THE ROLE OF AN ACCESSORY PROTEIN FAMILY IN TYPE VI SECRETION

THE ROLE OF AN ACCESSORY PROTEIN FAMILY IN TYPE VI SECRETION

BY KARTIK SACHAR

A thesis submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Master of Science

DESCRIPTIVE NOTE

McMaster University MASTER OF SCIENCE (2022) Hamilton, Ontario (Biochemistry and Biomedical Sciences)

Title: The role of an accessory protein family in type VI secretion

Author: Kartik Sachar (McMaster University)

Supervisor: Dr. John Whitney

Number of pages: xii, 75

LAY ABSTRACT

Microbial communities compete with their neighboring cells for resources and space. Species of Gram-negative bacteria can use a pathway called the type VI secretion system or T6SS to kill these competitor bacterial species. T6SS is a protein complex that facilitates the injection of toxic proteins, from attacking cells directly into target cells. The secretion of these toxic proteins requires additional accessory proteins, which recruit and stabilize the toxic proteins before secretion. In this work, I characterize an accessory protein family that functions as a chaperone for these toxic proteins. First, I showed that these accessory protein family genes are linked to genes that encode for the core T6SS apparatus proteins. Secondly, I was able to capture the first snapshot of this accessory protein family. This snapshot gives molecular details into the role of this accessory protein in T6SS secretion. Collectively, these results expand our knowledge of how bacteria use the T6SS to secrete toxic proteins. Uncovering a new mechanism for directly delivering proteins into different cell types has valuable potential for medical and industrial applications.

ABSTRACT

The type VI secretion system (T6SS) is a widely distributed protein nano-machinery found in Gram-negative bacteria. T6SSs are a harpoon-like apparatus that delivers toxic effector proteins directly into target cells in a contact-dependent manner. These effectors are associated with a diverse range of functions such as cytoskeletal modification, biofilm formation, and bacterial competition. Effectors are loaded onto the T6SS by interacting with its structural components. Often, additional accessory proteins are needed for the secretion of these effector proteins. Despite their importance, the direct role of these accessory proteins in the assembly of the T6SS warhead or its subsequent 'firing' is unclear. One such example is the putative DUF2169 family of proteins. These proteins are required for effector secretion, however, the molecular function of DUF2169 proteins remains unknown.

I have shown that DUF2169-encoding genes co-occur with the structural components of T6SS, such as VgrG and effectors that possess a specialized N-terminal PAAR-like domain. Using a *Pseudomonas aeruginosa* T6SS-associated gene cluster, I show the network of protein interactions with DUF2169 and these T6SS structural components. Furthermore, I was able to show that these gene clusters form a conserved synteny. Multiple proteins encoded within DUF2169 syntenies are predicted to mimic previously characterized proteins associated with T6SS function. Lastly, using X-ray crystallography I was able to solve the structure of DUF2169 from *Vibrio xiamenensis* as the first representative of the DUF2169 protein family. Additionally, using structural predictions I show the molecular interaction between DUF2169 and PAAR-like domains of effectors as a potential chaperone. Gaining structural insight into the role of this protein will not only enhance our understanding of T6SS function but also highlight the mechanistic differences between different T6SS warheads. Since T6SSs can deliver a variety of

proteins into both prokaryotic and eukaryotic cells, understanding how this delivery mechanism works expands our knowledge of how bacteria interact with diverse cell types. Furthermore, uncovering a new mechanism for directly delivering proteins into different cell types has valuable potential for medical and industrial applications.

ACKNOWLEDGMENTS

The completion of this work could not have been achieved without the support and encouragement of my friends and family. Firstly, I would like to thank my supervisor, Dr. John Whitney, for giving me the opportunity to work in his lab. Thank you for your guidance and support in this academic journey. I greatly appreciate your help with my scientific writing. This work would be incomprehensible without your help.

Secondly, to all my fellow Whitney lab members (both past and present), it was truly a pleasure to work alongside such brilliant minds. The senior students (Dr. Timothy Klein, Dr. Shehryar Ahmad, Nathan Bullen), I greatly appreciate the mentoring and support you have provided for me over the years. Dirk Grebenc and Andrea Alexei, thank you for the conversations we shared over coffee (even though it was mostly me venting about my experiments). A very special thank you to Amirahmad Azhieh for being an amazing friend. I will remember our days-long chess games to staying late in the lab, his continued friendship has been immensely appreciated.

To IDR support staff, Dr. Catie Luck, who helped me use all the X-ray equipment and set up protein crystal trays. Dr. Youngchang Kim for collecting the X-ray dataset and building the structure.

Furthermore, my family and friends (Dominic, Emily, Liam, and Alex) back in Winnipeg. Past mentors, Dr. Gerd Prehna and (soon to be Dr.) Neil Lorente Cobo, for encouraging me to further pursue science. It has been hard to only see each other a few times over the past years, but I've greatly appreciated that you guys have been there for me, nonetheless. Lastly, to my partner Shaylyn. It has truly been a pleasure to have you in my life and I could not have accomplished this without you.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AF2: AlphaFold2

CTD: C-terminal toxin domain

CFU: Colony Forming Units

CV: Column Volume

DTT: Dithiothreitol

DUF: Domain of Unknown Function

EAG: Effector Associated Genes

FATCAT: Flexible structure AlignmenT by Chaining Aligned fragment pairs allowing Twists

GNN: Genome Neighborhood Network

IPTG: Isopropyl β-D-1-Thiogalactopyranoside

LB: Luria Broth

MW: Molecular Weight

OB: Oligonucleotide/Oligosaccharide binding

PAAR: Pro-Ala-Ala-Arg

PAC: PAAR-PIPY Associated Chaperones

PCR: Polymerase Chain Reaction

Pld: Phospholipase D

PRP: Pentapeptide-Repeats Protein

SAXS: Small Angle X-ray Scattering

SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SEC: Size-Exclusion Chromatography

SOE: Splicing by Overlap Extension

SSN: Sequence Similarity Network

T6SS: Type 6 Secretion System

TAC: T6-Associated Gene Custer

TAE: T6 Amidase Effector

TAP: T6SS Adaptor Protein

TEC: T6 Effector Chaperone

TDE: T6 DNase Effectors

TGE: T6 Glycoside Hydrolase Effector

TLE: T6 Lipases Effector

TMD: Transmembrane domain

TME: T6 Membrane Disrupting Effector

TBS: Tris-Buffered Saline

TTR: Transthyretin-like

VgrG: Valine-glycine Repeat Protein G

VSV-G: Vesicular Stomatis Virus G

DECLARATION OF ACADEMIC ACHIEVEMENT

I, Kartik Sachar, declare that all work presented in this thesis was performed by myself. Any work performed by others contributing to this thesis has been truly appreciated and referenced appropriately. The X-ray data collection was performed by Dr. Youngchang Kim at the Advanced Photon Source Argonne National Laboratory.

Chapter 1: Introduction

1.1 Bacterial Antagonism

Microbial communities form complex interactive and competitive networks with their surrounding species. In these diverse communities, space and resource limitations encourage selective pressure to evolve antagonistic responses against other competing bacteria. Antagonistic behaviour of bacteria that includes interspecies competition is mediated through the production of specialized antimicrobials, such as secondary metabolites and toxic proteins (e.g. lysozymes, DNases, hydrolases) (Hibbing et al., 2010). This provides a competitive advantage for the antimicrobial-secreting bacterium, as secretion of these molecules into its surroundings can inhibit the growth of opposing species.

In comparison to the secretion of small molecules, the secretion of proteins is a complex and tightly regulated process. As such, bacteria have evolved large multi-protein nanomachinery for exporting proteins from the cytosol to the extracellular milieu. One mechanism by which Gram-negative bacteria secrete proteins is via a pathway called the Type 6 Secretion System (T6SS). The T6SS is a contractile protein nanomachine that allows Gram-negative bacteria to deliver toxic proteins, termed effectors, directly into neighbouring cells in a contact-dependent manner. "T6SS duelling" refers to the antagonistic behaviour of two neighbouring cells to deliver effector proteins directly into each other to inhibit their growth (Basler et al., 2013). Subsequent studies have shown that T6SS-mediated killing can be a response mechanism against antagonistic bacteria. As a result, two types of bacterial behaviour are observed: hostile bacteria (constitutively active T6SS) and defensive bacteria (controlled T6SS). As an example, *Pseudomonas aeruginosa* can sense T6SS attacks from *Vibrio cholera*e or *Acinetobacter baylyi,*

and activates its own T6SS to counterattack (Basler et al., 2013). This "tit for tat" model suggests that some bacteria are not constantly hostile, but rather use T6SS to ward off other hostile bacteria. With this in mind, it is of great importance to study the underlying mechanisms of T6SS.

1.2 Type 6 Secretion System Protein Nanomachinery

It is estimated that approximately >25% of Gram-negative bacteria possess at least one T6SS gene cluster (Bingle et al., 2008). While these bacteria share the same 14 core components, considerable variation can be observed between the function, regulation, and effector repertoire of T6SSs between organisms. The current model of T6SS apparatus structure and function sub-divides this nanomachine into three core components: a membrane anchoring complex, a phage tail-like structure, and a baseplate complex (Figure 1) (Zoued et al., 2014). The phage tail-like sheath tube is structurally homologous to the tail structure of T4 bacteriophage and consists of five proteins: Hemolysin Coregulated Protein (Hcp), Valine-glycine Repeat protein G (VgrG), TssB, TssC, and TssE (Silverman et al., 2013). TssB and TssC form an outer sheath which encloses the homo-hexameric interior Hcp tube. VgrGs are referred to as spike proteins because they form the tip of the apparatus and are the site for loading effector proteins. An additional protein PAAR (Pro-Ala-Ala-aRg) protein may also be required for the loading of the effector. PAAR proteins play an additional role of sharpening the VgrG spike, allowing the apparatus to penetrate the membrane of the target cell (Shneider et al., 2013). The contraction of TssB/TssC sheath proteins propels the Hcp rings and VgrG/PAAR to the exterior of the attacking cell and directly into the neighbouring cell (Basler & Mekalanos, 2012). Additionally, an AAA+ ATPase, ClpV, is required for recycling the contracted TssC/TssB sheath to restore the extended form of the T6SS for another round of secretion (Douzi et al., 2016). The membrane-anchoring complex

consists of three separate subunits that are encoded by *tssJ*, *tssM*, and *tssL*. TssJ is a lipoprotein that localizes to the periplasm and is anchored to the outer membrane. TssM interacts with TssJ via its C-terminal domain and is anchored to the inner membrane by three transmembrane domains. TssL is located in the cytoplasm where it interacts with the baseplate components of T6SS (Cianfanelli et al., 2016). Additionally, TssL extends into the periplasm and is anchored to the inner membrane via a single transmembrane domain, where it interacts with the baseplate of the T6SS (Cianfanelli et al., 2016). The last sub-complex, the baseplate, is made up of four proteins TssE, TssF, TssG, and TssK (Silverman et al., 2013). This structure serves as a nucleation point for the polymerization of the tail-like structure, which is a prerequisite step for the subsequent secretion of effector proteins.

1.3 T6SS Effector Proteins

The first repertoire of characterized T6SS effectors focused on the role of T6SSs in establishing pathogenesis and mediating host interaction (Schwarz et al., 2010). These effector proteins were associated with a diverse range of biological functions such as host cytoskeletal modification (e.g., VgrG-1^{VC} can cross-link cytosolic actin in macrophages), escaping immune cells (e.g., $KatN^{EHEC}$ is a Mn^{2+} catalase which can reduce levels of intracellular oxidative species following phagocytosis), membrane pore formation (e.g., $VasX^{VC}$ PH domain binds phosphatidylinositol phosphates leading to pore formation in amoebas) and mediating host cell membrane fusion (e.g., VgrG-5^{BT} is translocated in macrophages leading to the formation of multinucleated giant cells through cell membrane fusion) (Miyata et al., 2011; Pukatzki et al., 2007; Toesca et al., 2014; Wan et al., 2017). As the repertoire of known T6SS effectors has expanded over the past decade, it has become evident that T6SSs are primarily used for interbacterial competition. As it currently stands, these antimicrobial effectors are by far the most well-characterized class of

effector proteins. Antimicrobial effectors can be divided into two broad categories: periplasmic- and cytoplasmic-acting effectors. These T6SS-exported toxins have been grouped into families based on their biochemical function such as T6SS peptidoglycan amidase effectors (Tae), glycoside hydrolases (Tge), lipases/esterases (Tle), membrane disruptors/pore formers (Tme) or DNases (Tde). Antimicrobial effectors are often encoded beside a cognate immunity protein, which is typically highly specific for its associated toxin (Hood et al., 2010; Russell et al., 2012). Immunity proteins protect both toxin-secreting cells and nearby sister cells from the toxicity of these T6SS effectors. To date, *P. aeruginosa* has the best-characterized T6SS. As there are hundreds of characterized effectors, select representatives from *P. aeruginosa* will serve as examples to illustrate the biological functions of T6SS-associated effectors.

Overall, *P. aeruginosa* encodes three core T6SS gene clusters, termed the H(1-3)-T6SSs, along with several stand-alone "orphan" T6SS accessory gene clusters. It was first observed in *P. aeruginosa* PAO1 that the H1-T6SS functions to target bacteria via exported effector proteins that hydrolyze peptidoglycan in the target cell (Tse1 & Tse3) (Russell et al., 2011). As alluded to previously, these peptidoglycan hydrolases fall into two sub-categories, Tae and Tge, which cleave the peptide stems and glycan chains of peptidoglycan, respectively. Shortly after their identification, a crystal structure showed that Tae1 adopts a papain-like cysteine protease fold that facilitates the cleavage of peptidoglycan cross-links in the same way as previously characterized peptidoglycan D,L-endopeptidases (Benz et al., 2012). Meanwhile, Tge effectors such as Tge2 hydrolyze the N-acetylmuramic acid (NAM)-N-acetylglucosamine (NAG) backbone. Tge2 adopts a lysozyme-like fold and is neutralized by an immunity protein in a manner similar to that of other lysozyme inhibitors (Whitney et al., 2013). To prevent hydrolysis of sister cell peptidoglycan cross-links, *P. aeruginosa* also encodes an immunity protein, Tai1,

that inactivates Tae1 by blocking its active site. Together, Tae1 and Tge2 were the first T6SS effectors characterized which uncovered the role of T6SS in antibacterial competition.

