Thus, both constructs were expressed and harvested separately and then combined during the lysis step and the procedure is detailed in section 2.3.10 (Figure 6.7). Two independent pull-downs were performed, one using a HisTrap and immobilising TraN and the second using StrepTrap column and therefore immobilising TraG. The StrepTrap profile showed a

prominent elution peak (Figure 6.8A). For the HisTrap pull-down, elution was carried with 100% (v/v) buffer B which would correlate to 500 mM imidazole as this was a preliminary experiment to check if both TraG and TraN are eluting together, as supposed to trying to obtain pure samples (Figure 6.8B). An elution peak was observed. Both pull-downs were initially analysed by SDS-PAGE analysis but due to multiple bands existing in the elution fractions, it was difficult to accurately assess if both TraG and TraN were co-eluting. Therefore, a western blot was carried out and fractions from both pull-downs were probed by both anti-His and anti-Strep antibodies (Figure 6.8C). It appears that TraN and TraG do not interact as they are not co-eluting together. The elution fractions 15 and 16 from the HisTrap pull-down contains TraN as indicated by the band around 60 kDa in the anti-His blot. However, TraG is not present in the same fractions as highlighted by the absence of a band at around 100 kDa in the anti-Strep blot. Similarly, the StrepTrap elution fractions 13 and 14 contains TraG but not TraN. This is further supported by the observation that in the StrepTrap pull-down, TraN is being lost in the flow-through as indicated by the blot at approximately 60 kDa in the lane labelled FT in the anti-His western blot. This suggests the lack of an interaction between the two proteins, at least, in this experimental set-up.

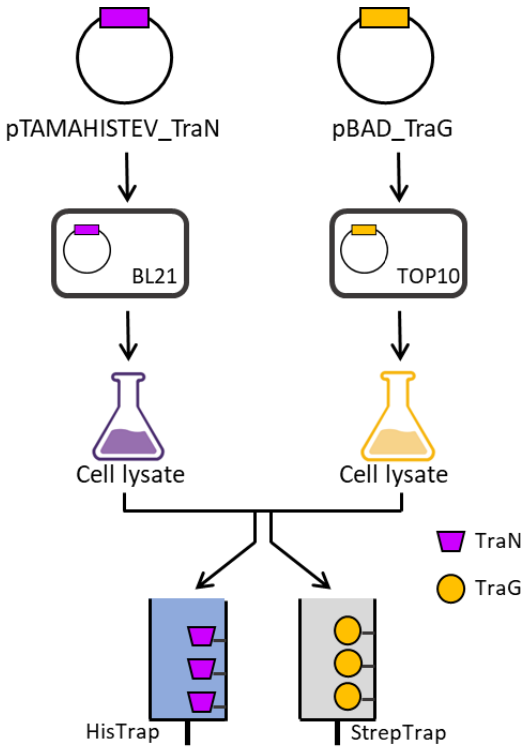

Figure 6.7: An overall schematic of the pull-down of TraN and TraG. The plasmids were co-transformed into different competent cells and expressed independently. The cell lysates were combined, and two independent pull-downs was carried out using either a StrepTrap or Histrap column. The purple inverted trapezoid represents TraN and the yellow circle represents TraG.

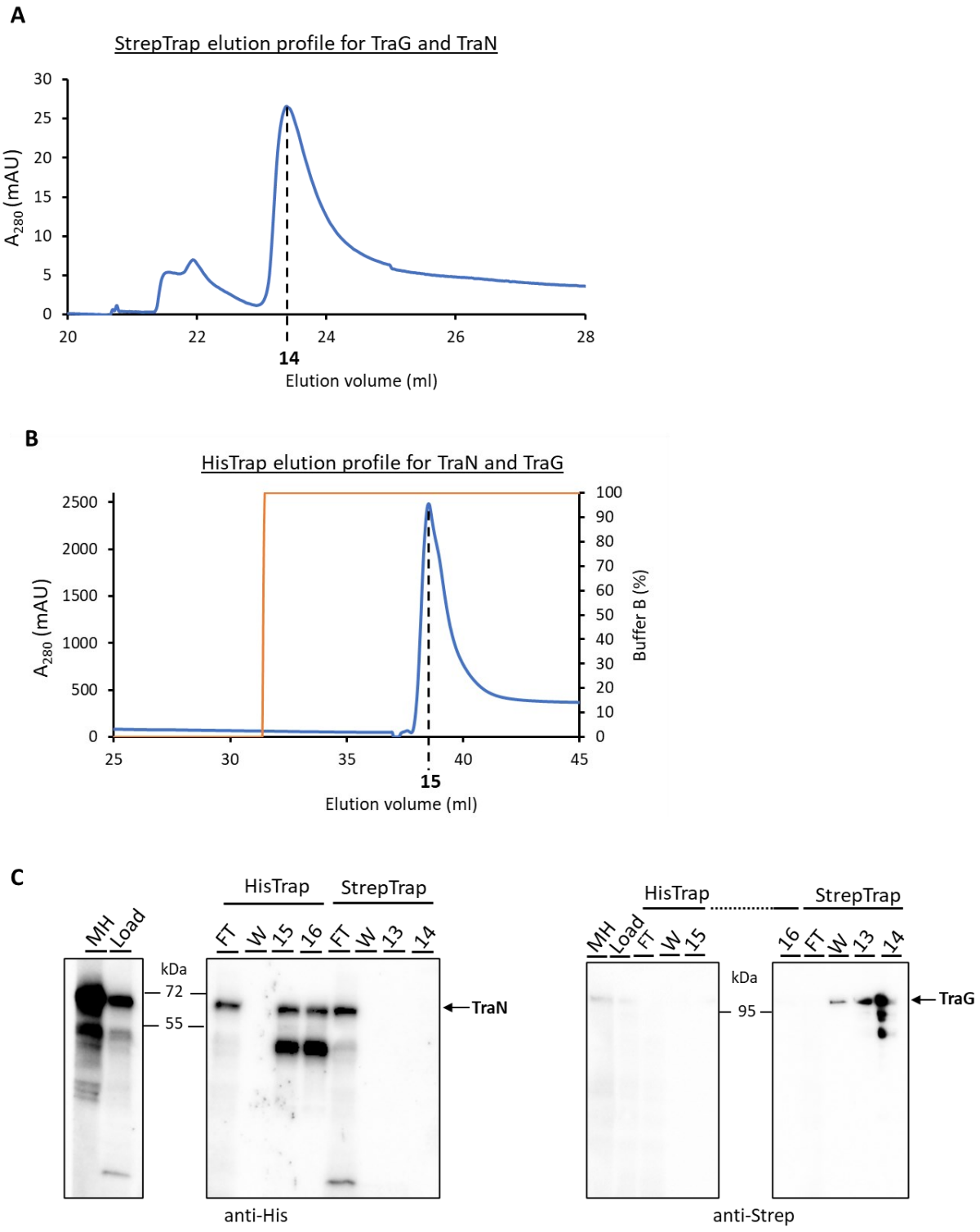

Figure 6.8: Pull-down to probe if TraG and TraN interact. (A) The elution profile of StrepTrap assisted pull-down of TraN with TraG. **(B)** The elution profile for HisTrap assisted pull-down of TraG with TraN. The 100% (v/v) buffer B corresponds to 500 mM imdiazole **(C)** An anti-His and anti-Strep western blot for the pull-down assays showing that TraG and TraN did not interact. MH (Membrane homogenised), FT (Flowthrough), W (Wash)

6.4 Discussion

In this Chapter, the recombinant expression of TraG was carried out and preliminary purifications were described. The TraG appears to assemble as a high order oligomer, potentially a hexamer. A pull-down was also performed with TraG and TraN, however, the results suggest that the two proteins do not strongly interact.

6.4.1 TraG appears to assemble as a hexamer

The gel filtration profile displayed in figure 6.3A and 6.4C indicates that TraG assembles as a high order oligomer and based on the elution volume of the peak, is suggestive of a hexameric protein. This oligomeric arrangement has also been seen in the pKM101 homolog, TraE and more recently in the R388 structure where VirB8 assembled as a homotrimer of hexamers (Casu et al., 2018, Mace et al., 2022). Here, the VirB8 was shown to form the periplasmic arches with the N-terminal alpha-helical structure forming part of the IMC, termed $VirB8_{tails,}$ and interacting with VirB4 (Mace et al., 2022). A similar periplasmic arch-like structure was also observed on the cryo-ET structure of the F-T4SS where it was labelled as 'collar' and displayed an overall hexameric structure (Hu et al., 2019a). The approximate diameter of this collar is 180 nm which would coincide with the size of the larger but rarer structures that were observed in Figure 6.5. Based on this, it is tempting to wonder whether the collar structure is formed by the hexameric TraG, possibly through its C-terminal portion. The oligomeric state of TraG will need to be further validated by other biochemical techniques such as SEC-MALS or by reconstruction of the high-resolution 3D structure. A fascinating aspect of TraG is that it is a fusion between VirB8 and VirB6. These two proteins have been shown to have different symmetrical arrangements in the context of the whole R388 machinery and therefore it will be interesting to study whether TraG accommodates this by tolerating a symmetry mismatch within the structure. Currently, it is too premature, to draw such conclusions based on the negative stain micrographs but this can be addressed by future cryo-EM work.

If indeed TraG is forming a high order oligomer, it poses the conundrum of how it is involved in Eex by recognition of cognate TraS in the recipient cell. A recurring model that has been proposed is that TraS recognition by TraG occurs via translocation of the protein into the recipient cell. Given that the NTD of TraG is homologous to VirB6, which is a crucial component needed for the transfer of DNA, and that the CTD is homologous to VirB8 and likely to be hexameric as indicated by the results in this Chapter, it is very

plausible that TraG forms an integral part of the core secretion channel. In this instance, it does not seem probable that such an important protein embedded within the machinery could be unattached and transferred to the recipient cell. Another theory that has been put forward is that a part of TraG is exposed to the periplasm of the recipient cell following mating bridge formation (Marrero and Waldor, 2005). Using a series of TraG chimeras, the region of TraG needed for recognition of TraS was identified to be the middle area of the protein. This was further proposed to be between aa 610 – 673 as this is the region where sequence dissimilarity is observed in the TraG from other F and F-like plasmids and has been proposed to infer specificity (Audette et al., 2007). These amino acids residues in the TraG AlphaFold model reside in the long stretch of unstructured region which was predicted with poor confidence (Figure 6.1A). This may be the part of the structure which facilitates Eex, however, whether it could be positioned in such a way that it is exposed to the TraS located in the inner membrane of the recipient cell seems doubtful. Nevertheless, a high-resolution structure of TraG would certainly help clarify how this region is incorporated within the oligomeric structure and will shed light on how the protein participates in Eex.