While the H1-T6SS has exclusively been shown to function as a bacterium targeting system, several anti-eukaryotic effectors have been identified and shown to be exported by H2-T6SS of *P. aeruginosa*. For example, two H2-T6SS associated phospholipase D (Pld) effectors, PldA (Tle5a) and PldB (Tle5b) possess dual HXKXXXXD catalytic motifs (Russell et al., 2013; Wettstadt et al., 2019). PldA is shown to have activity against a major phospholipid present in the bacterial membrane, phosphatidylethanolamine. It is therefore unsurprising that PldA can induce cell death by degrading phosphatidylethanolamine (Russell et al., 2013). Interestingly, both PldA and PldB are involved in promoting human cell evasion by activation of the phosphatidylinositol 3-kinase (PI3K)/Akt signalling pathway (Jiang et al., 2014). Therefore, PldA and PldB are considered to be "trans-kingdom" effectors, since they can target both eukaryotic and prokaryotic cells. It was later proposed that these Pld effectors may act synergistically with VgrG2b (Sana et al., 2016). H2-T6SS also injects VgrG2b which targets the microtubule network in macrophages, mainly the tubulin ring complex components (γ-TuRC) promoting microtubule nucleation at the membrane. Following VgrG2, the translocation of PldA and PldB and recruitment of Akt allows for internalization of *P. aeruginosa* into epithelial cells. These examples highlight the broad range of functions associated with T6SS effectors.

The most extensive family of characterized cytosolic T6SS effector are DNA-targeting effectors. Nuclease effectors are also a common form of antagonism found in other bacterial competition pathways, as they are an effective means of eliminating competitors. T6SS DNase effectors (Tde) are widely distributed and contain various DNase domains such as HNH, PoNe, and Tox-REase (Bernal et al., 2017; Bondage et al., 2016; H. Zhang et al., 2013). T6SS DNase effectors

most often belong to the HNH superfamily, which contains effectors with a (G/A/P)HH catalytic site. For this catalytic site, the first histidine is predicted to be involved in metal ion binding and the second histidine in deprotonation of a catalytic water molecule (Zhang et al. 2012). In this family of effectors, two proteins are relevant for this thesis: Tse7 from *P. aeruginosa* and VP1415 from *V. parahaemolyticus* (Pissaridou et al., 2018; Salomon et al., 2014). In both proteins, the C-terminal domain has a GHH nuclease domain that lies downstream of an N-terminal Domains of Unknown Function (DUFs) DUF4150 domain, which I hypothesize targets these effectors to the T6SS apparatus. The primary focus of my MSc project, as discussed in further detail below, is on the discovery and characterization of a novel accessory protein family required for the delivery of these effectors into competitor bacteria.

1.4 Structural Aspects of the T6SS Phage Tail-like Assembly

The mechanism of loading effectors onto the T6SS secretion apparatus remains mysterious. In all cases, T6SS effectors are genetically linked with one of the phage tail-like proteins Hcp, PAAR or VgrG proteins which form the tip of the T6SS apparatus. For this reason, effector proteins are often encoded in the proximity of these *hcp, vgrG*, and *paar* genes. VgrG proteins are structural homologs of T4 bacteriophage spike proteins gp5 and gp27 (Figure 2) (Uchida et al., 2014). These are trimeric proteins that consist of two main regions, a head and spike region. The head region is structurally homologous to gp27 and interacts at the top of stacked Hcp rings. The region between the head and the spike of VgrG contains an oligonucleotide/oligosaccharide binding (OB) fold motif. The OB-fold motif is a β-barrel made up of five β-strands with each one of the three β-barrels in the trimeric protein stacking laterally to form the central domain of VgrG (Spínola-Amilibia et al., 2016). The spike region is structurally homologous to the gp5 and consists of β-strands contributed by each monomer that intertwine with one another to form a

triple-stranded β-helix in a VgrG trimer. Some effectors also require an additional protein, termed PAAR, which forms a conical tip that sits on top of the flat surface formed by the end of the VgrG β-helix (Shneider et al., 2013). In PAAR, the conserved Pro-Ala-Ala-Arg motif stabilizes the central core of the conical structure where the three parts of the triangle intersect (Shneider et al., 2013). PAAR proteins have three histidines and one cysteine, all conserved residues, which form a Zn^{2+} binding site at the tip of the cone's vertex that stabilizes the overall fold. A comprehensive review published by Zhang et al. showed that the PAAR protein superfamily can be divided into 16 sub-types and doing so revealed an evolutionary relationship to the DUF4280 and DUF4150 (Zhang et al. 2021). These sub-type domains are predicted to resemble the same overall conical structure as PAAR but have different sequences and structural features making them unique. Phylogenetic analysis of *Francisella* species showed that DUF4280 proteins (also called IglG proteins) form a distinct clade from PAAR proteins (Lays et al., 2016). DUF4280 have four conserved cysteine residues near the vertex of the conical structure for the coordination of metals such as zinc and iron. DUF4150 (also called PAAR-like) proteins, however, have no conserved metal binding sites. Despite these structural differences, the interactions between PAAR, DUF4280, and PAAR-like with VgrG are all thought to be mediated by a flat hydrophobic surface formed by the VgrG β-helix and the bottom of the cone-shaped tip protein. PAAR-associated effectors are either directly fused to PAAR domains or encoded by genes surrounding an 'orphan' PAAR. Although not all effectors require PAAR proteins for their association with other T6SS protein components, it has been hypothesized that at least one PAAR protein is required for overall T6SS activity (Cianfanelli et al., 2016). For example, the deletion of all PAAR genes in *Acinetobacter baylyi* and *V. cholerae* prevents the secretion of Hcp into the extracellular milieu (Shneider et al., 2013). It is unclear whether PAAR

proteins regulate the recruitment of VgrG into the base of the T6SS apparatus and/or assist with the penetration of the T6SS across the outer membrane, but nonetheless, PAAR proteins are required for protein secretion via T6SS.

1.5 Mechanism of Effector Recruitment

The delivery mechanism of effector proteins into competitor cells is specific for each effector, however, there are a few general ways in which they are delivered. Some effectors, termed specialized effectors, are fused to a structural component. Alternatively, other effectors, termed cargo effectors, are held in place by physical interactions with a core component (i.e., PAAR, Hcp, or VgrG). These cargo effectors interact with different components of the T6SS, so when informatically analyzed as a group, no conserved signalling domain to target these cargo effectors to the T6 apparatus is apparent. In some cases, marker domains have been associated with cargo effectors, MIX (Marker for Type siX), and FIX (Found in type sIX) (Jana et al., 2019; Salomon et al., 2014). Another domain, RIX (aRginine-rich type sIX) has also been proposed as a similar marker domain similar to MIX and FIX (Kanarek et al., 2022). These marker domains are found in the N-terminus of effectors with a variable C-terminal toxin domain. Moreover, these marker domains are mutually exclusive as MIX and FIX domains are never found within the same effector. These marker domains are not exclusively found on cargo effectors, as certain PAAR-containing effectors also contain a MIX or FIX domain. It is unclear what the role of these marker domains are. The leading hypothesis is that these are modular units that serve to diversify the effector pool (Salomon, 2016). The MIX cassettes are believed to be horizontally transferred via a conjugative plasmid or DNA uptake from dead prey cells (Borgeaud et al., 2015).

Thus far, these marker domains are sought after to help with the identification of new T6SS effectors, rather than understanding their role in effector recruitment. There have been two identified adaptor domains that may function to physically link the above effector motifs to the T6SS spike: Transthyretin-like domains (TTR) and DUF2345. TTR domains are small globular structures composed of two stacked β-sheets that are found as C-terminal extensions of PAAR or VgrG (Flaugnatti et al., 2016, 2020). DUF2345 are VgrG C-terminal extensions and hence extend the length of the VgrG (Figure 2). DUF2345 is very similar to the β-helix of gp5, however, DUF2345 contains a 3-strand anti-parallel β-sheet which folds into a β-prism. The Tle1 effector of enteroaggregative *Escherichia coli* (EAEC) was the first example shown to require no additional adaptor protein or PAAR for the association of the effector with $VgrG^{EAEC}$ (Flaugnatti et al., 2020). The flexible N-terminus of Tle1 forms a β-barrel that interacts with the TTR domain of $VgrG^{EAEC}$. $VgrG^{EAEC}$ also contains a DUF2345 domain between TTR and gp5 β-helix that makes additional contact with Tle1. This interaction mechanism allows for cargo effector recruitment to the T6SS but also stabilizes and neutralizes the toxic effector prior to secretion. Each monomer of $VgrG^{EAEC}$ recruits one Tle1 effector and therefore the full trimeric $VgrG^{EAEC}$ recruits three Tle1 effectors for one secretion event. This discovery was an important milestone highlighting that multiple effectors can be delivered by the T6SS at the same time. All the previously highlighted domains are important for the association of one T6SS effector to the T6SS apparatus.

In comparison to cargo effectors, the loading of specialized effectors is far more straightforward. These specialized effectors are found as a fusion to core secreted components (Hcp/VgrG/PAAR), almost always as a C-terminal effector domain. These specialized core components are secreted as one would expect based on the inverted phage tail analogy and bring

with them their C-terminal effector domain. As an example, the *P. aeruginosa* VgrG2b spike protein contains similar domain architecture as $VgrG^{EAEC}$, except that the C-terminus of VgrG2b extends past the C-terminal TTR domain of $VgrG^{EAEC}$ and contains a metallopeptidase toxin domain that causes lysis in targeted bacteria (Wood et al., 2019). Alongside specialized VgrG effectors, PAAR C-terminal toxin extensions are also widespread. Such examples further emphasize the modular nature of T6SS effectors. The various effector functions and different modes of association with core components allow the T6SS to regulate which specific effectors are being secreted, in response to different targets and stimuli. Likewise, it was shown that H1-T6SS is able to deliver a variety of effectors to a target cell by having different variations of the tip complex, through the association of different PAAR effectors with VgrG (Hachani et al., 2014). The different domains which are associated with effector delivery suggest that multiple effectors are simultaneously delivered into a target cell. Altogether, the assembly and delivery of the T6SS effector is a complicated process that utilizes specific domain association between the effector and the core components to secrete specific effector into target cells. These interactions are summarized in a schematic (Figure 3A).

1.6 T6SS Chaperone and Adaptor Proteins

The above process of effector recruitment is made even more complicated because some T6SS effectors require additional chaperone proteins. These chaperone proteins are hypothesized to recruit and stabilize the effector prior to secretion though most have not been characterized experimentally. In the few characterized instances, these chaperone proteins are required for the secretion of their associated effector but they do not impact the overall T6SS activity (Liang et al., 2015; Unterweger et al., 2015). The concept of T6SS chaperones was first brought to light when it was identified that hexameric Hcp rings may function as a molecular chaperones

(Silverman et al., 2013). These stacked rings form a hollow tube with an internal diameter of ~40Å within which smaller (<20 kDa) cargo effectors can be recruited and exported. For example, *P. aeruginosa* effectors Tse1-3 bind and are stabilized by the inner lumen of the Hcp tube, revealing Hcp as more than just a passive channel for effector loading (Silverman et al., 2013). In addition to Hcp, there are four additional known or speculated chaperone proteins: DUF1795, DUF4123, DUF2875 and DUF2169. While proteins belonging to these domain families play a critical role in effector secretion, in contrast to the exported phage tail-like components (Hcp/VgrG/PAAR), these accessory proteins are themselves not secreted (Cianfanelli et al. 2016; Unterweger et al. 2015; Bondage et al. 2016). In all these DUF families, the most common gene organization places the chaperone encoding gene between *vgrG* and the genes encoding its associated effector-immunity pair (Figure 3B). This conserved gene synteny supports the somewhat limited experimental evidence that these domains may serve as chaperones during effector translocation. However, the similarities between these speculated chaperone domain families are far fewer than their differences, and as such, I hypothesize that each chaperone domain family is associated with a unique molecular function.

One of the first characterized chaperone proteins, Tap-1 (T6SS Adaptor Protein 1), is a DUF4123 domain-containing protein from *V. cholerae* (Unterweger et al., 2015)The authors of this study showed that Tap-1 recruits the effector, TseL, to VgrG-1 for secretion. It was initially hypothesized that the function of DUF4123 family proteins is to recruit effectors to the T6SS apparatus for secretion. However, this has expanded to also include chaperone activity (Pei et al., 2020). Therefore the Tap nomenclature for DUF4123 containing proteins was later replaced by Tec (T6 Effector Chaperone) as it more accurately represents their function (Liang et al., 2015; Pei et al., 2020). These Tec proteins are often encoded alongside effectors containing a MIX

motif within their N-terminus, suggesting that MIX may be specifically involved in Tec-effector interactions (Salomon, 2016). Although the exact nature of the molecular interactions between Tec and its cognate effector and VgrG is unclear, published data show that Tec proteins mediate the interaction between a C-terminal VgrG β-strand and the effector protein (Bondage et al. 2016). More recently, the *P. aeruginosa* DUF4123 containing protein TecT was shown to require a DUF3562 containing co-adaptor, named co-TecT, for the recruitment of the TseT effector to its cognate PAAR domain (Burkinshaw et al., 2018). Conversely, the same study showed that TecT recruits the effector to the C-terminus extension of the PAAR domain. Together, this highlights the ambiguity in DUF4123 domain association with core T6SS components, however, the biological function largely remains consistent as a chaperone protein required for effector secretion.