An inherent issue with TraG sample is that it appears to be unstable and could be dissociating, as indicated by Figure 6.4C and 6.4D. A plausible explanation could be that for proper stabilisation of TraG other Tra proteins are required. A likely candidate is the ATPase TraC which may interact with TraG as seen in the R388 structure whereby $VirB8_{tails}$ interacts with the outer VirB4 with two out of the three homotrimeric subunits making extensive contact with the ATPase (Mace et al., 2022). This is also supported by the observation that the periplasmic collar was no longer seen in the Δ*traC* mutant cells (Hu et al., 2019a). This suggests that perhaps TraC is required for stabilisation and proper hexameric arrangement of TraG. Moreover, VirB8 from *A. tumefaciens* has been suggested to interact with other components of the T4SS machinery (Das and Xie, 2000). This alludes to a web of interactions occurring amongst proteins which may be important for overall stability.

Furthermore, while cross-linking did not appear to be effective and resulted in the precipitation of the protein and elution of it in the void of the gel filtration run, this may be due to the experimental set-up rather than due to the cross-linker. In the future, it may be worth repeating the cross-linking but performing it at 4 °C as described for the core export apparatus from the T3SS (Kuhlen et al., 2018).

6.4.2 TraG and TraN did not interact

Both a StrepTrap and HisTrap assisted pulldown did not show an interaction occurring between the Strep-II tagged TraG and the hexahistidine tagged TraN. It was previously assumed that such an interaction could not be isolated due to a small number of interacting protein partners occurring in a mating pair. However, this hurdle should be overcome by over-expressing both proteins and thereby hypothetically allowing for an increase in interacting pairs. While the presence of detergents may have obstructed interactions between various Tra proteins in section 5.3.3, this is unlikely given that the region of TraG which is thought to interact with TraN is periplasmic and not membrane bound. Therefore, it should be free of any detergent shell (Lawley et al., 2003).

It may be the case that TraG and TraN do not interact directly or that the interaction is transient. It was proposed by Frankel and colleagues that interaction of TraN with its cognate Omp receptor may relay a signal to the core machinery which then increases efficiency of plasmid transfer (Low et al., 2022). A suitable protein which could play a role in receiving this signal could be TraG which while being involved in Mps is also likely to form a core structural aspect of the conjugative apparatus. How the signal is communicated was not alluded to, but perhaps could occur in a transient manner or through another protein. This may explain why to date, an interaction between TraG and TraN has not been demonstrated in mating cells by cross-linking, immunoprecipitation and/or bacterial two hybrid assays. The outcome of the pull-down described in this Chapter is furthermore in line with this. The periplasmic protein TraU has been suggested to be a putative Mps protein and while individually it was not required for pilus synthesis, it is involved in this function along with the periplasmic complex (Moore et al., 1990, Harris and Silverman, 2004). The possibility of an interaction between TraU and TraG was being explored by Frost and colleagues, however, has not yet been proven (Audette et al., 2007). Similarly, while an interaction between TraN and TraU has been speculated, this still remains to be demonstrated (Bragagnolo et al., 2020).

Another aspect which needs to be considered is how Eex with TraS occurs despite no sustained interaction occurring between TraG with TraN. In one of the models described in section 6.4.1, it was suggested that TraG may recognise the cognate TraS following establishment of the mating pair bridge (Marrero and Waldor, 2005). This was further elaborated by the hypothesis that through interaction with TraN, TraG can protrude through the mating pore and across the outer membrane of the recipient cell where it will then be in the vicinity of TraS (Marrero and Waldor, 2005). Given the lack of interaction observed in the pull-down described in this Chapter between TraN and TraG, it poses the question of how TraG recognises

TraS and therefore, this may occur in a different manner to what has been proposed. Furthermore, from the structure of the pKpQIL TraN and OmpK36, it showed that one TraN molecule extended into one subunit of the trimeric porin in such a way that its uncertain whether another polypeptide could also fit through (Low et al., 2022). Together, this suggests that the role of TraG in Mps and in Eex is likely to occur in another way to what has been hypothesised and makes TraG an interesting protein to further study.

6.4.3 Future perspectives

The results described in this Chapter could act as the basis for future structural work on TraG using cryo-EM and elucidating its oligomeric state. Following optimisation of the TraG sample, it will be amenable to structural determination. The 3D structure will be invaluable to understand how TraG is a fusion of two core VirB proteins but also to understand how one protein can carry out numerous functions.

Chapter 7: Conclusions

Within this thesis, various biochemical approaches and structural studies were carried out to further understand the mechanism underpinning bacterial conjugation. In particular, the following ideas were explored:

- Optimise purification of the F-OMCC such that it is amenable to structural studies using single particle cryo-EM
- Determine the structure of the F-OMCC to better understand its architecture
- Investigate if other Tra/Trb proteins interact with the F-OMCC
- Examine the structure of TraG to gain insight into its oligomeric state

7.1 The overview of the findings

In this thesis, the cryo-EM structure of the F-OMCC, formed by the three proteins TraK, TraV and TraB, was solved and described. The structure was solved at an overall resolution of 3.3 Å. The F-OMCC is formed from two concentric rings with two distinct symmetries; the outer ring has 13-fold symmetry while the inner ring adopts 17-fold symmetry. Together, the two rings form a complex with overall dimensions of 268 Å in diameter and 115 Å in height. The architecture of the F-OMCC was found to be intricate with a total of 69 polypeptides involved: 17 TraBs, 26 TraVs and 26 TraKs. The 17 copies of TraB span the inner ring, of which 13 extends down to shape the outer ring. All 26 TraK polypeptides shape the outer ring together with TraV. However, out of the 26 TraV chains, 17 traverse up to form part of the inner ring. Together, this forms a sophisticated and intertwining network of contacts between the proteins which contribute to the stability of the F-OMCC.

The symmetry mismatch within the complex is accommodated by TraB and TraV which span across both rings. However, direct linkages between the two rings could not be resolved due to those regions of the proteins being untraceable. A contributing factor is that these specific areas showed signs of flexibility. Thus, making the F-OMCC a dynamic complex. Two such regions were observed; one is at the O-layer and the second at the I-layer. At the O-layer, all but a few TraB protomers were without a continuous density that connects the outer ring β1 to the β2 strand at the inner ring. This lack of density was credited to

movement at the O-layer. Interestingly, the exit channel, which is shaped by TraB helices, has a diameter of 50 Å. This makes it too narrow to accommodate the mature pilus. Subsequently, it is likely that the structure represents the complex in the closed state. The observation that TraB facilitates the symmetry mismatch and mediates flexibility at the O-layer may explain how the F-OMCC undergoes structural changes which results in the opening of the exit channel. Hence, allowing conjugation to proceed. It will be interesting to investigate signals which mediates the movement of F-OMCC and how this is linked to the pilus, given that the outer membrane is where the pilus is presumably polymerised. A likely candidate that is involved in this signalling is the ATPases, and so it will be curious to explore the interplay within the F-OMCC, in particular TraB, with the ATPases. An understanding of how the movement of the F-OMCC is linked to the pilus will not only be intriguing mechanistically but may be useful in development of other therapeutics to complement antibiotics, such as bacteriophages. Given that for some bacteriophages the pilus is a point of entry, knowledge on how the pilus is extended will be vital and may be exploited to prolong pilus exposure to the virus.

Similarly, movement was observed at the I-layer where TraV protomers exhibit variable degrees of electron density related to the α1 helix connecting the outer and inner ring. The F-OMCC is the only OMCC solved to date that has TraV traversing a symmetry mismatch. In the other OMCCs, the symmetry mismatch is facilitated by VirB9 and its homologues, alongside VirB10. This prominent role of TraV may be important for the F-T4SS and may impart functional specificity. Furthermore, TraV has been suggested to form an interaction hub for the other Tra/Trb proteins to the F-OMCC (Arutyunov et al., 2010, Koch et al., 2020). Regions of TraV could not be assigned a sequence and furthermore, from a total of 26 polypeptides, 9 $TraV_{NTD}$ were absent in the cryo-EM structure. Hence, it is tempting to speculate whether these areas may participate in interaction with other F-T4SS proteins, and whether the presence of these proteins may stabilise regions of TraV. Such an interaction will be exciting as it will be F-T4SS specific and will provide insight into how the F-OMCC fits in with the rest of the T4SS and may also explain how the dynamics of the complex is coordinated.

Efforts were made to explore potential interactions between the F-OMCC and other Tra/Trb proteins which have been previously suggested in literature (Harris and Silverman, 2004, Koch et al., 2020). One such contact that was explored was between TraH and the F-OMCC. A potential interaction was observed between the proteins. However, the low concentration of the TraH prevented a thorough investigation. Subsequently, this remains to be confirmed. Another interaction which was proposed in this thesis was between TraN and the F-OMCC. The proteins co-eluted in both HisTrap and StrepTrap assisted pull-downs,

which indicates that a potential interaction is occurring. Interestingly, this was only observed when the pull-down was performed with the TraN being expressed on a vector encoding the *trbI* – *traG* stretch of the *tra* operon and co-expressed with the F-OMCC. The co-elution was no longer seen when the TraN and the F-OMCC were independently expressed. This suggests that perhaps the interaction between TraN and F-OMCC occurs via other Tra/Trb proteins. Nonetheless, currently it cannot be ruled out whether the results being observed is due to the proteins or due to the experimental set-up. Future work will focus on clarifying this.

Lastly, the inner membrane protein TraG was studied to gain insight into the oligomeric state it adopts. Based on the gel filtration profile and the corresponding negative stain electron micrographs, the TraG appears to form a high order oligomer, likely a hexamer. While this will require further work and confirmation using other biochemical techniques, it provides the first insight into the structure of TraG. A prominent issue with the TraG sample was lack of stability, with the protein dissociating. This may be improved by addition of stabilising compounds such as glycerol or optimising the cross-linking approach. Furthermore, a preliminary pull-down was unable to isolate an interaction between TraG and TraN. Since contact between the two proteins has been the prevailing mechanism to explain how Mps occurs, this provides an interesting scenario which needs to be further elaborated experimentally.

7.2 Future Outlook

The structure of the F-OMCC and investigation into the potential interactions that may occur between the complex and other proteins, has opened the door to exciting mechanistic insight underpinning conjugation. Symmetry mismatch has often been observed in the core complexes of T4SSs, but the significance of it still remains elusive. Given that the F-OMCC is compositionally smaller than the OMCCs from the expanded T4SS, it provides a great tool to study the biological basis of the symmetry mismatch. The details on the purification of the membrane complex together with the preliminary work on interaction studies will act as the foundation for future studies. Together, the findings of this thesis paves the way for future studies on the F-T4SS to broaden understanding of how the dynamic nature of the F-OMCC facilitates the transfer of DNA from one bacterium to another. From a medical standpoint, this will enable structure led design and development of inhibitors, in an effort to disrupt the nanomachine and prevent the spread of antibiotic resistance.

Bibliography

ACHTMAN, M., MORELLI, G. & SCHWUCHOW, S. 1978. Cell-cell interactions in conjugating Escherichia coli: role of F pili and fate of mating aggregates. *J Bacteriol,* 135**,** 1053-61.

ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L. W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr,* 66**,** 213-21.

ALEGRIA, M. C., SOUZA, D. P., ANDRADE, M. O., DOCENA, C., KHATER, L., RAMOS, C. H., DA SILVA, A. C. & FARAH, C. S. 2005. Identification of new protein-protein interactions involving the products of the chromosome- and plasmid-encoded type IV secretion loci of the phytopathogen Xanthomonas axonopodis pv. citri. *J Bacteriol,* 187**,** 2315-25.

ALVAREZ-MARTINEZ, C. E. & CHRISTIE, P. J. 2009. Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev,* 73**,** 775-808.

ALY, K. A. & BARON, C. 2007. The VirB5 protein localizes to the T-pilus tips in Agrobacterium tumefaciens. *Microbiology (Reading),* 153**,** 3766-3775.

AMIN, H., ILANGOVAN, A. & COSTA, T. R. D. 2021. Architecture of the outer-membrane core complex from a conjugative type IV secretion system. *Nat Commun,* 12**,** 6834.

AMRO, J., BLACK, C., JEMOUAI, Z., ROONEY, N., DANEAULT, C., ZEYTUNI, N., RUIZ, M., BUI, K. H. & BARON, C. 2023. Cryo-EM structure of the Agrobacterium tumefaciens T-pilus reveals the importance of positive charges in the lumen. *Structure,* 31**,** 375-384 e4.

ANTHONY, K. G., KLIMKE, W. A., MANCHAK, J. & FROST, L. S. 1999. Comparison of proteins involved in pilus synthesis and mating pair stabilization from the related plasmids F and R100-1: insights into the mechanism of conjugation. *J Bacteriol,* 181**,** 5149-59.

ARUTYUNOV, D., ARENSON, B., MANCHAK, J. & FROST, L. S. 2010. F plasmid TraF and TraH are components of an outer membrane complex involved in conjugation. *J Bacteriol,* 192**,** 1730-4.

ARUTYUNOV, D. & FROST, L. S. 2013. F conjugation: back to the beginning. *Plasmid,* 70**,** 18-32.

ATMAKURI, K., CASCALES, E., BURTON, O. T., BANTA, L. M. & CHRISTIE, P. J. 2007. Agrobacterium ParA/MinD-like VirC1 spatially coordinates early conjugative DNA transfer reactions. *EMBO J,* 26**,** 2540-51.

AUDETTE, G. F., MANCHAK, J., BEATTY, P., KLIMKE, W. A. & FROST, L. S. 2007. Entry exclusion in F-like plasmids requires intact TraG in the donor that recognizes its cognate TraS in the recipient. *Microbiology (Reading),* 153**,** 442-51.

BABIC, A., LINDNER, A. B., VULIC, M., STEWART, E. J. & RADMAN, M. 2008. Direct visualization of horizontal gene transfer. *Science,* 319**,** 1533-6.

BACKERT, S., FRONZES, R. & WAKSMAN, G. 2008. VirB2 and VirB5 proteins: specialized adhesins in bacterial type-IV secretion systems? *Trends Microbiol,* 16**,** 409-13.

BACKERT, S. & MEYER, T. F. 2006. Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr Opin Microbiol,* 9**,** 207-17.

BACKERT, S. & NAUMANN, M. 2010. What a disorder: proinflammatory signaling pathways induced by Helicobacter pylori. *Trends Microbiol,* 18**,** 479-86.

BAILEY, S., WARD, D., MIDDLETON, R., GROSSMANN, J. G. & ZAMBRYSKI, P. C. 2006. Agrobacterium tumefaciens VirB8 structure reveals potential protein-protein interaction sites. *Proc Natl Acad Sci U S A,* 103**,** 2582-7.

BANTA, L. M., KERR, J. E., CASCALES, E., GIULIANO, M. E., BAILEY, M. E., MCKAY, C., CHANDRAN, V., WAKSMAN, G. & CHRISTIE, P. J. 2011. An Agrobacterium VirB10 mutation conferring a type IV secretion system gating defect. *J Bacteriol,* 193**,** 2566-74.

BARON, C., LLOSA, M., ZHOU, S. & ZAMBRYSKI, P. C. 1997. VirB1, a component of the T-complex transfer machinery of Agrobacterium tumefaciens, is processed to a C-terminal secreted product, VirB1. *J Bacteriol,* 179**,** 1203-10.

BAYER-SANTOS, E., CENENS, W., MATSUYAMA, B. Y., OKA, G. U., DI SESSA, G., MININEL, I. D. V., ALVES, T. L. & FARAH, C. S. 2019. The opportunistic pathogen Stenotrophomonas maltophilia utilizes a type IV secretion system for interbacterial killing. *PLoS Pathog,* 15**,** e1007651.

BECKERS, M., MANN, D. & SACHSE, C. 2021. Structural interpretation of cryo-EM image reconstructions. *Prog Biophys Mol Biol,* 160**,** 26-36.

BERGER, B. R. & CHRISTIE, P. J. 1994. Genetic complementation analysis of the Agrobacterium tumefaciens virB operon: virB2 through virB11 are essential virulence genes. *J Bacteriol,* 176**,** 3646-60.

BERGER, K. H. & ISBERG, R. R. 1993. Two distinct defects in intracellular growth complemented by a single genetic locus in Legionella pneumophila. *Mol Microbiol,* 7**,** 7-19.

BLASER, M. J., PEREZ-PEREZ, G. I., KLEANTHOUS, H., COVER, T. L., PEEK, R. M., CHYOU, P. H., STEMMERMANN, G. N. & NOMURA, A. 1995. Infection with Helicobacter pylori strains possessing cagA is associated with an increased risk of developing adenocarcinoma of the stomach. *Cancer Res,* 55**,** 2111-5.

BOCK, D., HUSLER, D., STEINER, B., MEDEIROS, J. M., WELIN, A., RADOMSKA, K. A., HARDT, W. D., PILHOFER, M. & HILBI, H. 2021. The Polar Legionella Icm/Dot T4SS Establishes Distinct Contact Sites with the Pathogen Vacuole Membrane. *mBio,* 12**,** e0218021.

BRADLEY, D. E. 1980a. Morphological and serological relationships of conjugative pili. *Plasmid,* 4**,** 155-69.

BRADLEY, D. E. 1980b. Determination of pili by conjugative bacterial drug resistance plasmids of incompatibility groups B, C, H, J, K, M, V, and X. *J Bacteriol,* 141**,** 828-37.

BRAGAGNOLO, N., RODRIGUEZ, C., SAMARI-KERMANI, N., FOURS, A., KOROUZHDEHI, M., LYSENKO, R. & AUDETTE, G. F. 2020. Protein Dynamics in F-like Bacterial Conjugation. *Biomedicines,* 8.

BURSTEIN, D., ZUSMAN, T., DEGTYAR, E., VINER, R., SEGAL, G. & PUPKO, T. 2009. Genome-scale identification of Legionella pneumophila effectors using a machine learning approach. *PLoS Pathog,* 5**,** e1000508.

CAMBRONNE, E. D. & ROY, C. R. 2007. The Legionella pneumophila IcmSW complex interacts with multiple Dot/Icm effectors to facilitate type IV translocation. *PLoS Pathog,* 3**,** e188.

CASCALES, E., ATMAKURI, K., SARKAR, M. K. & CHRISTIE, P. J. 2013. DNA substrate-induced activation of the Agrobacterium VirB/VirD4 type IV secretion system. *J Bacteriol,* 195**,** 2691-704.

CASCALES, E. & CHRISTIE, P. J. 2004a. Agrobacterium VirB10, an ATP energy sensor required for type IV secretion. *Proc Natl Acad Sci U S A,* 101**,** 17228-33.

CASCALES, E. & CHRISTIE, P. J. 2004b. Definition of a bacterial type IV secretion pathway for a DNA substrate. *Science,* 304**,** 1170-3.

CASU, B., MARY, C., SVERZHINSKY, A., FOUILLEN, A., NANCI, A. & BARON, C. 2018. VirB8 homolog TraE from plasmid pKM101 forms a hexameric ring structure and interacts with the VirB6 homolog TraD. *Proc Natl Acad Sci U S A,* 115**,** 5950-5955.

CHAE, P. S., RASMUSSEN, S. G., RANA, R. R., GOTFRYD, K., KRUSE, A. C., MANGLIK, A., CHO, K. H., NURVA, S., GETHER, U., GUAN, L., LOLAND, C. J., BYRNE, B., KOBILKA, B. K. & GELLMAN, S. H. 2012. A new class of amphiphiles bearing rigid hydrophobic groups for solubilization and stabilization of membrane proteins. *Chemistry,* 18**,** 9485-90.

CHANDRAN DARBARI, V. & WAKSMAN, G. 2015. Structural Biology of Bacterial Type IV Secretion Systems. *Annu Rev Biochem,* 84**,** 603-29.