The DUF1795 family contains proteins such as EagT6 (Effector Associated Gene) *from P. aeruginosa* (Whitney et al. 2015), EagR from *Serratia marcescens* (Alcoforado Diniz et al. 2015) and SciW from *Salmonella* Typhimurium (Ahmad et al., 2020). Previously, our group has shown that Eag proteins are required for the secretion of transmembrane domain (TMD) containing PAAR effectors. Eags stabilize the TMDs of these effectors in the cytoplasm prior to their secretion from the cell (Ahmad et al., 2020; Quentin et al., 2018). In this context, Eags function to shield the hydrophobic effector TMDs from solvent during effector loading into the T6SS and allow for the proper folding of the effector's N-terminal PAAR domain. Although it is unclear why PAAR effectors require these TMDs in the first place, we hypothesized that the TMDs act as a molecular grease that facilitates the translocation of the T6SS spike across target cell membranes thus allowing eventual delivery of the effector into the target cell cytoplasm. In support of this model, all TMD containing PAAR effectors target essential molecules that exist

in the cytoplasm including metabolites such as NAD^+, $NADP^+$, ADP, and ATP, or essential macromolecules like the tubulin homolog, FtsZ, and the bacterial chromosome (Ahmad et al., 2020).

There is currently only a single report describing the role of DUF2875 as a potential chaperone protein (Berni et al., 2019a). The *P. aeruginosa* protein Tla3 contains two tandem DUF2875 domains that are required for the secretion of the effector Tle3 in an H2-T6SS manner. These tandem DUF2875 domains were shown to act as a chaperone for the recruitment of Tle3 to the VgrG2b spike protein by interacting with VgrG2b's TTR domain. Thus, Tla3 appears to be an intermediary that links VgrG2b to Tle3. DUF2875 is a relatively new class of T6SS chaperones and much more remains to be unveiled about this poorly characterized protein family, aspects of which I will touch on within the contents of this thesis.

Similar to the other characterized adaptor and chaperone families, DUF2169 proteins are required for T6SS effector-mediated killing but are not secreted, similar to the other characterized adapter/chaperone families. Bondage et al. also showed that the *Agrobacterium tumefaciens* Atu3641, which is a DUF2169 protein, is required for the delivery of the T6SS exported toxin Tde2 (Bondage et al. 2016). Similar results were obtained by Salomon et al., in which gene disruption of *VP1398* (DUF2169) resulted in attenuated virulence of the *V. parahaemolyticus* DNase toxin VP1415 (Ben-Yaakov & Salomon, 2019). While the underlying mechanism by which DUF2169 facilitates the delivery of the effector is unclear, these studies show that the DUF2169 proteins are required for specific T6SS effector delivery.

While all three of the aforementioned DUFs are considered to be associated with adaptor and/or chaperone function, only proteins from the DUF1795, DUF2875 and DUF4123 superfamilies have biochemical evidence that supports this notion. This has resulted in a lack of understanding

of the diversity of T6SS delivery mechanisms and the precise roles accessory proteins might play in the pre-secretion assembly of effector-spike protein complexes. While substantial work has been done to determine the molecular role of other chaperone families, the molecular function of DUF2169 proteins in T6S remains unknown. **I hypothesize that DUF2169 family proteins are a class of chaperone proteins that stabilize the effector prior to their delivery into target cells. The objective of my MSc thesis will be to elucidate the role of DUF2169 in effector translocation by the T6SS.**

Project Goals

This sets the background information necessary for the objective of my MSc thesis. The aim of my thesis will be to elucidate the answers to the following questions:

1) My first aim will be to identify and characterize protein-protein interaction partners of DUF2169. The *Pseudomonas aeruginosa* T6SS will be used as a model T6SS for this work because it encodes for a single DUF2619 gene, within its T6SS-associated gene cluster (TAC).
2) My second aim will be to structurally characterize DUF2169 proteins. To date, there are no structures of any member of this protein family; structural insights will provide a molecular mechanism by which these protein families execute their function.

Chapter 2: Material and Methods

2.1 Gene Neighborhood Analysis

Genome neighbourhood network (GNN) analysis on the DUF2169 protein family was done using an online EFI-GNT tool to generate Sequence Similarity Networks (SSN) using the DUF2169 protein family (Zallot et al., 2019). The online tool is available at https://efi.igb.illinois.edu/efi-gnt/. SSN groups related proteins together into clusters based on sequence similarity, which can then be used to assign function within a protein family. The histogram of "alignment lengths vs alignment score" was used in the next step of alignment score threshold determination. The SSN was limited to examining only single domain DUF2169 sequence alignments by limiting the sequence length to <400aa and AIScore of 80. The SSN retrieved from EFI-EST server was edited to isolate single clusters using Cytoscape. The final edited SSN was used as a query for EFI-GNT to visualize and calculate neighbouring domains that were associated with multiple clusters.

The second web tool, GeCoViz is available at https://gecoviz.cgmlab.org/ (Botas et al., 2022). I performed this analysis using Pfam domains as queries in a similar manner to EFI-GNT. While lacking the ability to edit the SSN like EFI-GNT, GeCoViz is much more user-friendly. Using Pfam queries for DUF2169 and DUF3540, I generated the multi-domain GNN for DUF2169-PRP.

2.2 Cloning and Plasmid Creation

Primers were synthesized and purified by Integrated DNA Technologies (IDT). Phusion polymerase, restriction enzymes and T4 DNA ligase were obtained from New England Biolabs

(NEB). The list of all plasmids used in the study are listed in Table 3. *E. coli* XL-1 Blue cells were used to transform and plasmid maintenance, SM10 for conjugal transfer of allelic exchange plasmids into *P. aeruginosa* PAO1, and BL21 Codon Plus for protein overexpression. Sanger sequencing was performed by Genewiz Incorporated.

Protein overexpression and purification

DUF2169 from *Vibrio parahaemolyticus* serotype O3:K6 (strain RIMD 2210633) and *Vibrio xiamenensis* strain G21 were cloned into pET29b vector (Table 3). *E. coli* BL21 Codon Plus cells harbouring these plasmids were grown to approximately an OD of 0.6 in 1L Fernbach flask containing Luria Broth (LB) media (10 g L^{-1} NaCl, 10 g L^{-1} tryptone, 5 g L^{-1} yeast extract) at 37°C. They were then allowed to cool down to 18°C before being induced with 1mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were further grown overnight at 18°C for approximately 16h. Cells were centrifuged and resuspended in lysis buffer (50 mM Tris pH 8, 750 mM NaCl, 15 mM Imidazole and 5% Glycerol). Next, the resuspended cells were lysed using a sonicator at 30Amp for 6 cycles of 30s. Lysed cells were centrifuged, and the supernatant was passed over expanded beads nickel-column for Ni-affinity chromatography. The beads were washed with 25 column volume (CV) of lysis buffer before being eluted using elution buffer (50 mM Tris-HCl pH 8.0, 750 mM NaCl, 400 mM Imidazole, 1 mM 2-mercaptoethanol, 5% Glycerol and 400 mM Imidazole).

Protein samples were further purified by size exclusion chromatography (SEC) using a HiLoad 16/600 Superdex200 or HiLoad 16/600 Superdex75 preparatory grade column. The final SEC buffer was 50 mM Tris-HCl pH 8.0, 750 mM NaCl, 5% Glycerol and 1 mM Dithiothreitol. Reducing agents were needed to abrogate any potential disulphide bond formation.

2.3 *Vibrio xiamenensis* DUF2169 protein crystallization

V. xiamenensis DUF2169 (SAMN04488136_12145) was concentrated to 15 mg mL^{-1} for initial screening using commercially available screens (Qiagen) by sitting-drop vapour diffusion using a Crystal Gryphon robot. The initial crystallization conditions were 15 mg mL^{-1} protein concentration with a 1:1 mixture of 0.2 M Calcium Acetate and 25% (v/v) PEG 3350 at 4°C. The micro-crystals were extracted and crushed using the Seed Beads™ kit from Hampton Research. These crystal seeds (diluted 1:100) were added to varying concentrations of Calcium Acetate vs PEG 3350 in a hanging-drop vapour diffusion screen. Crystals obtained from 0.75 mM Calcium Acetate and 22% (v/v) PEG 3350 underwent a second round of seeding using the Seed Beads™. Protein samples (15 mg mL^{-1}) containing crystal seeds (diluted 1:100) were added to 48 well hanging-drop vapour diffusion screens. Commercially available Additive Screen (Hampton Research) and 0.75 mM Calcium Acetate and 22% (v/v) PEG 3350 were tested for improved crystal morphology and size. Improved crystals obtained from 0.75 mM Calcium Acetate, 22% (v/v) PEG 3350 and 2% Benzamidine HCl were further optimized using macro-seeding *ex-situ*. Intact crystals were extracted from the previous hanging drop and placed into a fresh hanging drop containing crystallization condition 0.75 mM Calcium Acetate, 22% (v/v) PEG 3350 and 2% Benzamidine HCl and 15 mg mL^{-1} protein in a 1:1 mixture. This process was repeated until large enough crystals were obtained for diffraction. Crystals were prepared by cryo-protection in mother liquor plus 7% Ethylene Glycol and flash-frozen in liquid nitrogen. The X-ray data collection and structure building were performed by Dr. Youngchan Kim at the Advanced Photon Source Argonne National Laboratory.

2.4 TAC heterologous expression in *E. coli*

For co-expression of TAC cluster proteins, genes were cloned into pET29 with C-terminal Vesicular stomatitis virus G (VSV-G) tag or pETDuet with N-terminal His and/or FLAG tag (see Table 3). *E. coli* strains expressing the desired tagged proteins were grown in 50 mL of LB broth until OD_{600} 0.6–0.8. Protein expression was induced by the addition of 1 mM IPTG, and cells were further incubated overnight at 18°C. The next day, cell cultures were centrifuged at 4600 *g* for 10 min. Pellets from 50 mL culture were resuspended in 2 mL lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM imidazole) prior to sonication. Cell lysates were cleared by centrifugation at 21,000 *g* for 30 min and the soluble fraction was loaded onto a gravity flow Ni-NTA column that had been equilibrated in lysis buffer. Ni-NTA-bound complexes were washed twice with 10 mL of lysis buffer followed by elution in 2 mL of lysis buffer containing 400 mM imidazole instead of 10 mM imidazole. This elution was loaded on SDS-PAGE and detected by western blot.

Western blot analyses of protein samples were performed using SDS-PAGE gel and buffer system and a standard western blotting protocol. After SDS-PAGE separation, proteins were wet-transferred to 0.45µm nitrocellulose membranes (100 V for 30min, 4°C). The nitrocellulose membrane was incubated with 5% (w/v) blotting grade blocker (Bio-Rad) for either 2h at 37°C or overnight at 4°C. The membrane was washed 3x with Tris-Buffered Saline (TBS) + 5% Tween. The membrane was then incubated with tag-specific rabbit primary antibodies either α-His, α-FLAG or α-VSVG (1:5000, 1 hour). The membrane was washed again 3x with Tris-Buffered Saline (TBS) + 5% Tween, then incubated with secondary antibodies either α-mouse (1:5000; primary α-His) or α-rabbit (1:5000; primary α-FLAG/α-VSVG) for 30mins. Western

blots were developed using a chemiluminescent substrate (Bio-Rad) and imaged using Chemi-Doc Imaging System (Bio-Rad).

2.5 Small-angle X-ray scattering (SAXS)

Small-angle X-ray scattering (SAXS) was performed on the VP1398 DUF2169 from *V. parahaemolyticus*. Protein was purified using the procedure as described above and concentrated to 10 mg mL^{-1} protein in SEC buffer (50 mM Tris pH 8, 500 mM NaCl, 5 % glycerol and 1mM DTT). SAXS diffractions were collected with a BioSAXS-1000 Rigaku detector. Software SAXSLab 3.1.0 was used for data collection. BioXTas Raw software was used for Guinier analysis and pair-distance distribution data processing (Hopkins et al., 2017). *Ab initio* model was built using DAMMIF via RAW with these parameters: 20 number of reconstructions, slow mode, P1 symmetry, unknown anisometry and refinement with DAMMIN. The resulting PDB file was converted into a volume map using ChimeraX (Pettersen et al., 2021) molmap function with a resolution of 5Å. Alphafold2 model of VP1398 was fit into the map using ChimeraX fit to map function with auto rotate and shift model into the map.

2.6 Genetic manipulation of *P. aeruginosa* PAO1

Two-step allelic exchange was used to generate gene deletion mutants is described by Schweizer (Schweizer 1992). For the creation of Δ*tse7-tsi7* mutant, two sets of 500bp flanks downstream of *tsi7* and upstream of *tse7* were polymerase chain reaction (PCR) amplified and then joined together using splicing by overlap extension (SOE) PCR to create approximately 1kb nucleotide product. This 1kb product, which contains homologous regions surrounding the genes being deleted, was ligated into the suicide vector pEXG2.

This plasmid was then transformed into SM10 *E. coli* prior to being conjugated into *P. aeruginosa* PAO1 Δ*retS* strain. The plasmid pEXG2 cannot replicate in *P. aeruginosa* and so the

vector is inserted into the chromosome at a low frequency through homologous recombination, which is selected for using appropriate antibiotics. A second homologous recombination event removes the integrated vector creating the desired mutant. Successively, homologous recombinant mutants are counter-selected by growing them on sucrose-containing agar plates. pEXG2 contains a copy of *sacB* gene, SacB converts sucrose to levan which is toxic to the cell. Only cells which have excised the plasmid from the chromosome will survive on the sucrose plate. Finally, the deletions were confirmed by colony PCR in strains that were resistant to sucrose, but sensitive to gentamicin. The same procedure was used to create PAO1 Δ*retS* Δ*PA0097* (DUF2169) and PAO1 Δr*etS* Δ*PA0098* (ACP) strains. The creation of Δ*DUF2169min (PA0097)* gene deletion which retains 150 nucleotides from either terminus was created in the same manner. Gene complementation experiments were performed using plasmid pPSV39-CV harbouring PA0097 or PA0098. It is used for inducible protein expression in *P. aeruginosa*. Plasmid (pPSV39) encoding either PA0097-VSV-G or PA0098 was transformed into SM10 *E. coli* prior to being conjugated into *P. aeruginosa* deletion strain.

2.7 Bacterial competition assay

A tetracycline-resistant, *lacZ*-expression cassette was inserted into a neutral phage attachment site (*attB*) of recipient *P. aeruginosa* for blue-white cell screening. Recipient and donor cells were grown overnight night in LB media at 37°C. The next day, the two cultures were OD-matched and mixed at a 1:1 (v/v) ratio. The initial count of donors:recipients were determined by plating the mixture at the correct dilution on a solid 1.5% LB agar plate supplemented with 40μg mL^{-1} X-gal. The competition mixture was spot plated on 0.45μm nitrocellulose membrane overlaid on a 3% LB agar plate and incubated face up at 37°C for 20–24h. Competitions were then harvested by scraping cells off the agar plate and resuspending in LB. As before the final

ratio of donors:recipients were determined by plating on LB agar containing 40µg mL^{-1} X-gal. The final ratio of donors:recipients colony forming units (CFU) were normalized to the initial ratios of donor and recipient strains.

Complementation was performed identically with the exception that all media were supplemented with 0.5 mM IPTG. To test for expression from pPSV39 plasmid, pull-down using VSV-G agarose beads were used. Stationary-phase overnight cultures of *P. aeruginosa* were grown at 37°C until OD 0.6 – 0.8. Upon reaching the desired OD, all samples were centrifuged at 7600 *g* for 3 min. The cell pellet was resuspended in lysis buffer (Tris 50 mM pH 8, 500 mM NaCl and 5% Glycerol). Lysed cells were centrifuged at 21,000 *g* for 25min, the supernatant was removed and incubated with VSV-G antibody linked agarose beads for 1hr. After incubation, beads were washed with lysis buffer and resuspended in a final volume of 2 CV of the initial bead volume. The beads were treated with 4X Laemmli SDS loading dye and subjected to boiling at 90°C for 5mins. The pull-down proteins were blotted using anti-VSV-G antibodies using the western blot procedure described above.

2.8 Protein alignment, structure prediction and visualization

All sequence alignments were performed using Omega Clustal (Madeira et al. 2022) and visualized using ESPript 3.0 (Robert & Gouet, 2014). The secondary structural features and structural predictions were made using AlphaFold2 Google Collab available at https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/beta/AlphaFold2_advanced.ipynb. All structural alignment and visualization were made using ChimeraX (Pettersen et al., 2021). Online server FATCAT (Flexible structure AlignmenT by Chaining Aligned fragment pairs allowing Twists) is available at https://fatcat.godziklab.org/ (Li et al., 2020).

Chapter 3: Results

3.1 Genomic Neighborhood Analysis of DUF2169

Prior publications reported the co-occurrence of DUF2169 domains with other components of the T6SS including VgrG and DUF4150 (henceforth referred to as PAAR-like) (Bayer-Santos et al., 2019; Bondage et al., 2016; Pissaridou et al., 2018). To formally consolidate these findings, I performed a GNN analysis of the DUF2169 protein family using the online tools GeCoWiz (Botas et al., 2022) and EFI-GNT (Zallot et al., 2019), which revealed three distinct gene syntenies associated with DUF2169 (Figure 4). These three syntenies also brought to light the existence of two distinct domain arrangements for DUF2169 proteins: single domain (synteny type A and B) and multidomain (synteny type C). In all cases, DUF2169 encoding genes are frequently located downstream of VgrG and upstream of PAAR-like (DUF4150) encoding genes. This gene order is generally consistent with that of other T6SS chaperone proteins, in which chaperone genes are often located in between *vgrG* and PAAR (Figure 2B).

Using GeCoWiz (Botas et al., 2022), 550 representative genomes (mostly in phylum Proteobacteria) encoding a DUF2169 domain were examined. In these genomes, 88% encoded a PAAR-like gene and 75% encoded a *vgrG* gene within five genes of DUF2169. As previously mentioned, PAAR-like genes are found in two forms: 1) ‘orphan', which only encodes the PAAR-like DUF4150 domain, and 2) ‘specialized’ PAAR-like genes that have a 3' extension encoding a toxin domain. Overall, this analysis revealed that DUF2169 is genetically linked to ‘orphan’ PAAR-like and ‘specialized’ PAAR-like fused effectors. Similar results were obtained using ‘PAAR-like’ (DUF4150) as a search query; 82% of examined genomes encode a DUF2169 protein within five genes of PAAR-like. The rate of co-occurrence between these two queries (88% vs 82%) suggests an approximate 1:1 relationship between these domain families.

In a small subset of genomes, the number of PAAR-like encoding genes exceeds the number of DUF2169 genes, which accounts for some of the discrepancies in the co-occurrence between these gene families. In genomes encoding one or more PAAR-like gene, they often encode at least one DUF2169 gene, but this gene may not be located within proximity to both PAAR-like genes, which potentially accounts for the remaining 12% of cases. These results are similar to a well-characterized T6SS chaperone family Eag, in which ~10% of *eag* genes are not located in the same operon as their cognate effector (Ahmad et al., 2020). In summary, these bioinformatic observations indicate a likely functional relationship between DUF2169 and PAAR-like domains, which suggests that DUF2169 proteins are required for the function of PAAR-like containing effectors.

For my gene co-occurrence workflow, a total of 1100 genomes were retrieved by EFI-GNT and from these genomes, I determined that 75% and 16% of single domain DUF2169s are found in type A and type B gene arrangements, respectively, while the remaining gene clusters were not as well defined and thus difficult to categorize. The type B arrangement follows the more common synteny seen in other chaperone families, in which DUF2169 is located directly between VgrG and PAAR-like encoding genes. The major difference between the two patterns of gene synteny is the occurrence of a predicted acyl-carrier protein (ACP) encoding gene located upstream of DUF2169 in the type A arrangement. These ACP genes are predicted as such based on their primary amino acid sequence, which is characterized by a conserved GxxSL/I motif (Lambalot et al., 1996). ACPs are scaffold proteins that play a role in the biosynthesis of fatty acids and function to bind fatty acyl intermediates during their synthesis. To date, there is no known correlation between fatty acid biosynthesis and the T6SS, so the role of this protein remains unclear.

Recently, an article published by Liu et al. similarly found that approximately 74.4% of these T6SS-linked ACPs are located adjacent to DUF2169 encoding genes (Liu et al. 2020). Liu et al. attempted to re-classify these ACPs as another adaptor family of proteins, which they refer to as PRK06147 (Liu et al. 2020). To stay consistent with prior literature, I will henceforth refer to DUF2169-associated ACPs as PRK06147. The bioinformatics analysis of Liu et al. found that PRK06147 proteins are typically located directly upstream of a predicted effector and downstream of VgrG, sometimes in the absence of PAAR-like and DUF2169. In their analysis, they relied solely on bioinformatic approaches to suggest an adaptor function for this protein family, and thus there remains sparse experimental work characterizing the function of these T6SS-associated PRK06147.

Beyond PRK06147 proteins, Liu et al. did not identify the existence of a small gene (~135-170aa) that is located upstream of DUF2169 and co-occurs with PRK06147 in type A DUF2169 gene arrangements (Liu et al., 2020). This small gene is sometimes annotated as an OB-fold domain (also referred to as DUF6484), the same protein fold that comprises the neck region of VgrG spike proteins. This is unusual because there is no precedent for an isolated neck region of a VgrG functioning in T6SS. In all genomes containing this OB-fold domain, it is located directly downstream of a *vgrG* gene and is near DUF2169 and PRK06147 genes. While experimental data is needed, this genetic co-occurrence suggests that these unusual OB-fold proteins may have a functional relationship with PRK06147 and potentially DUF2169 proteins encoded by type A gene arrangements.

Based on their predicted structural similarity with a region of VgrG, I hypothesize that these OB-fold proteins also form a homotrimer complex as is the case with all characterized VgrGs. Using Alphafold2 (AF2) (Jumper et al., 2021), I generated a high-confidence trimeric model of the

PA0096 OB-fold protein from *P. aeruginosa*. This predicted structure shows the protein likely resembles the gp5 needle-like structure of VgrG corresponding to a three-stranded β-prism. In some full-length VgrG proteins, this region is known as DUF2345, and it also contains the OB domain found in the middle of VgrG (Figure 5B). As shown in Figure 5B, structural alignment of the predicted OB-fold of PA0096 and DUF2345 from $VgrG^{EAEC}$ (PDB: 6Sk0) shows a nearly identical β-prism arrangement. This suggests that these OB-fold proteins may act as a molecular 'extension rod' for their cognate VgrG protein, but how the OB-fold domain interacts with VgrG remains to be determined through experimental characterization.

Interestingly, DUF2875 T6SS adaptor families have been shown to interact with the VgrG DUF2345 extension domain for the recruitment of effectors (Figure 5C) (Berni et al., 2019b). AF2-generated models for PRK06147 and the adaptor, DUF2875, reveals that PRK06147 proteins are structurally similar to DUF2875 proteins (Figure 5B). Both structures are composed of two anti-parallel β-sheets and two α-helices that occur in between the β-sheets. Additionally, there are two additional α-helices on each end of the β-sheets resulting in six total α-helices: two in the middle and four on opposite sides. Using the online server FATCAT (Flexible structure AlignmenT by Chaining Aligned fragment pairs allowing Twists), structural alignment of these two domain families indicates that they are significantly similar to one another with a P-value of $1.08x10^{-03}$ and a RMSD of 2.93Å (Li et al., 2020). A sequence alignment between PA0098 (PRK06147), PA0259 (DUF2875), and KAS II, a bonafide ACP from *E. coli*, shows low sequence identity between these proteins (31-38%) even though they all adopt the same fold (Figure 5A). This structural similarity between PRK06147 and DUF2875 combined with their association with the VgrG β-prism extension suggests that PRK06147 and DUF2875 may perform an identical function as an adaptor that connects VgrG proteins to effector proteins

(Berni et al., 2019b). One possibility is that the PRK06147 protein physically interacts with the OB-fold protein in a manner that is analogous to the interaction of the Tla3 (DUF2875) with the DUF2345 and TTR domain at the C-terminus of $VgrG^{EAEC}$ (Figure 5C).

In my previous GNN, only single domain DUF2169 genes were used as query to reveal two sets of distinct gene synteny within the DUF2169 protein family. While ~70% of the total DUF2169 genes can be characterized this way, the remainder of DUF2169-containing genes were multidomain, and in almost all cases have a Pentapeptide-Repeats Protein (PRP) domain at their C-terminus (Figure 4). DUF2169-PRP fusions are located immediately upstream of another PRP encoding gene, followed by either DUF4150 or DUF3540 genes. This gene synteny has been previously observed by Shyntym et al. who refer to DUF2169-PRP and PRP tandem domain architectures as TagAB and TagB, respectively (Shyntum et al., 2014). However, their analysis failed to show that gene layout is strongly conserved. Nearly a third of all DUF2169 encode a PRP extension and I calculated approximately ~88% of all DUF2169-PRP genes encode a PRP, PAAR-like and DUF3540 domain within ± 5 genes. PRPs are widely distributed in nature and the function of PRP is uncertain in most proteins. One characterized PRP in MfpA, shows PRP can act as DNA gyrase inhibitor (Hegde et al., 2005). Structural evidence suggests that PRP are able to achieve this by mimicking the structure of DNA allowing it to bind and inhibit DNA gyrase (Hegde et al., 2005). To date, no association with DUF2169-PRP inhibiting DNA gyrase has been reported. Meanwhile, DUF3540 encoding genes are only found within T6SS gene clusters with DUF2169-PRP synteny (TagAB and TagB).

3.2 Interbacterial Competition assays using *Pseudomonas aeruginosa* PAO1 TAC

To identify the function of DUF2169 proteins, it is first necessary to establish whether these proteins are required for the secretion of T6SS effectors. Currently, there are three characterized

effectors that are genetically linked to DUF2169 proteins: Tse7 from *P. aeruginosa*, Tde2 from *A. tumefaciens*, and VP1415 from *V. parahaemolyticus* (Ben-Yaakov & Salomon, 2019; Bondage et al., 2016; Pissaridou et al., 2018). The *P. aeruginosa* T6SS will be used as a model system for this work because DUF2169 is encoded within the very well-characterized H1-T6SS of this organism. More specifically, it lies within a synteny A type DUF2169 gene cluster that is historically referred to as T6-Associated gene Custer (TAC) as it lies immediately adjacent to the structural genes of the H1-T6SS apparatus.

The domains associated with TAC genes are described in Figure 6A. As stated above, TAC resembles type A synteny because it contains an OB-fold and PRK06147 encoding genes. VgrG1b, which is the first gene in the seven-gene cluster, possesses substantial homology to the two other VgrGs exported by the *P. aeruginosa* H1-T6SS, VgrG1c and VgrG1a (71% amino acid similarity). The third gene encodes for a single domain DUF2169 protein whereas the fifth and sixth genes encode for the Tse7-Tsi7 effector-immunity pair. Tse7 contains an N-terminal PAAR-like domain and a C-terminal DNase toxin domain (CTD), the activity of which is neutralized by Tsi7 (Pissaridou et al., 2018). Finally, the last gene within the cluster encodes a heat-repeat protein. Heat-repeat domains generally adopt a helix-forming semi-circular helical bundle shape. Some orthologues of these heat repeat proteins are found in TAC-like clusters but are not as well conserved as the previously described DUF2169 associated genes.