CHANDRAN, V., FRONZES, R., DUQUERROY, S., CRONIN, N., NAVAZA, J. & WAKSMAN, G. 2009. Structure of the outer membrane complex of a type IV secretion system. *Nature,* 462**,** 1011-5.
CHANG, Y. W., RETTBERG, L. A., TREUNER-LANGE, A., IWASA, J., SOGAARD-ANDERSEN, L. & JENSEN, G. J. 2016. Architecture of the type IVa pilus machine. *Science,* 351**,** aad2001.
CHANG, Y. W., SHAFFER, C. L., RETTBERG, L. A., GHOSAL, D. & JENSEN, G. J. 2018. In Vivo Structures of the Helicobacter pylori cag Type IV Secretion System. *Cell Rep,* 23**,** 673-681.
CHAPTAL, V., DELOLME, F., KILBURG, A., MAGNARD, S., MONTIGNY, C., PICARD, M., PRIER, C., MONTICELLI, L., BORNERT, O., AGEZ, M., RAVAUD, S., ORELLE, C., WAGNER, R., JAWHARI, A., BROUTIN, I., PEBAY-PEYROULA, E., JAULT, J. M., KABACK, H. R., LE MAIRE, M. & FALSON, P. 2017. Quantification of Detergents Complexed with Membrane Proteins. *Sci Rep,* 7**,** 41751.
CHEAH, K. C. & SKURRAY, R. 1986. The F plasmid carries an IS3 insertion within finO. *J Gen Microbiol,* 132**,** 3269-75.
CHERNYATINA, A. A. & LOW, H. H. 2019. Core architecture of a bacterial type II secretion system. *Nat Commun,* 10**,** 5437.
CHETRIT, D., HU, B., CHRISTIE, P. J., ROY, C. R. & LIU, J. 2018. A unique cytoplasmic ATPase complex defines the Legionella pneumophila type IV secretion channel. *Nat Microbiol,* 3**,** 678-686.
CHRISTIE, P. J. 2016. The Mosaic Type IV Secretion Systems. *EcoSal Plus,* 7.
CHRISTIE, P. J. & VOGEL, J. P. 2000. Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol,* 8**,** 354-60.
CHRISTIE, P. J., WHITAKER, N. & GONZALEZ-RIVERA, C. 2014. Mechanism and structure of the bacterial type IV secretion systems. *Biochim Biophys Acta,* 1843**,** 1578-91.
CHUNG, J. M., SHEEDLO, M. J., CAMPBELL, A. M., SAWHNEY, N., FRICK-CHENG, A. E., LACY, D. B., COVER, T. L. & OHI, M. D. 2019. Structure of the Helicobacter pylori Cag type IV secretion system. *Elife,* 8.
CITOVSKY, V., GURALNICK, B., SIMON, M. N. & WALL, J. S. 1997. The molecular structure of agrobacterium VirE2-single stranded DNA complexes involved in nuclear import. *J Mol Biol,* 271**,** 718-27.
CLARKE, M., MADDERA, L., HARRIS, R. L. & SILVERMAN, P. M. 2008. F-pili dynamics by live-cell imaging. *Proc Natl Acad Sci U S A,* 105**,** 17978-81.
COLOM, J., BATISTA, D., BAIG, A., TANG, Y., LIU, S., YUAN, F., BELKHIRI, A., MARCELINO, L., BARBOSA, F., RUBIO, M., ATTERBURY, R., BERCHIERI, A. & BARROW, P. 2019. Sex pilus specific bacteriophage to drive bacterial population towards antibiotic sensitivity. *Sci Rep,* 9**,** 12616.
COSTA, T. R., FELISBERTO-RODRIGUES, C., MEIR, A., PREVOST, M. S., REDZEJ, A., TROKTER, M. & WAKSMAN, G. 2015. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat Rev Microbiol,* 13**,** 343-59.
COSTA, T. R. D., HARB, L., KHARA, P., ZENG, L., HU, B. & CHRISTIE, P. J. 2021. Type IV secretion systems: Advances in structure, function, and activation. *Mol Microbiol,* 115**,** 436-452.
COSTA, T. R. D., ILANGOVAN, A., UKLEJA, M., REDZEJ, A., SANTINI, J. M., SMITH, T. K., EGELMAN, E. H. & WAKSMAN, G. 2016. Structure of the Bacterial Sex F Pilus Reveals an Assembly of a Stoichiometric Protein-Phospholipid Complex. *Cell,* 166**,** 1436-1444 e10.
COVER, T. L. 2016. Helicobacter pylori Diversity and Gastric Cancer Risk. *mBio,* 7**,** e01869-15.
COVER, T. L., LACY, D. B. & OHI, M. D. 2020. The Helicobacter pylori Cag Type IV Secretion System. *Trends Microbiol,* 28**,** 682-695.
DARBARI, V. C., CICCONE, J., PATEL, J. S., ISLAM, B., AGARWAL, P. K. & HAIDER, S. 2020. Electrostatic Switching Controls Channel Dynamics of the Sensor Protein VirB10 in A. tumefaciens Type IV Secretion System. *ACS Omega,* 5**,** 3271-3281.
DAS, A. & XIE, Y. H. 2000. The Agrobacterium T-DNA transport pore proteins VirB8, VirB9, and VirB10 interact with one another. *J Bacteriol,* 182**,** 758-63.

DE LA CRUZ, F., FROST, L. S., MEYER, R. J. & ZECHNER, E. L. 2010. Conjugative DNA metabolism in Gram-negative bacteria. *FEMS Microbiol Rev,* 34**,** 18-40.

DENG, W., CHEN, L., PENG, W. T., LIANG, X., SEKIGUCHI, S., GORDON, M. P., COMAI, L. & NESTER, E. W. 1999. VirE1 is a specific molecular chaperone for the exported single-stranded-DNA-binding protein VirE2 in Agrobacterium. *Mol Microbiol,* 31**,** 1795-807.

DIX, S. R., OWEN, H. J., SUN, R., AHMAD, A., SHASTRI, S., SPIEWAK, H. L., MOSBY, D. J., HARRIS, M. J., BATTERS, S. L., BROOKER, T. A., TZOKOV, S. B., SEDELNIKOVA, S. E., BAKER, P. J., BULLOUGH, P. A., RICE, D. W. & THOMAS, M. S. 2018. Structural insights into the function of type VI secretion system TssA subunits. *Nat Commun,* 9**,** 4765.

DORAN, T. J., LOH, S. M., FIRTH, N. & SKURRAY, R. A. 1994. Molecular analysis of the F plasmid traVR region: traV encodes a lipoprotein. *J Bacteriol,* 176**,** 4182-6.

DUBOCHET, J., ADRIAN, M., CHANG, J.-J., HOMO, J.-C., LEPAULT, J., MCDOWALL, A. W. AND SCHULTZ, P. (1988) “Cryo-electron microscopy of vitrified specimens,” Quarterly Reviews of Biophysics. Cambridge University Press, 21(2), pp. 129–228.

DURIE, C. L., SHEEDLO, M. J., CHUNG, J. M., BYRNE, B. G., SU, M., KNIGHT, T., SWANSON, M., LACY, D. B. & OHI, M. D. 2020. Structural analysis of the Legionella pneumophila Dot/Icm type IV secretion system core complex. *Elife,* 9.

EISENBRANDT, R., KALKUM, M., LAI, E. M., LURZ, R., KADO, C. I. & LANKA, E. 1999. Conjugative pili of IncP plasmids, and the Ti plasmid T pilus are composed of cyclic subunits. *J Biol Chem,* 274**,** 22548-55.

ELHENAWY, W., HORDIENKO, S., GOULD, S., OBERC, A. M., TSAI, C. N., HUBBARD, T. P., WALDOR, M. K. & COOMBES, B. K. 2021. High-throughput fitness screening and transcriptomics identify a role for a type IV secretion system in the pathogenesis of Crohn's disease-associated Escherichia coli. *Nat Commun,* 12**,** 2032.

EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr,* 66**,** 486-501.

FAN, X., WANG, J., ZHANG, X., YANG, Z., ZHANG, J. C., ZHAO, L., PENG, H. L., LEI, J. & WANG, H. W. 2019. Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. *Nat Commun,* 10**,** 2386.

FERNANDEZ, D., DANG, T. A., SPUDICH, G. M., ZHOU, X. R., BERGER, B. R. & CHRISTIE, P. J. 1996. The Agrobacterium tumefaciens virB7 gene product, a proposed component of the T-complex transport apparatus, is a membrane-associated lipoprotein exposed at the periplasmic surface. *J Bacteriol,* 178**,** 3156-67.

FIRTH, N. & SKURRAY, R. 1992. Characterization of the F plasmid bifunctional conjugation gene, traG. *Mol Gen Genet,* 232**,** 145-53.

FISCHER, W. 2011. Assembly and molecular mode of action of the Helicobacter pylori Cag type IV secretion apparatus. *FEBS J,* 278**,** 1203-12.

FISCHER, W., TEGTMEYER, N., STINGL, K. & BACKERT, S. 2020. Four Chromosomal Type IV Secretion Systems in Helicobacter pylori: Composition, Structure and Function. *Front Microbiol,* 11**,** 1592.

FRANKEN, L. E., GRUNEWALD, K., BOEKEMA, E. J. & STUART, M. C. A. 2020. A Technical Introduction to Transmission Electron Microscopy for Soft-Matter: Imaging, Possibilities, Choices, and Technical Developments. *Small,* 16**,** e1906198.

FRASER, D. W., TSAI, T. R., ORENSTEIN, W., PARKIN, W. E., BEECHAM, H. J., SHARRAR, R. G., HARRIS, J., MALLISON, G. F., MARTIN, S. M., MCDADE, J. E., SHEPARD, C. C. & BRACHMAN, P. S. 1977. Legionnaires' disease: description of an epidemic of pneumonia. *N Engl J Med,* 297**,** 1189-97.

FRICK-CHENG, A. E., PYBURN, T. M., VOSS, B. J., MCDONALD, W. H., OHI, M. D. & COVER, T. L. 2016. Molecular and Structural Analysis of the Helicobacter pylori cag Type IV Secretion System Core Complex. *mBio,* 7**,** e02001-15.