As previously mentioned, *P. aeruginosa* uses its H1-T6SS ‘defensively’ by only assembling the apparatus when detecting an attack from another bacterium. To mitigate the phenotypic heterogeneity that could arise from using the abovementioned approaches to activate H1-T6SS gene expression, our lab and many others use a parent strain lacking the *retS* gene. Deletion of *retS* induces the Gac/Rsm signal transduction cascade, which activates the transcription of H1-

T6SS genes (Goodman et al., 2004). To initiate my studies, I performed a bacterial competition assay between PAO1 Δ*retS* Δ*tse7*-Δ*tsi7* and its parental strain, PAO1 Δ*retS*. The results of this experiment showed that the Δ*retS* Δ*tse7*-Δ*tsi7* strain was significantly outcompeted by its parental Δ*retS* strain (Figure 6B). As the recipient strain lacks Tsi7, the Δ*retS* strain can outcompete the Δ*retS* Δ*tse7*-Δ*tsi7* strain via Tse7-mediated killing. These results are consistent with the results reported by Pissaridou et al., in which *P. aeruginosa* PA14 Tse7 provides a fitness advantage when this organism is competed against *Pseudomonas putida* in T6SS-dependent bacterial competition (Pissaridou et al., 2018).

Having reconfirmed that Tse7 is a T6SS effector involved in bacterial killing, I next tested Δ*DUF2169* and Δ*PRK06147* strains to examine whether these mutants were still able to kill susceptible recipients in a Tse7-dependent manner. Bacterial competition assays showed that both Δ*DUF2169* and Δ*PRK06147* deletion mutants displayed no fitness advantage over the *Δtse7-tsi7* recipient strain (Figure 6B). This finding suggests that these mutants are either unable to secrete Tse7 or that Tse7 is unstable in the absence of these gene products. Together, these results suggest that DUF2169 and PRK06147 proteins are required for Tse7-mediated killing.

To confirm the requirement of these genes, I conducted gene complementation by expressing the genes from an exogenous plasmid. Deleted genes were cloned into IPTG inducible pPSV39-CV vector (see Table 3), for expression in *P. aeruginosa.* Interbacterial competitions were repeated with the Δ*DUF216* and Δ*PRK06147* strains expressing the deleted genes in pPSV39-CV vector. Upon expression, the levels of Tse7 mediated killing should have been restored to levels similar to the parental Δ*retS* strain. However, upon complementation of these genes, interbacterial competitions showed that the levels of killing were not restored (Figure 6B). To validate that DUF2169 was being properly expressed, DUF2169 was cloned into pPSV39-CV with a C-

terminal VSV-G epitope tag. Western blot analysis using α-VSV-G antibody showed a strong antibody-specific band at around 44 kDa, which corresponds to the molecular weight of DUF2169-VSV-G (Figure 6B). While this gave me confidence that the DUF2169 protein was being expressed; however, the complemented strain was still unable to kill via Tse7 (Figure 6B). These results led me to consider that my full gene deletion may be disrupting cis-regulatory elements that affect the expression of downstream genes, including the effector *tse7*.

To mitigate these potential polar effects, I created a new gene deletion construct referred to as Δ*DUF2169min*, which would retain 150 nucleotides from either ends of the DUF2169 gene instead of the usual 12-15 nucleotides. While this deletion strain was still defective in Tse7 mediated killing, *in trans* expression of DUF2169 still failed to complement to the parental level of killing (Figure 6B). As before, I validated the expression of DUF2169 through western blot detection of the DUF2169-VSV-G (Figure 6B). The reason for the lack of complementation remains unclear but could possibly still be the result of a polar effect that disrupts the expression of downstream genes in the TAC cluster. The remaining body of work focusses predominantly on DUF2169 structure and function. Therefore, my efforts in troubleshooting DUF2169 complementation in *P. aeruginosa* remain unfinished. A possible future resolution to this issue would be to create a *DUF2169* gene truncation by inserting an early stop codon within the *DUF2169* ORF. This would minimize the impact of these gene deletions on any downstream genes.

3.3 Identifying potential interacting partners of TAC gene products

Biochemistry studies of the T6SS chaperones from DUF1795, DUF2875 and DUF4123 protein families have demonstrated direct interactions between these adaptors and their cognate VgrG and/or PAAR proteins (Ahmad et al., 2020; Berni et al., 2019b; Bondage et al., 2016).

Identification of DUF2169 interactions within the T6SS phage tail-like complex remains a vital aspect to uncovering its precise function. I created several *E. coli* co-expression constructs that would test the potential interactions of DUF2169 with other proteins encoded by the TAC cluster (Table 3). In a pulldown assay using Tse7 as bait, co-expression of DUF2169 with Tse7-Tsi7 or $Tse7_{CTD}$-Tsi7 (Tsi7 immunity protein is additionally required to neutralize the toxin) showed that DUF2169 pulls down both full-length Tse7 and $Tse7_{CTD}$ (Figure 7A). Initially, I used the $Tse7_{CTD}$ of this protein because I thought I would find that DUF2169 only interacts with the N-terminal PAAR-like domain of Tse7 but instead, my data suggest that it interacts with both domains. The N-terminal PAAR-like domain of Tse7 does not express well, so its overall levels of expression are much less than Tse7 and $Tse7_{CTD}$. Using DUF2169 as bait to pulldown on the PAAR-like domain of Tse7, I reciprocally showed that DUF2169 also pulls down the PAAR-like domain (Figure 7B).

Building on these results, I next created a co-expression system for VgrG1b and the PAAR-like domain of Tse7. Pissaridou et al. hypothesized potential interactions using molecular docking of the PAAR-like domain onto the tip of VgrG1b (Pissaridou et al., 2018). Although there is no structural evidence to show that these PAAR-like domains interact with their cognate VgrGs, their mutagenesis studies suggest that PAAR-like proteins, like 'true' PAARs, sit on top of the flat VgrG β-helix. Together with earlier studies on PAAR, this evidence suggests that these PAAR-like domains interact with the C-terminus of VgrG in a similar manner (Figure 3C) (Pissaridou et al., 2018; Shneider et al., 2013). I tested this potential interaction by expressing FLAG-tagged PAAR-like domain of Tse7 with VgrG1b, $VgrG1b_{\Delta C\text{-term}}$, which is a truncation of its C-terminal β-helix domain, or $VgrG1b_{C\text{-term}}$, which is its C-terminal fragment in isolation. Fortuitously, all three of these VgrG1b constructs show stable protein expression. Expression of

several other VgrG1b$_{\Delta C\text{-term}}$ constructs with different truncation lengths were attempted but only the truncation, VgrG (1-566), showed minimal degradation. Using these VgrG constructs as bait, all three constructs were co-purified with the PAAR-like domain of Tse7 (Figure 7C). I expected the PAAR-like domain to only bind VgrG1b and VgrG1b$_{C\text{-term}}$, but it was also able to bind VgrG1b$_{\Delta C term}$. I suspect that my unanticipated results may be due to non-specific hydrophobic interactions between VgrG1b and the PAAR-like domain, but specificity controls are needed to test this out further. Considering VgrG1b$_{\Delta C\text{-term}}$ was still able to bind the PAAR-like domain and the existence of earlier work from Pissaridou et al. suggesting that PAAR-like specifically binds the C-terminus of VgrG, it is possible that DUF2169 directs PAAR-like to the C-terminus of VgrG (Pissaridou et al., 2018). In the same manner as before, my three VgrG constructs were next used as bait to pull down on DUF2169-VSV-G (Figure 7D). These results show that DUF2169 interacts with full length VgrG1b. The VgrG1b$_{\Delta C\text{-terminus}}$ truncation seemingly disrupted the interaction between VgrG1b and DUF2169, suggesting this interaction occurs through the C-terminus of VgrG1b. The interaction was restored upon expressing only the C-terminus of VgrG1b (VgrG1b$_{C\text{-term}}$) with DUF2169. This supports the previous results, signifying DUF2169 specifically binds the C-terminus of VgrG1b. Finally, I wanted to test whether the interaction of PAAR-like with VgrG1b disrupts its interaction with DUF2169. To do this, I examined if DUF2169 interacts with both VgrG1b and Tse7's PAAR-like domain when both proteins are present in the experiment. As before, DUF2169 interacts with VgrG1b and VgrG1b$_{C\text{-term}}$ but not VgrG1b$_{\Delta C\text{-term}}$ whereas the PAAR-like domain interacts with all three constructs (Figure 7E).

3.4 Structure of DUF2169 from *V. xiamenensis*

The second major aim of my thesis is the structural characterization of the DUF2169 protein family. My initial attempts focused on crystallizing DUF2169 from *P. aeruginosa* PAO1, but these were unsuccessful due to poor solubility, stability, and expression of the protein. Unable to overcome these challenges, I sought out homologous proteins encoded by bacterial genomes available to me in the lab. I cloned four DUF2169 proteins: PFL_3002 from *Pseudomonas protegens* Pf-5, a previously unannotated gene from *Burkholderia cenocepacia* K56 that contained a DUF2169 domain, VP1398 from *V. parahaemolyticus* RIMD 2210633, and BTH_II1892 from *Burkholderia thailandensis* E264. BTH_II1892 DUF2169 contained a C-terminal PRP extension so I made two constructs: one with and one without its PRP extension. Among these constructs, only *V. parahaemolyticus* DUF2169 had high levels of expression and it purified as a single species using Ni-NTA chromatography. My attempts at crystallizing VP1398 eventually yielded crystals through multiple rounds of optimization. Unfortunately, however, these crystals were too small and clustered together, so it was impossible to loop an individual crystal for diffraction testing. I next attempted to crystallize close homologs of *V. parahaemolyticus* DUF2169 from *Vibrio* spp. One of these DUF2169 homolog from *V. xiamenensis* (SAMN04488136_12145) produced good quality crystals that diffracted, and I was able to solve the first experimental structure of a DUF2169 protein.

The crystal structure of DUF2169 from *V. xiamenensis* was resolved to a resolution of 1.85Å with R_{free}/R_{work} values of 0.217 and 0.182, respectively (see Table 1). Phasing was resolved by molecular replacement using an AF2 model of SAMN04488136_12145 as a search model (Figure 8B). The structure of *V. xiamenensis* DUF2169 forms a homodimer with a domain-swapped region that consists of a loop and a short α-helix contributed by each monomer (Figure

8A). This α-helix acts like a claw that grabs onto the opposite monomer. To reflect this interesting domain-swapped arrangement, these loops will henceforth be referred to as locking loops and the α-helix as the locking claw. The interface between the two monomers is largely hydrophobic with aromatic residues forming key π stacking interactions. Notably, Phe210 to Phe210 and Trp214 to Trp214 π stack to form the core of the interaction. The exterior of the claw helix directs acidic residues Glu216 and Asp212 toward a shallow groove where they form hydrogen bonds with Gln9 and Tyr230, respectively. The overall structure of DUF2169 is primarily β-stranded, which forms three anti-parallel β-sheets within the core of the protein. The three β-sheets fold into two β-sandwiches, in which two anti-parallel β-sheets stack against the same face of a third elongated anti-parallel β-sheet. This results in the protein adopting a flat oblong structure with the β-sheets forming the core of the protein and their intervening loops forming the exterior. A DALI webserver search of my structural model shows no similar structures in the Protein Data Bank (all Z values < 5), therefore DUF2169 possesses a unique fold (Holm, 2022). Interestingly, several DALI hits showed structural aligned with one pair of β-sandwiches, which when taken out of context with the rest of the protein, resembles an immunoglobulin (Ig) fold. Ig-folds typically consist of 7-9 β-strands forming two layers of anti-parallel β-sheets, resulting in a Greek key topology (Bork et al., 1994). While DUF2169 lacks the Greek key topology, it retains the same anti-parallel β-strand sandwich fold which resembles an Ig-like fold. These two structural features (locking loops and Ig-fold) are the conserved regions at a sequence level. I will expand on this observation and describe their potential function and the overall role of DUF2169 proteins in chapter 3.6.

To validate the structure of *V. xiamenensis* DUF2169, additional experiments are needed to biochemically show that the static crystal structure is biologically relevant. However due to time

limitation, I was unable to continue further to validate this structure. Instead, I will be using data generated from my earlier work with DUF2169 from *V. parahaemolyticus* (VP1398) which bears 78% sequence similarity and thus likely behaves similarly to *V. xiamenensis* DUF2169. Using purified *V. parahaemolyticus* DUF2169, I performed small-angle X-ray scattering (SAXS) to examine the structural properties of this protein in solution. Guinier and pair-distance distribution analysis of the SAXS data is summarized in Table 2. SAXS estimated the molecular weight (MW) of the protein to be 41.0 kDa (Fischer method) or 44.7 kDa (95% C.I., 39.8 - 47.1 kDa) using Bayesian approach, which is significantly lower than the calculated 77.6 kDa MW of the homodimer DUF2169. Meanwhile, the theoretical MW of *V. parahaemolyticus* DUF2169 monomer is 38.8 kDa, suggesting that this protein is a monomer in solution.

As such, I examined whether structural differences between *V. parahaemolyticus* and *V. xiamenensis* DUF2169 could explain this discrepancy. Structural alignment of the AF2 model of *V. parahaemolyticus* DUF2169 with my experimental crystal structure of *V. xiamenensis* DUF2169 shows they are nearly identical with a C-alpha RMSD of 0.69Å (Figure 9A). To validate my earlier suspicion that DUF2169 are monomers, I generated *ab initio* molecular envelope of the *V. parahaemolyticus* DUF2169 using the SAXS data. The monomeric *V. parahaemolyticus* DUF2169 structure fits well into the *ab initio* molecular envelope generated from SAXS with a mean correlation coefficient of 0.78 between the AF2 simulated envelope and SAXS envelope (Figure 9C). In the SAXS envelope, there was volume near the locking loops domain which could not be filled by the AF2 model. I hypothesize that the crystallized dimer may be an artifact of crystal packing and this dimeric arrangement forces the locking loop into a non-native conformation via interaction with another monomer. In solution, this loop may be highly dynamic and flexible, extending away from the core of the structure. My calculated SAXS

envelope supports this idea, as there is a significant amount of available volume surrounding the locking loop. Based on this, I hypothesize this loop exhibits large conformational flexibility prior to binding a protein partner and my previous pull-down data suggests that this partner is the cone-shaped PAAR-like domain of Tse7.