FRONZES, R., SCHAFER, E., WANG, L., SAIBIL, H. R., ORLOVA, E. V. & WAKSMAN, G. 2009. Structure of a type IV secretion system core complex. *Science,* 323**,** 266-8.
FROST, L. S., IPPEN-IHLER, K. & SKURRAY, R. A. 1994. Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol Rev,* 58**,** 162-210.
FROST, L. S. & KORAIMANN, G. 2010. Regulation of bacterial conjugation: balancing opportunity with adversity. *Future Microbiol,* 5**,** 1057-71.
GELVIN, S. B. 2012. Traversing the Cell: Agrobacterium T-DNA's Journey to the Host Genome. *Front Plant Sci,* 3**,** 52.
GENETELLO, C., VAN LAREBEKE, N., HOLSTERS, M., DE PICKER, A., VAN MONTAGU, M. & SCHELL, J. 1977. Ti plasmids of Agrobacterium as conjugative plasmids. *Nature,* 265**,** 561-3.
GHIGO, J. M. 2001. Natural conjugative plasmids induce bacterial biofilm development. *Nature,* 412**,** 442-5.
GHOSAL, D., CHANG, Y. W., JEONG, K. C., VOGEL, J. P. & JENSEN, G. J. 2017. In situ structure of the Legionella Dot/Icm type IV secretion system by electron cryotomography. *EMBO Rep,* 18**,** 726-732.
GHOSAL, D., JEONG, K. C., CHANG, Y. W., GYORE, J., TENG, L., GARDNER, A., VOGEL, J. P. & JENSEN, G. J. 2019. Molecular architecture, polar targeting and biogenesis of the Legionella Dot/Icm T4SS. *Nat Microbiol,* 4**,** 1173-1182.
GILLESPIE, J. J., PHAN, I. Q., SCHEIB, H., SUBRAMANIAN, S., EDWARDS, T. E., LEHMAN, S. S., PIITULAINEN, H., RAHMAN, M. S., RENNOLL-BANKERT, K. E., STAKER, B. L., TAIRA, S., STACY, R., MYLER, P. J., AZAD, A. F. & PULLIAINEN, A. T. 2015. Structural Insight into How Bacteria Prevent Interference between Multiple Divergent Type IV Secretion Systems. *mBio,* 6**,** e01867-15.
GODDARD, T. D., HUANG, C. C., MENG, E. C., PETTERSEN, E. F., COUCH, G. S., MORRIS, J. H. & FERRIN, T. E. 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci,* 27**,** 14-25.
GOMIS-RUTH, F. X., MONCALIAN, G., PEREZ-LUQUE, R., GONZALEZ, A., CABEZON, E., DE LA CRUZ, F. & COLL, M. 2001. The bacterial conjugation protein TrwB resembles ring helicases and F1-ATPase. *Nature,* 409**,** 637-41.
GREEN, E. R. & MECSAS, J. 2016. Bacterial Secretion Systems: An Overview. *Microbiol Spectr,* 4.
GROHMANN, E., CHRISTIE, P. J., WAKSMAN, G. & BACKERT, S. 2018. Type IV secretion in Gram-negative and Gram-positive bacteria. *Mol Microbiol,* 107**,** 455-471.
GUZMAN, L. M., BELIN, D., CARSON, M. J. & BECKWITH, J. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol,* 177**,** 4121-30.
HALLGREN, J., TSIRIGOS, K. D., PEDERSEN, M. D., ALMAGRO ARMENTEROS, J. J., MARCATILI, P., NIELSEN, H., KROGH, A. & WINTHER, O. 2022. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv***,** 2022.04.08.487609.
HARE, S., BAYLISS, R., BARON, C. & WAKSMAN, G. 2006. A large domain swap in the VirB11 ATPase of Brucella suis leaves the hexameric assembly intact. *J Mol Biol,* 360**,** 56-66.
HARRIS, R. L., HOMBS, V. & SILVERMAN, P. M. 2001. Evidence that F-plasmid proteins TraV, TraK and TraB assemble into an envelope-spanning structure in Escherichia coli. *Mol Microbiol,* 42**,** 757-66.
HARRIS, R. L. & SILVERMAN, P. M. 2002. Roles of internal cysteines in the function, localization, and reactivity of the TraV outer membrane lipoprotein encoded by the F plasmid. *J Bacteriol,* 184**,** 3126-9.
HARRIS, R. L. & SILVERMAN, P. M. 2004. Tra proteins characteristic of F-like type IV secretion systems constitute an interaction group by yeast two-hybrid analysis. *J Bacteriol,* 186**,** 5480-5.
HATAKEYAMA, M. 2004. Oncogenic mechanisms of the Helicobacter pylori CagA protein. *Nat Rev Cancer,* 4**,** 688-94.

HENDERSON, R. 1995. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q Rev Biophys,* 28**,** 171-93.
HOFREUTER, D., ODENBREIT, S. & HAAS, R. 2001. Natural transformation competence in Helicobacter pylori is mediated by the basic components of a type IV secretion system. *Mol Microbiol,* 41**,** 379-91.
HORWITZ, M. A. 1983. The Legionnaires' disease bacterium (Legionella pneumophila) inhibits phagosome-lysosome fusion in human monocytes. *J Exp Med,* 158**,** 2108-26.
HORWITZ, M. A. & MAXFIELD, F. R. 1984. Legionella pneumophila inhibits acidification of its phagosome in human monocytes. *J Cell Biol,* 99**,** 1936-43.
HOSPENTHAL, M. K., COSTA, T. R. D. & WAKSMAN, G. 2017. A comprehensive guide to pilus biogenesis in Gram-negative bacteria. *Nat Rev Microbiol,* 15**,** 365-379.
HOWARD, E. A., ZUPAN, J. R., CITOVSKY, V. & ZAMBRYSKI, P. C. 1992. The VirD2 protein of A. tumefaciens contains a C-terminal bipartite nuclear localization signal: implications for nuclear uptake of DNA in plant cells. *Cell,* 68**,** 109-18.
HU, B., KHARA, P. & CHRISTIE, P. J. 2019a. Structural bases for F plasmid conjugation and F pilus biogenesis in Escherichia coli. *Proc Natl Acad Sci U S A,* 116**,** 14222-14227.
HU, B., KHARA, P., SONG, L., LIN, A. S., FRICK-CHENG, A. E., HARVEY, M. L., COVER, T. L. & CHRISTIE, P. J. 2019b. In Situ Molecular Architecture of the Helicobacter pylori Cag Type IV Secretion System. *mBio,* 10.
HU, B., LARA-TEJERO, M., KONG, Q., GALAN, J. E. & LIU, J. 2017. In Situ Molecular Architecture of the Salmonella Type III Secretion Machine. *Cell,* 168**,** 1065-1074 e10.
HUANG, L., BOYD, D., AMYOT, W. M., HEMPSTEAD, A. D., LUO, Z. Q., O'CONNOR, T. J., CHEN, C., MACHNER, M., MONTMINY, T. & ISBERG, R. R. 2011. The E Block motif is associated with Legionella pneumophila translocated substrates. *Cell Microbiol,* 13**,** 227-45.
ILANGOVAN, A., KAY, C. W. M., ROIER, S., EL MKAMI, H., SALVADORI, E., ZECHNER, E. L., ZANETTI, G. & WAKSMAN, G. 2017. Cryo-EM Structure of a Relaxase Reveals the Molecular Basis of DNA Unwinding during Bacterial Conjugation. *Cell,* 169**,** 708-721 e12.
ISAAC, D. T. & ISBERG, R. 2014. Master manipulators: an update on Legionella pneumophila Icm/Dot translocated substrates and their host targets. *Future Microbiol,* 9**,** 343-59.
JAKUBOWSKI, S. J., KERR, J. E., GARZA, I., KRISHNAMOORTHY, V., BAYLISS, R., WAKSMAN, G. & CHRISTIE, P. J. 2009. Agrobacterium VirB10 domain requirements for type IV secretion and T pilus biogenesis. *Mol Microbiol,* 71**,** 779-94.
JEONG, K. C., GHOSAL, D., CHANG, Y. W., JENSEN, G. J. & VOGEL, J. P. 2017. Polar delivery of Legionella type IV secretion system substrates is essential for virulence. *Proc Natl Acad Sci U S A,* 114**,** 8077-8082.
JUHAS, M., CROOK, D. W. & HOOD, D. W. 2008. Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. *Cell Microbiol,* 10**,** 2377-86.
JUMPER, J., EVANS, R., PRITZEL, A., GREEN, T., FIGURNOV, M., RONNEBERGER, O., TUNYASUVUNAKOOL, K., BATES, R., ZIDEK, 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. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature,* 596**,** 583-589.
KELLEY, L. A., MEZULIS, S., YATES, C. M., WASS, M. N. & STERNBERG, M. J. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc,* 10**,** 845-58.
KHARA, P., SONG, L., CHRISTIE, P. J. & HU, B. 2021. In Situ Visualization of the pKM101-Encoded Type IV Secretion System Reveals a Highly Symmetric ATPase Energy Center. *mBio,* 12**,** e0246521.

KIM, H., KUBORI, T., YAMAZAKI, K., KWAK, M. J., PARK, S. Y., NAGAI, H., VOGEL, J. P. & OH, B. H. 2020. Structural basis for effector protein recognition by the Dot/Icm Type IVB coupling protein complex. *Nat Commun,* 11**,** 2623.

KISHIDA, K., BOSSERMAN, R. E., HARB, L., KHARA, P., SONG, L., HU, B., ZENG, L. & CHRISTIE, P. J. 2022. Contributions of F-specific subunits to the F plasmid-encoded type IV secretion system and F pilus. *Mol Microbiol,* 117**,** 1275-1290.

KLIMKE, W. A. & FROST, L. S. 1998. Genetic analysis of the role of the transfer gene, traN, of the F and R100-1 plasmids in mating pair stabilization during conjugation. *J Bacteriol,* 180**,** 4036-43.

KLIMKE, W. A., RYPIEN, C. D., KLINGER, B., KENNEDY, R. A., RODRIGUEZ-MAILLARD, J. M. & FROST, L. S. 2005. The mating pair stabilization protein, TraN, of the F plasmid is an outer-membrane protein with two regions that are important for its function in conjugation. *Microbiology (Reading),* 151**,** 3527-3540.

KOCH, B., CALLAGHAN, M. M., TELLECHEA-LUZARDO, J., SEEGER, A. Y., DILLARD, J. P. & KRASNOGOR, N. 2020. Protein interactions within and between two F-type type IV secretion systems. *Mol Microbiol,* 114**,** 823-838.

KOHLER, P. L., CHAN, Y. A., HACKETT, K. T., TURNER, N., HAMILTON, H. L., CLOUD-HANSEN, K. A. & DILLARD, J. P. 2013. Mating pair formation homologue TraG is a variable membrane protein essential for contact-independent type IV secretion of chromosomal DNA by Neisseria gonorrhoeae. *J Bacteriol,* 195**,** 1666-79.

KORAIMANN, G. 2018. Spread and Persistence of Virulence and Antibiotic Resistance Genes: A Ride on the F Plasmid Conjugation Module. *EcoSal Plus,* 8.

KREIDA, S., NARITA, A., JOHNSON, M. D., TOCHEVA, E. I., DAS, A., GHOSAL, D. & JENSEN, G. J. 2023. Cryo-EM structure of the Agrobacterium tumefaciens T4SS-associated T-pilus reveals stoichiometric protein-phospholipid assembly. *Structure,* 31**,** 385-394 e4.

KÜHLBRANDT, W. 2014. The Resolution Revolution. *Science,* 343**,** 1443-1444.

KUHLEN, L., ABRUSCI, P., JOHNSON, S., GAULT, J., DEME, J., CAESAR, J., DIETSCHE, T., MEBRHATU, M. T., GANIEF, T., MACEK, B., WAGNER, S., ROBINSON, C. V. & LEA, S. M. 2018. Structure of the core of the type III secretion system export apparatus. *Nat Struct Mol Biol,* 25**,** 583-590.