3.5 Structural differences between PAAR and PAAR-like

The results of my GNN showed that DUF2169 are exclusively linked to PAAR-like domains, and as such, there are expected to be distinguishing features between PAAR, and PAAR-like which mediated this exclusivity. As mentioned earlier, DUF4150 (PAAR-like) proteins are predicted to be structurally very similar to PAAR proteins (Shneider et al., 2013). However, there are several important distinctions between these two protein families as well. PAAR proteins contain conserved histidine and cysteine residues which form a zinc-binding site at tip of their conical structure whereas PAAR-like lack these residues (Shneider et al., 2013). Additionally, PAAR-like proteins lack the PAAR motifs from which 'true' PAAR proteins derive their name (Figure 10A). To explore the relationship between these protein families further, I wanted to examine the structural differences that make PAAR-like proteins distinct from 'true' PAAR.

Using AF2, I predicted the structures of PAAR and PAAR-like proteins from many different Proteobacteria in order to identify the structural differences between these two domains. One previously overlooked unique feature of PAAR-like proteins that I uncovered is a solvent-exposed hydrophobic region located near the vertex of the conical structure that is absent in 'true' PAAR. Representative structures of PAAR-like protein from *Vibrio azureus* and a 'true' PAAR protein from *V. cholerae* are shown in Figure 10 to summarize this analysis. This unique hydrophobic region of PAAR-like is conserved across all examined PAAR-like predicted

structures. A multiple sequence alignment of diverse PAAR-like proteins obtained from the reference proteome database revealed a conserved motif P(I/V)P(Y/F) that comprises this region of the structure (Figure 10A). Overlaying sequence conservation onto the AF2 model of PAAR-like from *V. parahaemolyticus* RIMD 2210633 shows that this hydrophobic region is very conserved (Figure 10B). Generally speaking, it is unusual to find large regions of surface-exposed hydrophobic residues on a globular protein because it is entropically unfavourable due to the shell-like arrangement of water molecules that form on surfaces such as this. More often, hydrophobic patches like this are suggestive of a protein interaction partner that when bound, would shield this region of the protein from solvent. It is unclear what the precise role of this hydrophobic region might be, but nonetheless it distinguishes PAAR-like proteins from PAAR proteins and thus may play a role in the delivery of PAAR-like containing effectors such as Tse7.

3.6 DUF2169 PAAR-like interface shows proline-aromatic residue interactions

As my previous work suggests, genes encoding DUF2169 domains are exclusively found with genes encoding PAAR-like domains whereas other chaperone families, such as Eag, are genetically linked to ‘true’ PAAR domains. The identification of this hydrophobic region in PAAR-like along with their co-occurrence in bacterial genomes led me to speculate that DUF2169 proteins might be required to shield these hydrophobic regions of PAAR-like via direct protein-protein interaction. In support of this, I performed an analysis of hydrophobic and conserved residues on DUF2169 proteins and found a similarly sized conserved hydrophobic region on DUF2169 (Figure 10B). In the *V. parahaemolyticus* DUF2169, this region is formed by two loops, residues 52-71 and 195-233 (locking loop) which form a shallow cavity. As discussed previously, my SAXS data suggests that this region is very dynamic and thus this region of DUF2169 likely extends out from the core of the protein. I speculate that this cavity

provides a binding pocket for PAAR-like proteins due to their shared hydrophobic surface properties and complementary shape.

Using AF2, I was able to generate a high-confidence structural prediction of a potential complex formed by DUF2169 and DUF4150 PAAR-like domains from *V. parahaemolyticus* (Figure 11A). The interface between these proteins is largely hydrophobic with the aforementioned hydrophobic surface of PAAR-like burying itself in the hydrophobic groove of DUF2169. In the model of PAAR-like, buried residues 21-44 interact with the locking loop of DUF2169 formed by residues 195-233. Interestingly, the locking loop of DUF2169, which facilitated dimer formation in my crystal structure, similarly acts like a claw that wraps around PAAR-like in the predicted complex. There are numerous key residues identified in this predicted interaction, which I speculate facilitate this hydrophobic interaction as shown in Figure 11B.

One of the main interactions facilitating this interaction is formed by the conserved motif P(I/V)P(Y/F). DUF2169 residues Phe208 and Trp212 form CH/π interactions with PAAR-like Pro38 and Pro40. CH/π interactions are a unique type of molecular interaction that have been characterized through studies on the Trp-cage mini protein (Neidigh & Andersen, 2002). Briefly, CH/π interactions typically occur when a short α-helix with two aromatic residues packs against a polyproline helix (Neidigh et al., 2002). The aromatic-proline interaction is mediated through the negatively charged π face of the aromatic residues and the partially positively charged proline ring (Pal & Chakrabarti, 1999). In the presence of aromatic residues, prolines are more likely to adopt a cis-prolyl amide bond, which is suggestive of favourable interactions between aromatic residues and prolines (Zondlo, 2013). Prolines are also more likely to adopt a cis amide bond when encoded next to an aromatic residue, which may, in part, explain the P(I/V)P(Y/F) motif having a conserved Phe/Tyr. The aromatic residues Phe208 and Trp212 are also strongly

conserved within DUF2169 (Figure 11C). Also at this interface, other residues may be involved in additional proline-aromatic interactions. Specifically, DUF2169 residues Phe224, Phe197, and Phe198 are all within proximity to two conserved prolines of PAAR-like. PAAR-like residue Pro25 may also CH/π stack with these aromatic residues, providing additional proline-aromatic interactions. In summary, the predicted DUF2169 interaction with PAAR-like is primarily mediated through contacting proline-aromatic residues. Interestingly, proline-aromatic interactions have been shown to be very important for protein stability and for protein-protein interactions (Zondlo, 2013), which leads me to speculate that DUF2169 functions as a molecular chaperone that facilitates the proper folding of its cognate PAAR-like protein.

In summary, these results have led to the generation of many testable hypotheses that can now be experimentally examined. As previously mentioned, gene disruption of *vp1398* (DUF2169) shows attenuated killing against a strain that is susceptible to killing by the VP1415 effector (Ben-Yaakov & Salomon, 2019). In this case, the PAAR-like containing effector gene, *vp1415*, is located 17 genes downstream of its potential cognate DUF2169 gene, *vp1398*. However, the authors did not examine any potential interaction between this DUF2169 protein and PAAR-like effector. Similar to what I did with the DUF2169 and the Tse7 PAAR-like containing effector from *P. aeruginosa*, I co-expressed VP1398 (DUF2169) and the PAAR-like domain of VP1415 from *V. parahaemolyticus* for co-expression in *E. coli* to test their potential interaction. The results of this experiment indicate that the PAAR-like domain pulls down when the DUF2169 is used as bait, suggesting that these two proteins interact with each other (Figure 11D). This provides some preliminary evidence that DUF2169 domains bind PAAR-like domains potentially through shielding the hydrophobic patch on their surface. Altogether the results of my

thesis provide evidence for the role of DUF2169 as a chaperone domain that is required for the delivery of PAAR-like containing effectors used in T6SS-mediated bacterial killing.

Chapter 4: Discussion

The complex network of protein domains involved in the function of the T6SS still largely remains poorly understood at the molecular level. The recruitment of effectors to the T6SS pre-injection complex involves several interacting protein domains that ensure their proper delivery of the effector into target cells. Retaining the modular nature of T6SS effectors while simultaneously maintaining the interactions between core components often requires additional accessory proteins. There have been multiple domains identified that function as chaperones to stabilize effectors including Hcp, DUF1795 (Eag), and DUF4123 (Tec). Additionally, other proteins domains such as DUF2875 act as molecular adaptors for effector recruitment to the T6SS apparatus. Together, these domains aid in the diversification of the T6SS effectors which enhance bacterial fitness (Salomon, 2016). There have been several studies characterizing the role of most of these domain families; however, the role of DUF2169 was previously unclear. This work aims to establish the role of DUF2169 in the T6SS. Through domain association predictions and structural data, I have generated data that strongly suggest that DUF2169 functions as a chaperone for PAAR-like DUF4150 'spike tip' proteins.

Through GNN analysis, I identified three distinct gene syntenies associated with DUF2169 genes. Congruent with previous studies, DUF2169 encoding genes co-occur with genes encoding PAAR-like domains and VgrGs (Bayer-Santos et al., 2019; Bondage et al., 2016; Pissaridou et al., 2018). While other known adaptor families (DUF2875, DUF4123 and DUF1795) are associated with PAAR containing effectors, I showed that DUF2169 is exclusively genetically linked to the evolutionarily distinct PAAR-like gene family. Additionally, PAAR-like proteins and DUF2169 occur at approximately the same ratio in bacterial genomes. However, there are cases in which PAAR-like genes exceed the amount of DUF2169 genes. For example, *V. azureus*

NBRC 104587 contains four PAAR-like genes and a single copy of DUF2169. A sequence alignment of the PAAR-like domains shows that these genes are 99% identical, only differing by a single nucleotide on the boundary of PAAR-like domain, suggesting that in this instance, a single copy of DUF2169 may act as a chaperone for multiple PAAR-like proteins. In comparison, the genome of *Myxococcus stipitatus* DSM 14675 contains seven copies of PAAR-like and seven copies of DUF2169, and expectedly, these paralogous genes have lower pairwise sequence identity between one another (72-83%). Thus, the total number of DUF2169 encoding genes seems to be dependent on the number of PAAR-like domains except in circumstances where paralogous PAAR-like genes are near identical. Further supporting this trend and proposed function of DUF2169, I found no instances where the number of DUF2169s exceeded the number of PAAR-like domains. This finding suggests that all PAAR-like domains need a DUF2169 and in some cases, a single DUF2169 may serve as a chaperone for multiple highly similar PAAR-like domain containing effectors.

DUF2169 and PAAR-like gene co-occurrence is consistent across the three gene syntenies, which further implies a probable functional relationship between these two domain families. This begs the question: what drives the differences in gene content observed between the three syntenies? In their review of the PAAR superfamily, Zhang et al. showed that PAAR-like DUF4150 domains are split into three sub-families (Zhang et al., 2021). While their analysis falls short of identifying any meaningful difference that could explain these PAAR-like subtypes, it is possible that the three gene syntenies identified are formed by the three sub-types of PAAR-like. One of the sub-families (E1) largely consists of so-called ‘orphan’ PAAR-like domains (i.e., genes encoding only PAAR-like domain), which is reminiscent of DUF2169 type C synteny that I identified. Empirically, DUF2169-PRP fusions were observed to co-occur more frequently with

orphan PAAR-like. At the moment, it is unclear why DUF2169-PRP shows a preference towards orphan PAAR.

At the moment, any potential functional differences between 'regular' DUF2169s and DUF2169-PRP fusions is unknown. One group found that both DUF2169-PRP and a downstream 'orphan' PRP (TagAB and TagB) genes are required for *Burkholderia glumae* T6SS-dependent virulence in rice plants (Shyntum et al., 2014). Similarly, *B. thialandensis* virulence in mice was also attenuated when TagAB and TagB gene deletions are introduced (Lennings et al., 2019). These findings suggest that the C-terminal PRP domain is an essential component for the function of DUF2169 proteins that harbour this extension, but also that the PRP domain may not be required for all DUF2169 proteins because most of them lack this domain. Prior structural work has shown that PRP domains form a quadrilateral β-helix (Hegde et al., 2005). This is superficially reminiscent of the β-helix formed by VgrG spike proteins, the top of which serves as a platform for interaction with the β-sheet containing PAAR-like domain. In contrast to VgrG trimers, which form a three-sided triangular β-helix, quadrilateral β-helices form a rectangular shape and thus these domains are symmetry mismatched. This geometric distinction between the VgrG spike and PRP domains is an important consideration for hypothesis generation for the potential function of PRP. PAAR and PAAR-like have flat triangular face that allow them to sit on top of the VgrG spike (Shneider et al., 2013). My data showing that DUF2169 likely wraps around the tip of PAAR-like makes it difficult to speculate what the role of a PRP domain's quadrilateral β-helix might be. My bioinformatic data also showed that these DUF2169-PRP fusions co-occur with DUF3540, which based on AF2 prediction, are suggested to form a triangular β-helical structure similar to the OB-fold domain of VgrG. One possibility is that this symmetry-matched DUF3540 protein may function as a spike 'extender' that extends the trigonal β-helix of the

VgrG spike. In this model, PAAR-like would interact with the flat surface of the extender rather than VgrG itself, thus creating an extra-long spike complex. The strict co-occurrence of DUF2169-PRP with the DUF3540 extender may occur because the PRP domain functions as a 'fastener' to hold the DUF3540 extender in place (Figure 12).

Domain co-occurrence also has important implications for how all of these proteins interact with the T6SS apparatus itself. Characterized chaperone/adaptor domains interact with different components of the T6SS spike complex: Eag chaperones (DUF1795) bind transmembrane domains of PAAR fused effectors while Tec adaptors (DUF4123) bind the C-terminus of PAAR or VgrG. Yet these chaperone/adaptor families rarely co-occur with the same T6SS spike complex (< 5% co-occurrence) and have never been shown to interact with the same effector, presumably because they have overlapping and potentially redundant binding domains. In contrast, however, DUF2169 frequently co-occurs with another predicted chaperone/adaptor protein, PRK06147 (degenerate ACP), as shown in type A gene synteny. While these proteins are often annotated as acyl-carrier proteins, they are unusual in that they are associated with T6SS effectors. ACP superfamily proteins are involved in fatty acid synthesis and beta-oxidation and to date, there is no known correlation between fatty acid biosynthesis and the T6SS. So the role of this domain has been tentatively assigned as another adaptor/chaperone function called PRK06147 (Liu et al., 2020). Using structural informatics, I showed that PRK06147 domains are predicted to be structurally similar to DUF2875 proteins, which have been shown to interact with the C-terminal TTR domain and triangular β-prism extension (DUF2345) of some VgrG proteins. Therefore, it is likely that PRK06147 and DUF2169 interact with different regions of VgrG spike complexes, with the data presented in my thesis supporting the model of DUF2169 interacting with PAAR-like (Figure 12).