KWAK, M. J., KIM, J. D., KIM, H., KIM, C., BOWMAN, J. W., KIM, S., JOO, K., LEE, J., JIN, K. S., KIM, Y. G., LEE, N. K., JUNG, J. U. & OH, B. H. 2017. Architecture of the type IV coupling protein complex of Legionella pneumophila. *Nat Microbiol,* 2**,** 17114.

LAI, E. M., CHESNOKOVA, O., BANTA, L. M. & KADO, C. I. 2000. Genetic and environmental factors affecting T-pilin export and T-pilus biogenesis in relation to flagellation of Agrobacterium tumefaciens. *J Bacteriol,* 182**,** 3705-16.

LAI, E. M. & KADO, C. I. 1998. Processed VirB2 is the major subunit of the promiscuous pilus of Agrobacterium tumefaciens. *J Bacteriol,* 180**,** 2711-7.

LANG, S., GRUBER, K., MIHAJLOVIC, S., ARNOLD, R., GRUBER, C. J., STEINLECHNER, S., JEHL, M. A., RATTEI, T., FROHLICH, K. U. & ZECHNER, E. L. 2010. Molecular recognition determinants for type IV secretion of diverse families of conjugative relaxases. *Mol Microbiol,* 78**,** 1539-55.

LARREA, D., DE PAZ, H. D., MATILLA, I., GUZMAN-HERRADOR, D. L., LASSO, G., DE LA CRUZ, F., CABEZON, E. & LLOSA, M. 2017. Substrate translocation involves specific lysine residues of the central channel of the conjugative coupling protein TrwB. *Mol Genet Genomics,* 292**,** 1037-1049.

LAWLEY, T. D., KLIMKE, W. A., GUBBINS, M. J. & FROST, L. S. 2003. F factor conjugation is a true type IV secretion system. *FEMS Microbiol Lett,* 224**,** 1-15.

LEDERBERG, J. & TATUM, E. L. 1946. Gene recombination in Escherichia coli. *Nature,* 158**,** 558.

LEE, C. R., LEE, J. H., PARK, K. S., KIM, Y. B., JEONG, B. C. & LEE, S. H. 2016. Global Dissemination of Carbapenemase-Producing Klebsiella pneumoniae: Epidemiology, Genetic Context, Treatment Options, and Detection Methods. *Front Microbiol,* 7**,** 895.

LIU, X., KHARA, P., BAKER, M. L., CHRISTIE, P. J. & HU, B. 2022. Structure of a type IV secretion system core complex encoded by multi-drug resistance F plasmids. *Nat Commun,* 13**,** 379.

LLOSA, M. & ALKORTA, I. 2017. Coupling Proteins in Type IV Secretion. *Curr Top Microbiol Immunol,* 413**,** 143-168.

LLOSA, M., BOLLAND, S. & DE LA CRUZ, F. 1994. Genetic organization of the conjugal DNA processing region of the IncW plasmid R388. *J Mol Biol,* 235**,** 448-64.

LLOSA, M., ZUPAN, J., BARON, C. & ZAMBRYSKI, P. 2000. The N- and C-terminal portions of the Agrobacterium VirB1 protein independently enhance tumorigenesis. *J Bacteriol,* 182**,** 3437-45.

LOCKWOOD, D. C., AMIN, H., COSTA, T. R. D. & SCHROEDER, G. N. 2022. The Legionella pneumophila Dot/Icm type IV secretion system and its effectors. *Microbiology (Reading),* 168.

LOW, H. H., GUBELLINI, F., RIVERA-CALZADA, A., BRAUN, N., CONNERY, S., DUJEANCOURT, A., LU, F., REDZEJ, A., FRONZES, R., ORLOVA, E. V. & WAKSMAN, G. 2014. Structure of a type IV secretion system. *Nature,* 508**,** 550-553.

LOW, W. W., WONG, J. L. C., BELTRAN, L. C., SEDDON, C., DAVID, S., KWONG, H. S., BIZEAU, T., WANG, F., PENA, A., COSTA, T. R. D., PHAM, B., CHEN, M., EGELMAN, E. H., BEIS, K. & FRANKEL, G. 2022. Mating pair stabilization mediates bacterial conjugation species specificity. *Nat Microbiol,* 7**,** 1016-1027.

LU, J., MANCHAK, J., KLIMKE, W., DAVIDSON, C., FIRTH, N., SKURRAY, R. A. & FROST, L. S. 2002. Analysis and characterization of the IncFV plasmid pED208 transfer region. *Plasmid,* 48**,** 24-37.

LU, J., PENG, Y., WAN, S., FROST, L. S., RAIVIO, T. & GLOVER, J. N. M. 2019. Cooperative Function of TraJ and ArcA in Regulating the F Plasmid tra Operon. *J Bacteriol,* 201.

LU, J., WONG, J. J., EDWARDS, R. A., MANCHAK, J., FROST, L. S. & GLOVER, J. N. 2008. Structural basis of specific TraD-TraM recognition during F plasmid-mediated bacterial conjugation. *Mol Microbiol,* 70**,** 89-99.

LUSIAK-SZELACHOWSKA, M., MIEDZYBRODZKI, R., DRULIS-KAWA, Z., CATER, K., KNEZEVIC, P., WINOGRADOW, C., AMARO, K., JONCZYK-MATYSIAK, E., WEBER-DABROWSKA, B., REKAS, J. & GORSKI, A. 2022. Bacteriophages and antibiotic interactions in clinical practice: what we have learned so far. *J Biomed Sci,* 29**,** 23.

MACE, K., VADAKKEPAT, A. K., REDZEJ, A., LUKOYANOVA, N., OOMEN, C., BRAUN, N., UKLEJA, M., LU, F., COSTA, T. R. D., ORLOVA, E. V., BAKER, D., CONG, Q. & WAKSMAN, G. 2022. Cryo-EM structure of a type IV secretion system. *Nature,* 607**,** 191-196.

MAFFEI, B., FRANCETIC, O. & SUBTIL, A. 2017. Tracking Proteins Secreted by Bacteria: What's in the Toolbox? *Frontiers in Cellular and Infection Microbiology,* 7.

MAGORI, S. & CITOVSKY, V. 2011. Agrobacterium counteracts host-induced degradation of its effector F-box protein. *Sci Signal,* 4**,** ra69.

MANEEWANNAKUL, S., MANEEWANNAKUL, K. & IPPEN-IHLER, K. 1991. Characterization of trbC, a new F plasmid tra operon gene that is essential to conjugative transfer. *J Bacteriol,* 173**,** 3872-8.

MANEEWANNAKUL, S., MANEEWANNAKUL, K. & IPPEN-IHLER, K. 1992. Characterization, localization, and sequence of F transfer region products: the pilus assembly gene product TraW and a new product, TrbI. *J Bacteriol,* 174**,** 5567-74.

MANNING, P. A., MORELLI, G. & ACHTMAN, M. 1981. traG protein of the F sex factor of Escherichia coli K-12 and its role in conjugation. *Proc Natl Acad Sci U S A,* 78**,** 7487-91.

MANOIL, C. & ROSENBUSCH, J. P. 1982. Conjugation-deficient mutants of Escherichia coli distinguish classes of functions of the outer membrane OmpA protein. *Mol Gen Genet,* 187**,** 148-56.

MARRA, A., BLANDER, S. J., HORWITZ, M. A. & SHUMAN, H. A. 1992. Identification of a Legionella pneumophila locus required for intracellular multiplication in human macrophages. *Proc Natl Acad Sci U S A,* 89**,** 9607-11.

MARRERO, J. & WALDOR, M. K. 2005. Interactions between inner membrane proteins in donor and recipient cells limit conjugal DNA transfer. *Dev Cell,* 8**,** 963-70.

MCCLAIN, M. S., VOSS, B. J. & COVER, T. L. 2020. Lipoprotein Processing and Sorting in Helicobacter pylori. *mBio,* 11.

MCDADE, J. E., SHEPARD, C. C., FRASER, D. W., TSAI, T. R., REDUS, M. A. & DOWDLE, W. R. 1977. Legionnaires' disease: isolation of a bacterium and demonstration of its role in other respiratory disease. *N Engl J Med,* 297**,** 1197-203.

MCGANN, P., SNESRUD, E., MAYBANK, R., COREY, B., ONG, A. C., CLIFFORD, R., HINKLE, M., WHITMAN, T., LESHO, E. & SCHAECHER, K. E. 2016. Escherichia coli Harboring mcr-1 and blaCTX-M on a Novel IncF Plasmid: First Report of mcr-1 in the United States. *Antimicrob Agents Chemother,* 60**,** 4420-1.

MEIR, A., CHETRIT, D., LIU, L., ROY, C. R. & WAKSMAN, G. 2018. Legionella DotM structure reveals a role in effector recruiting to the Type 4B secretion system. *Nat Commun,* 9**,** 507.

MEIR, A., MACE, K., LUKOYANOVA, N., CHETRIT, D., HOSPENTHAL, M. K., REDZEJ, A., ROY, C. & WAKSMAN, G. 2020. Mechanism of effector capture and delivery by the type IV secretion system from Legionella pneumophila. *Nat Commun,* 11**,** 2864.

MOORE, D., MANEEWANNAKUL, K., MANEEWANNAKUL, S., WU, J. H., IPPEN-IHLER, K. & BRADLEY, D. E. 1990. Characterization of the F-plasmid conjugative transfer gene traU. *J Bacteriol,* 172**,** 4263-70.

MOORE, D., SOWA, B. A. & IPPEN-IHLER, K. 1981. The effect of tra mutations on the synthesis of the F-pilin membrane polypeptide. *Mol Gen Genet,* 184**,** 260-4.

MOYNIE, L., MILENKOVIC, S., MISLIN, G. L. A., GASSER, V., MALLOCI, G., BACO, E., MCCAUGHAN, R. P., PAGE, M. G. P., SCHALK, I. J., CECCARELLI, M. & NAISMITH, J. H. 2019. The complex of ferric-enterobactin with its transporter from Pseudomonas aeruginosa suggests a two-site model. *Nat Commun,* 10**,** 3673.