In this work, I reported the first experimental structure of a DUF2169 protein, which is encoded by *V. xiamenensis*. The crystal structure shows a homodimer arrangement. However, my earlier SAXS experiments on a similar protein from *V. parahaemolyticus* suggests that these proteins exist as monomers in solution. Unfortunately, due to time limitations, I was unable to perform SAXS on the *V. xiamenensis* DUF2169 protein. These two proteins are closely related (78% sequence similarity), and as such, the oligomeric states are expected to be the same. Therefore, it is likely that homodimerization is an artifact of crystal packing that occurred through the stabilization of domain swapped locking loops from each protomer. The SAXS data further suggests that these locking loops are highly dynamic in solution, capable of extending in and out from the core of the protein. My previous crystallization efforts on numerous other DUF2169 proteins were foiled by solubility issues. In the few instances where soluble protein was obtained, crystallography was foiled by heterogeneous crystal morphology. Conformationally flexible regions of proteins are a well-documented hindrance to protein crystallization (Sousa, 1995). Thus, one explanation for why DUF2169 from *V. xiamenensis* crystallized while other DUF2169s did not was because of a fortuitous crystal packing artefact that only occurs with this particular protein. In the context of the T6SS spike complex, this flexible loop may act as the thumb region of a catcher's mitt that functions to grab onto the PAAR-like domain and lock it into the hydrophobic groove that exists at the palm of the mitt. In this model, prior to effector secretion, DUF2169 would shield the hydrophobic region of PAAR-like by burying it into the hydrophobic groove of DUF2169, which my modelling suggests is highly complementary to the shape surface properties of PAAR-like. Future studies that mutate key residues involved in the potential interaction interface between PAAR-like and DUF2169 would provide supporting evidence for the binding model proposed for these two domains.

Penetration of membranes by the T6SS is a sensitive task given that the T6SS-wielding bacterium needs to ensure that the proteins involved in this process don't compromise their own membranes. In my model DUF2169 shields the hydrophobic region of PAAR-like and potentially prevents it from erroneously inserting into the membranes of the producing cell. Because the function of PAAR and PAAR-like is widely thought to be membrane penetration of target cells, the existence of a molecular 'gun safety' (DUF2169) would ensure that the T6SS-containing bacterium's own membranes are accidentally compromised. To penetrate target cell membranes via the T6SS, cone-shaped proteins such as PAAR require structural features like a zinc ion and transmembrane domains to 'sharpen' and 'grease' the tip of the T6SS spike complex for penetration of the target cell (Ahmad et al. 2020; Shneider et al. 2013). Structural predictions show that the PAAR-like DUF4150 protein family has a conserved P(I/V)P(Y/F) motif in place of the canonical zinc binding site in 'true' PAAR proteins. The highly conserved Zn binding site in conical PAAR proteins consists of three histidines and one cysteine residue, which when bound to zinc, are hypothesized to stabilize the integrity of the conical structure (Shneider et al., 2013). Other distantly related PAARs such as DUF4280 found in *Bacteroides* spp., are also predicted to contain a metal binding site via four highly conserved cysteine residues (Lays et al., 2016). Meanwhile, PAAR-like domains only have one highly conserved cysteine residue which is oriented into the interior of the vertex and is usually surrounded by aspartic acid residues (adjacent to the - P(I/V)P(Y/F)- motif). Residues such as arginine, lysine and glutamines are also known to coordinate metal binding in other PAAR homologs (Shneider et al., 2013). However, the positioning of these residues in PAAR-like relative to the conserved cysteine suggests that it is highly unlikely PAAR-like coordinates a metal ion. One possibility is that T6SS spike tip proteins that lack a metal binding site, such as PAAR-like, require additional

proteins such as DUF2169 to stabilize their conical fold prior to spike ejection from the cell via the T6SS.

Although it is uncertain whether PAAR-like proteins coordinate metal ions, the tip of PAAR-like proteins contains a hydrophobic - P(I/V)P(Y/F)- motif that contributes to one of the key predicted points of contact between DUF2169 and PAAR-like. In general, hydrophobic residues are buried in the interior of proteins where solvent accessibility is low. This concept is a cornerstone of our understanding of the thermodynamics of protein folding (Rose et al., 1985). Surface exposed hydrophobic residues, like those present on PAAR-like domains, would be expected to confer instability to the overall structure. Therefore, there must be an added benefit provided by this hydrophobic region to compensate for the potential structural instability. I propose that this hydrophobic region is analogous to the TMDs found on some PAAR effectors (Ahmad et al., 2020). The TMDs within PAAR effectors are hydrophobic α-helices that are tethered to PAAR and stabilized prior to secretion by the T6SS by Eag chaperone proteins. In this model, upon secretion from the cell, chaperone-free TMDs act as molecular "grease" for effective penetration of the target cell membrane. This molecular "grease" concept can also be applied to the hydrophobic region of PAAR-like domains in which DUF2169 proteins, like Eags, function to shield the hydrophobic region until it is functionally required, i.e., once the needle complex leaves the attacking cell. One potential benefit of such a hydrophobic spike could be the direct delivery of effectors into cytoplasm, a prediction that is supported by the frequent association of both DUF1795 (Eag) and DUF2169 with effectors possessing activities that function in the target cell cytoplasm (Liu et al., 2020). Future molecular dynamics studies of effector translocation events could help in examining the role of these hydrophobic features as important determinants of effector delivery.

In the model that I've proposed, the DUF2169 interaction with PAAR-like extends to the interface of the PAAR-like and VgrG spike-tip complex. As I mentioned briefly, DUF2169 has two conserved regions: the locking loop and the Ig-like fold. While I discussed in depth the locking loop of DUF2169 potentially functioning to shield the hydrophobic region of PAAR-like, the role of the Ig-like fold lacked discourse. Currently, there are two described Ig-like folds within T6SS proteins: TTR domains and the N-terminal plug domain of Rhs cage effectors (Jurėnas et al., 2021; Shneider et al., 2013). Interestingly, both TTRs and the Rhs N-terminus plug have been shown to interact with the C-terminus β-helix of VgrG (Berni et al., 2019b; Günther et al., 2022). My pull-down experiments indicate that PA0097 (DUF2169) may also interact with the C-terminus of VgrG1b. In the complex prediction, the Ig-like fold extends to the interface between the PAAR-like and VgrG tip-spike complex. Although it is unclear whether it interacts with VgrG, the Ig-like domain of DUF2169 could serve as a recognition marker for VgrG β-helix that directs the PAAR-like capped DUF2169 complex to the VgrG spike. Another consideration would be the DUF2169 Ig-fold may interact with PRK06147 adaptors, as this may mimic the interaction between TTR domain and DUF2875. β-sandwich folds such as the Ig-fold in domains such as TTR and N-terminal plug of Rhs mimic similar interactions with VgrG currently lacks experimental evidence but certainly warrants further investigation.

As a final note on nomenclature, in this report I have used the term PAAR-like to represent DUF4150. This has primarily been done to stay consistent with the field in the naming of these DUFs and for ease of reading. In doing this however, one runs the risk of being misleading as there are other poorly characterized proteins domains that are also called PAAR-like such as DUF4280, which is frequently found in *Bacteroides* spp. (Lays et al., 2016). Adding to the complexity of this naming system are the PAAR-like sub-types, which follow the

PAAR_(A,B,C…H) nomenclature (Zhang et al., 2021). I believe the nomenclature should reflect the distinction between the different DUFs (PAAR, DUF4150, and DUF4280) rather than an arbitrary letter designation. Keeping in mind that the original 'true' PAAR domain was named after three conserved, Pro-Ala-Ala-Arg motifs, I propose that DUF4150 should be renamed to PAAR-PIPY to reflect the conserved DUF4150 signature -Pro-(Ile/Val)-Pro-Tyr- motif (shown in Figure 10). Additionally, the AF2 model of the DUF2169-DUF4150 complex shows that DUF2169 binds the PIPY motif. As the first comprehensive report to show this association, I also propose naming DUF2169 proteins PACs (PAAR-PIPY Associated Chaperones). This nomenclature is consistent with other T6SS chaperone families (Eag and Tap/Tec) in which their name provides a reminder of their molecular function.

Protein secretion is a vital part of bacterial physiology and as such some bacteria utilize protein secretion nanomachines to execute antagonistic behaviour against other bacteria. This in turn provides a fitness advantage to the bacterium but comes at a metabolic cost to the attacker due to the complexity of the protein architecture of the nanomachinery (Basler et al., 2013). Prior to their secretion, T6SS effectors are loaded onto the apparatus through conserved domain association with the T4 phage tail-like spike-tube complex. These domain associations are conserved at a gene synteny level, and as such, DUF2169 genes co-occur with *paar-like* and *vgrG* genes. In this thesis, I was able to determine the first experimental structure of a DUF2169 protein. This provides structural evidence to suggest that DUF2169 proteins are chaperones that stabilize the hydrophobic region of PAAR-like. Gaining structural insight into the role of this protein enhances our understanding of T6SS function and highlights potential mechanistic differences between different T6SS warheads. In the longer term, this may provide the necessary

molecular information needed to engineer specialized T6SSs that are able to deliver any desired protein cargo into any target cell type.

References

Ahmad, S., Tsang, K. K., Sachar, K., Quentin, D., Tashin, T. M., Bullen, N. P., Raunser, S., McArthur, A. G., Prehna, G., & Whitney, J. C. (2020). Structural basis for effector transmembrane domain recognition by type VI secretion system chaperones. *ELife*, *9*, e62816. https://doi.org/10.7554/eLife.62816

Basler, M., Ho, B. T., & Mekalanos, J. J. (2013). Tit-for-Tat: Type VI Secretion System Counterattack during Bacterial Cell-Cell Interactions. *Cell*, *152*(4), 884–894. https://doi.org/10.1016/j.cell.2013.01.042

Basler, M., & Mekalanos, J. J. (2012). Type 6 Secretion Dynamics Within and Between Bacterial Cells. *Science*, *337*(6096), 815–815. https://doi.org/10.1126/science.1222901

Bayer-Santos, E., Ceseti, L. de M., Farah, C. S., & Alvarez-Martinez, C. E. (2019). Distribution, Function and Regulation of Type 6 Secretion Systems of Xanthomonadales. *Frontiers in Microbiology*, *10*. https://www.frontiersin.org/articles/10.3389/fmicb.2019.01635

Ben-Yaakov, R., & Salomon, D. (2019). The regulatory network of *VIBRIO PARAHAEMOLYTICUS* type VI secretion system 1. *Environmental Microbiology*, *21*(7), 2248–2260. https://doi.org/10.1111/1462-2920.14594

Benz, J., Sendlmeier, C., Barends, T. R. M., & Meinhart, A. (2012). Structural Insights into the Effector – Immunity System Tse1/Tsi1 from Pseudomonas aeruginosa. *PLoS ONE*, *7*(7), e40453. https://doi.org/10.1371/journal.pone.0040453

Bernal, P., Allsopp, L. P., Filloux, A., & Llamas, M. A. (2017). The Pseudomonas putida T6SS is a plant warden against phytopathogens. *The ISME Journal*, *11*(4), 972–987. https://doi.org/10.1038/ismej.2016.169

Berni, B., Soscia, C., Djermoun, S., Ize, B., & Bleves, S. (2019a). A Type VI Secretion System Trans-Kingdom Effector Is Required for the Delivery of a Novel Antibacterial Toxin in Pseudomonas aeruginosa. *Frontiers in Microbiology*, *10*, 1218. https://doi.org/10.3389/fmicb.2019.01218

Berni, B., Soscia, C., Djermoun, S., Ize, B., & Bleves, S. (2019b). A Type VI Secretion System Trans-Kingdom Effector Is Required for the Delivery of a Novel Antibacterial Toxin in Pseudomonas aeruginosa. *Frontiers in Microbiology*, *10*, 1218. https://doi.org/10.3389/fmicb.2019.01218

Bingle, L. E., Bailey, C. M., & Pallen, M. J. (2008). Type VI secretion: A beginner's guide. *Current Opinion in Microbiology*, *11*(1), 3–8. https://doi.org/10.1016/j.mib.2008.01.006

Bondage, D. D., Lin, J.-S., Ma, L.-S., Kuo, C.-H., & Lai, E.-M. (2016). VgrG C terminus confers the type VI effector transport specificity and is required for binding with PAAR and adaptor–effector complex. *Proceedings of the National Academy of Sciences*, *113*(27). https://doi.org/10.1073/pnas.1600428113

Borgeaud, S., Metzger, L. C., Scrignari, T., & Blokesch, M. (2015). The type VI secretion system of *Vibrio cholerae* fosters horizontal gene transfer. *Science*, *347*(6217), 63–67. https://doi.org/10.1126/science.1260064

Bork, P., Holm, L., & Sander, C. (1994). The Immunoglobulin Fold. *Journal of Molecular Biology*, *242*(4), 309–320. https://doi.org/10.1006/jmbi.1994.1582

Botas, J., Rodríguez del Río, Á., Giner-Lamia, J., & Huerta-Cepas, J. (2022). GeCoViz: Genomic context visualisation of prokaryotic genes from a functional and evolutionary

perspective. *Nucleic Acids Research*, *50*(W1), W352–W357. https://doi.org/10.1093/nar/gkac367