MURRAY, C. J. L., IKUTA, K. S., SHARARA, F., SWETSCHINSKI, L., ROBLES AGUILAR, G., GRAY, A., HAN, C., BISIGNANO, C., RAO, P., WOOL, E., JOHNSON, S. C., BROWNE, A. J., CHIPETA, M. G., FELL, F., HACKETT, S., HAINES-WOODHOUSE, G., KASHEF HAMADANI, B. H., KUMARAN, E. A. P., MCMANIGAL, B., AGARWAL, R., AKECH, S., ALBERTSON, S., AMUASI, J., ANDREWS, J., ARAVKIN, A., ASHLEY, E., BAILEY, F., BAKER, S., BASNYAT, B., BEKKER, A., BENDER, R., BETHOU, A., BIELICKI, J., BOONKASIDECHA, S., BUKOSIA, J., CARVALHEIRO, C., CASTAÑEDA-ORJUELA, C., CHANSAMOUTH, V., CHAURASIA, S., CHIURCHIÙ, S., CHOWDHURY, F., COOK, A. J., COOPER, B., CRESSEY, T. R., CRIOLLO-MORA, E., CUNNINGHAM, M., DARBOE, S., DAY, N. P. J., DE LUCA, M., DOKOVA, K., DRAMOWSKI, A., DUNACHIE, S. J., ECKMANNS, T., EIBACH, D., EMAMI, A., FEASEY, N., FISHER-PEARSON, N., FORREST, K., GARRETT, D., GASTMEIER, P., GIREF, A. Z., GREER, R. C., GUPTA, V., HALLER, S., HASELBECK, A., HAY, S. I., HOLM, M., HOPKINS, S., IREGBU, K. C., JACOBS, J., JAROVSKY, D., JAVANMARDI, F., KHORANA, M., KISSOON, N., KOBEISSI, E., KOSTYANEV, T., KRAPP, F., KRUMKAMP, R., KUMAR, A., KYU, H. H., LIM, C., LIMMATHUROTSAKUL, D., LOFTUS, M. J., LUNN, M., MA, J., MTURI, N., MUNERA-HUERTAS, T., MUSICHA, P., MUSSI-PINHATA, M. M., NAKAMURA, T., NANAVATI, R., NANGIA, S., NEWTON, P., NGOUN, C., NOVOTNEY, A., NWAKANMA, D., OBIERO, C. W., OLIVAS-MARTINEZ, A., OLLIARO, P., OOKO, E., et al. 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet,* 399**,** 629-655.

NAGAI, H., CAMBRONNE, E. D., KAGAN, J. C., AMOR, J. C., KAHN, R. A. & ROY, C. R. 2005. A C-terminal translocation signal required for Dot/Icm-dependent delivery of the Legionella RalF protein to host cells. *Proc Natl Acad Sci U S A,* 102**,** 826-31.

NAKANO, N., KUBORI, T., KINOSHITA, M., IMADA, K. & NAGAI, H. 2010. Crystal structure of Legionella DotD: insights into the relationship between type IVB and type II/III secretion systems. *PLoS Pathog,* 6**,** e1001129.

NAS, M. Y., WHITE, R. C., DUMONT, A. L., LOPEZ, A. E. & CIANCIOTTO, N. P. 2019. Stenotrophomonas maltophilia Encodes a VirB/VirD4 Type IV Secretion System That Modulates Apoptosis in Human Cells and Promotes Competition against Heterologous Bacteria, Including Pseudomonas aeruginosa. *Infect Immun,* 87.

NOGALES, E. & SCHERES, S. H. 2015. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Mol Cell,* 58**,** 677-89.

OKA, G. U., SOUZA, D. P., CENENS, W., MATSUYAMA, B. Y., CARDOSO, M. V. C., OLIVEIRA, L. C., DA SILVA LIMA, F., CUCCOVIA, I. M., GUZZO, C. R., SALINAS, R. K. & FARAH, C. S. 2022. Structural basis for effector recognition by an antibacterial type IV secretion system. *Proc Natl Acad Sci U S A,* 119.

ORLOVA, E. V. & SAIBIL, H. R. 2011. Structural analysis of macromolecular assemblies by electron microscopy. *Chem Rev,* 111**,** 7710-48.

PANSEGRAU, W., MIELE, L., LURZ, R. & LANKA, E. 1987. Nucleotide sequence of the kanamycin resistance determinant of plasmid RP4: homology to other aminoglycoside 3'-phosphotransferases. *Plasmid,* 18**,** 193-204.

PANSEGRAU, W., SCHOUMACHER, F., HOHN, B. & LANKA, E. 1993. Site-specific cleavage and joining of single-stranded DNA by VirD2 protein of Agrobacterium tumefaciens Ti plasmids: analogy to bacterial conjugation. *Proc Natl Acad Sci U S A,* 90**,** 11538-42.

PARK, D., CHETRIT, D., HU, B., ROY, C. R. & LIU, J. 2020. Analysis of Dot/Icm Type IVB Secretion System Subassemblies by Cryoelectron Tomography Reveals Conformational Changes Induced by DotB Binding. *mBio,* 11.

PATKOWSKI, J. B., DAHLBERG, T., AMIN, H., GAHLOT, D. K., VIJAYRAJRATNAM, S., VOGEL, J. P., FRANCIS, M. S., BAKER, J. L., ANDERSSON, M. & COSTA, T. R. D. 2023. The F-pilus biomechanical adaptability accelerates conjugative dissemination of antimicrobial resistance and biofilm formation. *Nat Commun,* 14**,** 1879.

PENCZEK, P. A. 2010. Fundamentals of three-dimensional reconstruction from projections. *Methods Enzymol,* 482**,** 1-33.

PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem,* 25**,** 1605-12.

PREVOST, M. S. & WAKSMAN, G. 2018. X-ray crystal structures of the type IVb secretion system DotB ATPases. *Protein Sci,* 27**,** 1464-1475.

PUNJANI, A., RUBINSTEIN, J. L., FLEET, D. J. & BRUBAKER, M. A. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods,* 14**,** 290-296.

PURTSCHERT-MONTENEGRO, G., CARCAMO-OYARCE, G., PINTO-CARBO, M., AGNOLI, K., BAILLY, A. & EBERL, L. 2022. Pseudomonas putida mediates bacterial killing, biofilm invasion and biocontrol with a type IVB secretion system. *Nat Microbiol,* 7**,** 1547-1557.

RASHKOVA, S., SPUDICH, G. M. & CHRISTIE, P. J. 1997. Characterization of membrane and protein interaction determinants of the Agrobacterium tumefaciens VirB11 ATPase. *J Bacteriol,* 179**,** 583-91.

REDZEJ, A., UKLEJA, M., CONNERY, S., TROKTER, M., FELISBERTO-RODRIGUES, C., CRYAR, A., THALASSINOS, K., HAYWARD, R. D., ORLOVA, E. V. & WAKSMAN, G. 2017. Structure of a VirD4 coupling protein bound to a VirB type IV secretion machinery. *EMBO J,* 36**,** 3080-3095.

RIPOLL-ROZADA, J., ZUNZUNEGUI, S., DE LA CRUZ, F., ARECHAGA, I. & CABEZON, E. 2013. Functional interactions of VirB11 traffic ATPases with VirB4 and VirD4 molecular motors in type IV secretion systems. *J Bacteriol,* 195**,** 4195-201.

RIVERA-CALZADA, A., FRONZES, R., SAVVA, C. G., CHANDRAN, V., LIAN, P. W., LAEREMANS, T., PARDON, E., STEYAERT, J., REMAUT, H., WAKSMAN, G. & ORLOVA, E. V. 2013. Structure of a bacterial type IV secretion core complex at subnanometre resolution. *EMBO J,* 32**,** 1195-204.

ROHOU, A. & GRIGORIEFF, N. 2015. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol,* 192**,** 216-21.

ROSENTHAL, P. B. & HENDERSON, R. 2003. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol,* 333**,** 721-45.

SAIBIL, H. R. 2000. Macromolecular structure determination by cryo-electron microscopy. *Acta Crystallogr D Biol Crystallogr,* 56**,** 1215-22.

SCHEIFFELE, P., PANSEGRAU, W. & LANKA, E. 1995. Initiation of Agrobacterium tumefaciens T-DNA processing. Purified proteins VirD1 and VirD2 catalyze site- and strand-specific cleavage of superhelical T-border DNA in vitro. *J Biol Chem,* 270**,** 1269-76.

SCHRODINGER, LLC 2015. The PyMOL Molecular Graphics System, Version 1.8.

SCHULZ, S., WILKES, M., MILLS, D. J., KUHLBRANDT, W. & MEIER, T. 2017. Molecular architecture of the N-type ATPase rotor ring from Burkholderia pseudomallei. *EMBO Rep,* 18**,** 526-535.

SEGAL, G. & SHUMAN, H. A. 1999. Possible origin of the Legionella pneumophila virulence genes and their relation to Coxiella burnetii. *Mol Microbiol,* 33**,** 669-70.

SEXTON, J. A. & VOGEL, J. P. 2002. Type IVB secretion by intracellular pathogens. *Traffic,* 3**,** 178-85.

SEXTON, J. A., YEO, H. J. & VOGEL, J. P. 2005. Genetic analysis of the Legionella pneumophila DotB ATPase reveals a role in type IV secretion system protein export. *Mol Microbiol,* 57**,** 70-84.

SGRO, G. G., COSTA, T. R. D., CENENS, W., SOUZA, D. P., CASSAGO, A., COUTINHO DE OLIVEIRA, L., SALINAS, R. K., PORTUGAL, R. V., FARAH, C. S. & WAKSMAN, G. 2018. Cryo-EM structure of the bacteria-killing type IV secretion system core complex from Xanthomonas citri. *Nat Microbiol,* 3**,** 1429-1440.

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. 2019. Bacteria-Killing Type IV Secretion Systems. *Front Microbiol,* 10**,** 1078.

SHALA-LAWRENCE, A., BRAGAGNOLO, N., NOWROOZI-DAYENI, R., KHEYSON, S. & AUDETTE, G. F. 2018. The interaction of TraW and TrbC is required to facilitate conjugation in F-like plasmids. *Biochem Biophys Res Commun,* 503**,** 2386-2392.

SHEEDLO, M. J., CHUNG, J. M., SAWHNEY, N., DURIE, C. L., COVER, T. L., OHI, M. D. & LACY, D. B. 2020. Cryo-EM reveals species-specific components within the Helicobacter pylori Cag type IV secretion system core complex. *Elife,* 9.

SHEEDLO, M. J., DURIE, C. L., CHUNG, J. M., CHANG, L., ROBERTS, J., SWANSON, M., LACY, D. B. & OHI, M. D. 2021. Cryo-EM reveals new species-specific proteins and symmetry elements in the Legionella pneumophila Dot/Icm T4SS. *Elife,* 10.

SHEEDLO, M. J., OHI, M. D., LACY, D. B. & COVER, T. L. 2022. Molecular architecture of bacterial type IV secretion systems. *PLoS Pathog,* 18**,** e1010720.

SHIRASU, K. & KADO, C. I. 1993. Membrane location of the Ti plasmid VirB proteins involved in the biosynthesis of a pilin-like conjugative structure on Agrobacterium tumefaciens. *FEMS Microbiol Lett,* 111**,** 287-94.

SILVERMAN, P. M. 1997. Towards a structural biology of bacterial conjugation. *Mol Microbiol,* 23**,** 423-9.

SKURRAY, R. A., HANCOCK, R. E. & REEVES, P. 1974. Con--mutants: class of mutants in Escherichia coli K-12 lacking a major cell wall protein and defective in conjugation and adsorption of a bacteriophage. *J Bacteriol,* 119**,** 726-35.

SLOTBOOM, D. J., DUURKENS, R. H., OLIEMAN, K. & ERKENS, G. B. 2008. Static light scattering to characterize membrane proteins in detergent solution. *Methods,* 46**,** 73-82.

SMITH, M. A., COINCON, M., PASCHOS, A., JOLICOEUR, B., LAVALLEE, P., SYGUSCH, J. & BARON, C. 2012. Identification of the binding site of Brucella VirB8 interaction inhibitors. *Chem Biol,* 19**,** 1041-8.

SOBTI, M., WALSHE, J. L., WU, D., ISHMUKHAMETOV, R., ZENG, Y. C., ROBINSON, C. V., BERRY, R. M. & STEWART, A. G. 2020. Cryo-EM structures provide insight into how E. coli F1Fo ATP synthase accommodates symmetry mismatch. *Nat Commun,* 11**,** 2615.

SOUZA, D. P., ANDRADE, M. O., ALVAREZ-MARTINEZ, C. E., ARANTES, G. M., FARAH, C. S. & SALINAS, R. K. 2011. A component of the Xanthomonadaceae type IV secretion system combines a VirB7 motif with a N0 domain found in outer membrane transport proteins. *PLoS Pathog,* 7**,** e1002031.

SOUZA, D. P., OKA, G. U., ALVAREZ-MARTINEZ, C. E., BISSON-FILHO, A. W., DUNGER, G., HOBEIKA, L., CAVALCANTE, N. S., ALEGRIA, M. C., BARBOSA, L. R., SALINAS, R. K., GUZZO, C. R. & FARAH, C. S. 2015. Bacterial killing via a type IV secretion system. *Nat Commun,* 6**,** 6453.

SPANKIE, T. J., HAYWOOD, A. L., DOTTORINI, T., BARROW, P. A. & HIRST, J. D. 2020. Interaction of the maturation protein of the bacteriophage MS2 and the sex pilus of the Escherichia coli F plasmid. *J Mol Graph Model,* 101**,** 107723.

STACHEL, S. E. & ZAMBRYSKI, P. C. 1986. virA and virG control the plant-induced activation of the T-DNA transfer process of A. tumefaciens. *Cell,* 46**,** 325-33.

STANLEY, S. A., RAGHAVAN, S., HWANG, W. W. & COX, J. S. 2003. Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system. *Proc Natl Acad Sci U S A,* 100**,** 13001-6.

SUNDBERG, C., MEEK, L., CARROLL, K., DAS, A. & REAM, W. 1996. VirE1 protein mediates export of the single-stranded DNA-binding protein VirE2 from Agrobacterium tumefaciens into plant cells. *J Bacteriol,* 178**,** 1207-12.

SUTHERLAND, M. C., NGUYEN, T. L., TSENG, V. & VOGEL, J. P. 2012. The Legionella IcmSW complex directly interacts with DotL to mediate translocation of adaptor-dependent substrates. *PLoS Pathog,* 8**,** e1002910.

TAKEKAWA, N., KAWAMOTO, A., SAKUMA, M., KATO, T., KOJIMA, S., KINOSHITA, M., MINAMINO, T., NAMBA, K., HOMMA, M. & IMADA, K. 2021. Two Distinct Conformations in 34 FliF Subunits Generate Three Different Symmetries within the Flagellar MS-Ring. *mBio,* 12.

TERRADOT, L., BAYLISS, R., OOMEN, C., LEONARD, G. A., BARON, C. & WAKSMAN, G. 2005. Structures of two core subunits of the bacterial type IV secretion system, VirB8 from Brucella suis and ComB10 from Helicobacter pylori. *Proc Natl Acad Sci U S A,* 102**,** 4596-601.

TEUFEL, F., ALMAGRO ARMENTEROS, J. J., JOHANSEN, A. R., GISLASON, M. H., PIHL, S. I., TSIRIGOS, K. D., WINTHER, O., BRUNAK, S., VON HEIJNE, G. & NIELSEN, H. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol,* 40**,** 1023-1025.

TIMMINS, P. A., LEONHARD, M., WELTZIEN, H. U., WACKER, T. & WELTE, W. 1988. A physical characterization of some detergents of potential use for membrane protein crystallization. *FEBS Letters,* 238**,** 361-368.

TINLAND, B., SCHOUMACHER, F., GLOECKLER, V., BRAVO-ANGEL, A. M. & HOHN, B. 1995. The Agrobacterium tumefaciens virulence D2 protein is responsible for precise integration of T-DNA into the plant genome. *EMBO J,* 14**,** 3585-95.

VAN HEEL, M. 1987. Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy,* 21**,** 111-23.

VARGA, M. G., SHAFFER, C. L., SIERRA, J. C., SUAREZ, G., PIAZUELO, M. B., WHITAKER, M. E., ROMERO-GALLO, J., KRISHNA, U. S., DELGADO, A., GOMEZ, M. A., GOOD, J. A., ALMQVIST, F., SKAAR, E. P.,

CORREA, P., WILSON, K. T., HADJIFRANGISKOU, M. & PEEK, R. M. 2016. Pathogenic Helicobacter pylori strains translocate DNA and activate TLR9 via the cancer-associated cag type IV secretion system. *Oncogene,* 35**,** 6262-6269.

VINCENT, C. D., FRIEDMAN, J. R., JEONG, K. C., BUFORD, E. C., MILLER, J. L. & VOGEL, J. P. 2006. Identification of the core transmembrane complex of the Legionella Dot/Icm type IV secretion system. *Mol Microbiol,* 62**,** 1278-91.

VINCENT, C. D., FRIEDMAN, J. R., JEONG, K. C., SUTHERLAND, M. C. & VOGEL, J. P. 2012. Identification of the DotL coupling protein subcomplex of the Legionella Dot/Icm type IV secretion system. *Mol Microbiol,* 85**,** 378-91.

VIROLLE, C., GOLDLUST, K., DJERMOUN, S., BIGOT, S. & LESTERLIN, C. 2020. Plasmid Transfer by Conjugation in Gram-Negative Bacteria: From the Cellular to the Community Level. *Genes (Basel),* 11.

VOGEL, J. P., ANDREWS, H. L., WONG, S. K. & ISBERG, R. R. 1998. Conjugative transfer by the virulence system of Legionella pneumophila. *Science,* 279**,** 873-6.

VOTH, D. E., BROEDERDORF, L. J. & GRAHAM, J. G. 2012. Bacterial Type IV secretion systems: versatile virulence machines. *Future Microbiol,* 7**,** 241-57.

WAGNER, A., WHITAKER, R. J., KRAUSE, D. J., HEILERS, J. H., VAN WOLFEREN, M., VAN DER DOES, C. & ALBERS, S. V. 2017. Mechanisms of gene flow in archaea. *Nat Rev Microbiol,* 15**,** 492-501.

WAKSMAN, G. 2019. From conjugation to T4S systems in Gram-negative bacteria: a mechanistic biology perspective. *EMBO Rep,* 20.

WHITE, C. E. & WINANS, S. C. 2007. Cell-cell communication in the plant pathogen Agrobacterium tumefaciens. *Philos Trans R Soc Lond B Biol Sci,* 362**,** 1135-48.

WONG, J. J., LU, J., EDWARDS, R. A., FROST, L. S. & GLOVER, J. N. 2011. Structural basis of cooperative DNA recognition by the plasmid conjugation factor, TraM. *Nucleic Acids Res,* 39**,** 6775-88.

WORRALL, L. J., HONG, C., VUCKOVIC, M., DENG, W., BERGERON, J. R. C., MAJEWSKI, D. D., HUANG, R. K., SPRETER, T., FINLAY, B. B., YU, Z. & STRYNADKA, N. C. J. 2016. Near-atomic-resolution cryo-EM analysis of the Salmonella T3S injectisome basal body. *Nature,* 540**,** 597-601.

XU, J., XU, D., WAN, M., YIN, L., WANG, X., WU, L., LIU, Y., LIU, X., ZHOU, Y. & ZHU, Y. 2017. Structural insights into the roles of the IcmS-IcmW complex in the type IVb secretion system of Legionella pneumophila. *Proc Natl Acad Sci U S A,* 114**,** 13543-13548.

YUAN, Q., CARLE, A., GAO, C., SIVANESAN, D., ALY, K. A., HOPPNER, C., KRALL, L., DOMKE, N. & BARON, C. 2005. Identification of the VirB4-VirB8-VirB5-VirB2 pilus assembly sequence of type IV secretion systems. *J Biol Chem,* 280**,** 26349-59.

ZHENG, S. Q., PALOVCAK, E., ARMACHE, J. P., VERBA, K. A., CHENG, Y. & AGARD, D. A. 2017. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods,* 14**,** 331-332.

ZHENG, W., PENA, A., LOW, W. W., WONG, J. L. C., FRANKEL, G. & EGELMAN, E. H. 2020. Cryoelectron-Microscopic Structure of the pKpQIL Conjugative Pili from Carbapenem-Resistant Klebsiella pneumoniae. *Structure,* 28**,** 1321-1328 e2.

ZHU, J., OGER, P. M., SCHRAMMEIJER, B., HOOYKAAS, P. J., FARRAND, S. K. & WINANS, S. C. 2000. The bases of crown gall tumorigenesis. *J Bacteriol,* 182**,** 3885-95.

ZIVANOV, J., NAKANE, T., FORSBERG, B. O., KIMANIUS, D., HAGEN, W. J., LINDAHL, E. & SCHERES, S. H. 2018. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife,* 7.

ZUPAN, J., HACKWORTH, C. A., AGUILAR, J., WARD, D. & ZAMBRYSKI, P. 2007. VirB1* promotes T-pilus formation in the vir-Type IV secretion system of Agrobacterium tumefaciens. *J Bacteriol,* 189**,** 6551-63.