Burkinshaw, B. J., Liang, X., Wong, M., Le, A. N. H., Lam, L., & Dong, T. G. (2018). A type VI secretion system effector delivery mechanism dependent on PAAR and a chaperone–co-chaperone complex. *Nature Microbiology*, *3*(5), 632–640. https://doi.org/10.1038/s41564-018-0144-4

Cianfanelli, F. R., Alcoforado Diniz, J., Guo, M., De Cesare, V., Trost, M., & Coulthurst, S. J. (2016). VgrG and PAAR Proteins Define Distinct Versions of a Functional Type VI Secretion System. *PLOS Pathogens*, *12*(6), e1005735. https://doi.org/10.1371/journal.ppat.1005735

Douzi, B., Brunet, Y. R., Spinelli, S., Lensi, V., Legrand, P., Blangy, S., Kumar, A., Journet, L., Cascales, E., & Cambillau, C. (2016). Structure and specificity of the Type VI secretion system ClpV-TssC interaction in enteroaggregative Escherichia coli. *Scientific Reports*, *6*(1), 34405. https://doi.org/10.1038/srep34405

Flaugnatti, N., Le, T. T. H., Canaan, S., Aschtgen, M.-S., Nguyen, V. S., Blangy, S., Kellenberger, C., Roussel, A., Cambillau, C., Cascales, E., & Journet, L. (2016). A phospholipase A $_1$ antibacterial Type VI secretion effector interacts directly with the C-terminal domain of the VgrG spike protein for delivery: Phospholipase A $_1$ antibacterial T6SS effector. *Molecular Microbiology*, *99*(6), 1099–1118. https://doi.org/10.1111/mmi.13292

Flaugnatti, N., Rapisarda, C., Rey, M., Beauvois, S. G., Nguyen, V. A., Canaan, S., Durand, E., Chamot-Rooke, J., Cascales, E., Fronzes, R., & Journet, L. (2020). Structural basis for

loading and inhibition of a bacterial T6 SS phospholipase effector by the VgrG spike. *The EMBO Journal*, *39*(11). https://doi.org/10.15252/embj.2019104129

Goodman, A. L., Kulasekara, B., Rietsch, A., Boyd, D., Smith, R. S., & Lory, S. (2004). A Signaling Network Reciprocally Regulates Genes Associated with Acute Infection and Chronic Persistence in Pseudomonas aeruginosa. *Developmental Cell*, *7*(5), 745–754. https://doi.org/10.1016/j.devcel.2004.08.020

Günther, P., Quentin, D., Ahmad, S., Sachar, K., Gatsogiannis, C., Whitney, J. C., & Raunser, S. (2022). Structure of a bacterial Rhs effector exported by the type VI secretion system. *PLOS Pathogens*, *18*(1), e1010182. https://doi.org/10.1371/journal.ppat.1010182

Hachani, A., Allsopp, L. P., Oduko, Y., & Filloux, A. (2014). The VgrG Proteins Are "à la Carte" Delivery Systems for Bacterial Type VI Effectors. *Journal of Biological Chemistry*, *289*(25), 17872–17884. https://doi.org/10.1074/jbc.M114.563429

Hegde, S. S., Vetting, M. W., Roderick, S. L., Mitchenall, L. A., Maxwell, A., Takiff, H. E., & Blanchard, J. S. (2005). A Fluoroquinolone Resistance Protein from *Mycobacterium tuberculosis* That Mimics DNA. *Science*, *308*(5727), 1480–1483. https://doi.org/10.1126/science.1110699

Hibbing, M. E., Fuqua, C., Parsek, M. R., & Peterson, S. B. (2010). Bacterial competition: Surviving and thriving in the microbial jungle. *Nature Reviews Microbiology*, *8*(1), 15–25. https://doi.org/10.1038/nrmicro2259

Holm, L. (2022). Dali server: Structural unification of protein families. *Nucleic Acids Research*, *50*(W1), W210–W215. https://doi.org/10.1093/nar/gkac387

Hood, R. D., Singh, P., Hsu, F., Güvener, T., Carl, M. A., Trinidad, R. R. S., Silverman, J. M., Ohlson, B. B., Hicks, K. G., Plemel, R. L., Li, M., Schwarz, S., Wang, W. Y., Merz, A. J., Goodlett, D. R., & Mougous, J. D. (2010). A Type VI Secretion System of Pseudomonas aeruginosa Targets a Toxin to Bacteria. *Cell Host & Microbe*, *7*(1), 25–37. https://doi.org/10.1016/j.chom.2009.12.007

Hopkins, J. B., Gillilan, R. E., & Skou, S. (2017). *BioXTAS RAW*: Improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *Journal of Applied Crystallography*, *50*(5), 1545–1553. https://doi.org/10.1107/S1600576717011438

Jana, B., Fridman, C. M., Bosis, E., & Salomon, D. (2019). A modular effector with a DNase domain and a marker for T6SS substrates. *Nature Communications*, *10*(1), 3595. https://doi.org/10.1038/s41467-019-11546-6

Jiang, F., Waterfield, N. R., Yang, J., Yang, G., & Jin, Q. (2014). A Pseudomonas aeruginosa Type VI Secretion Phospholipase D Effector Targets Both Prokaryotic and Eukaryotic Cells. *Cell Host & Microbe*, *15*(5), 600–610. https://doi.org/10.1016/j.chom.2014.04.010

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Jurėnas, D., Rosa, L. T., Rey, M., Chamot-Rooke, J., Fronzes, R., & Cascales, E. (2021). Mounting, structure and autocleavage of a type VI secretion-associated Rhs polymorphic toxin. *Nature Communications*, *12*(1), 6998. https://doi.org/10.1038/s41467-021-27388-0

Kanarek, K., Fridman, C. M., Bosis, E., & Salomon, D. (2022). *A new class of polymorphic T6SS effectors and tethers* [Preprint]. Microbiology. https://doi.org/10.1101/2022.10.27.514009

Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. A., Reid, R., Khosla, C., & Walsh, C. T. (1996). A new enzyme superfamily—The phosphopantetheinyl transferases. *Chemistry & Biology*, *3*(11), 923–936. https://doi.org/10.1016/S1074-5521(96)90181-7

Lays, C., Tannier, E., & Henry, T. (2016). Francisella IglG protein and the DUF4280 proteins: PAAR-like proteins in non-canonical Type VI secretion systems? *Microbial Cell*, *3*(11), 576–578. https://doi.org/10.15698/mic2016.11.543

Lennings, J., West, T. E., & Schwarz, S. (2019). The Burkholderia Type VI Secretion System 5: Composition, Regulation and Role in Virulence. *Frontiers in Microbiology*, *9*, 3339. https://doi.org/10.3389/fmicb.2018.03339

Li, Z., Jaroszewski, L., Iyer, M., Sedova, M., & Godzik, A. (2020). FATCAT 2.0: Towards a better understanding of the structural diversity of proteins. *Nucleic Acids Research*, *48*(W1), W60–W64. https://doi.org/10.1093/nar/gkaa443

Liang, X., Moore, R., Wilton, M., Wong, M. J. Q., Lam, L., & Dong, T. G. (2015). Identification of divergent type VI secretion effectors using a conserved chaperone domain. *Proceedings of the National Academy of Sciences*, *112*(29), 9106–9111. https://doi.org/10.1073/pnas.1505317112

Liu, Y., Zhang, Z., Wang, F., Li, D., & Li, Y. (2020). Identification of type VI secretion system toxic effectors using adaptors as markers. *Computational and Structural Biotechnology Journal*, *18*, 3723–3733. https://doi.org/10.1016/j.csbj.2020.11.003

Miyata, S. T., Kitaoka, M., Brooks, T. M., McAuley, S. B., & Pukatzki, S. (2011). Vibrio cholerae Requires the Type VI Secretion System Virulence Factor VasX To Kill Dictyostelium discoideum. *Infection and Immunity*, *79*(7), 2941–2949. https://doi.org/10.1128/IAI.01266-10

Neidigh, J. W., & Andersen, N. H. (2002). Peptide conformational changes induced by tryptophan-phosphocholine interactions in a micelle. *Biopolymers*, *65*(5), 354–361. https://doi.org/10.1002/bip.10272

Neidigh, J. W., Fesinmeyer, R. M., & Andersen, N. H. (2002). Designing a 20-residue protein. *Nature Structural Biology*, *9*(6), 425–430. https://doi.org/10.1038/nsb798

Pal, D., & Chakrabarti, P. (1999). Cis peptide bonds in proteins: Residues involved, their conformations, interactions and locations 1 1Edited by J. M. Thornton. *Journal of Molecular Biology*, *294*(1), 271–288. https://doi.org/10.1006/jmbi.1999.3217

Pei, T.-T., Li, H., Liang, X., Wang, Z.-H., Liu, G., Wu, L.-L., Kim, H., Xie, Z., Yu, M., Lin, S., Xu, P., & Dong, T. G. (2020). Intramolecular chaperone-mediated secretion of an Rhs effector toxin by a type VI secretion system. *Nature Communications*, *11*(1), 1865. https://doi.org/10.1038/s41467-020-15774-z

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF CHIMERAX: Structure visualization for researchers,

educators, and developers. *Protein Science*, *30*(1), 70–82. https://doi.org/10.1002/pro.3943

Pissaridou, P., Allsopp, L. P., Wettstadt, S., Howard, S. A., Mavridou, D. A. I., & Filloux, A. (2018). The *Pseudomonas aeruginosa* T6SS-VgrG1b spike is topped by a PAAR protein eliciting DNA damage to bacterial competitors. *Proceedings of the National Academy of Sciences*, *115*(49), 12519–12524. https://doi.org/10.1073/pnas.1814181115

Pukatzki, S., Ma, A. T., Revel, A. T., Sturtevant, D., & Mekalanos, J. J. (2007). Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proceedings of the National Academy of Sciences*, *104*(39), 15508–15513. https://doi.org/10.1073/pnas.0706532104

Quentin, D., Ahmad, S., Shanthamoorthy, P., Mougous, J. D., Whitney, J. C., & Raunser, S. (2018). Mechanism of loading and translocation of type VI secretion system effector Tse6. *Nature Microbiology*, *3*(10), 1142–1152. https://doi.org/10.1038/s41564-018-0238-z

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, *42*(W1), W320–W324. https://doi.org/10.1093/nar/gku316

Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., & Zehfus, M. H. (1985). Hydrophobicity of Amino Acid Residues in Globular Proteins. *Science*, *229*(4716), 834–838. https://doi.org/10.1126/science.4023714

Russell, A. B., Hood, R. D., Bui, N. K., LeRoux, M., Vollmer, W., & Mougous, J. D. (2011). Type VI secretion delivers bacteriolytic effectors to target cells. *Nature*, *475*(7356), 343–347. https://doi.org/10.1038/nature10244

Russell, A. B., LeRoux, M., Hathazi, K., Agnello, D. M., Ishikawa, T., Wiggins, P. A., Wai, S. N., & Mougous, J. D. (2013). Diverse type VI secretion phospholipases are functionally plastic antibacterial effectors. *Nature*, *496*(7446), 508–512. https://doi.org/10.1038/nature12074

Russell, A. B., Singh, P., Brittnacher, M., Bui, N. K., Hood, R. D., Carl, M. A., Agnello, D. M., Schwarz, S., Goodlett, D. R., Vollmer, W., & Mougous, J. D. (2012). A Widespread Bacterial Type VI Secretion Effector Superfamily Identified Using a Heuristic Approach. *Cell Host & Microbe*, *11*(5), 538–549. https://doi.org/10.1016/j.chom.2012.04.007

Salomon, D. (2016). MIX and match: Mobile T6SS MIX-effectors enhance bacterial fitness. *Mobile Genetic Elements*, *6*(1), e1123796. https://doi.org/10.1080/2159256X.2015.1123796

Salomon, D., Kinch, L. N., Trudgian, D. C., Guo, X., Klimko, J. A., Grishin, N. V., Mirzaei, H., & Orth, K. (2014). Marker for type VI secretion system effectors. *Proceedings of the National Academy of Sciences*, *111*(25), 9271–9276. https://doi.org/10.1073/pnas.1406110111

Sana, T. G., Berni, B., & Bleves, S. (2016). The T6SSs of Pseudomonas aeruginosa Strain PAO1 and Their Effectors: Beyond Bacterial-Cell Targeting. *Frontiers in Cellular and Infection Microbiology*, *6*. https://doi.org/10.3389/fcimb.2016.00061

Schwarz, S., West, T. E., Boyer, F., Chiang, W.-C., Carl, M. A., Hood, R. D., Rohmer, L., Tolker-Nielsen, T., Skerrett, S. J., & Mougous, J. D. (2010). Burkholderia Type VI Secretion Systems Have Distinct Roles in Eukaryotic and Bacterial Cell Interactions. *PLoS Pathogens*, *6*(8), e1001068. https://doi.org/10.1371/journal.ppat.1001068

Shneider, M. M., Buth, S. A., Ho, B. T., Basler, M., Mekalanos, J. J., & Leiman, P. G. (2013). PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. *Nature*, *500*(7462), 350–353. https://doi.org/10.1038/nature12453

Shyntum, D. Y., Venter, S. N., Moleleki, L. N., Toth, I., & Coutinho, T. A. (2014). Comparative genomics of type VI secretion systems in strains of Pantoea ananatisfrom different environments. *BMC Genomics*, *15*(1), 163. https://doi.org/10.1186/1471-2164-15-163

Silverman, J. M., Agnello, D. M., Zheng, H., Andrews, B. T., Li, M., Catalano, C. E., Gonen, T., & Mougous, J. D. (2013). Haemolysin Coregulated Protein Is an Exported Receptor and Chaperone of Type VI Secretion Substrates. *Molecular Cell*, *51*(5), 584–593. https://doi.org/10.1016/j.molcel.2013.07.025

Sousa, R. (1995). Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallization. *Acta Crystallographica Section D Biological Crystallography*, *51*(3), 271–277. https://doi.org/10.1107/S0907444994014009

Spínola-Amilibia, M., Davó-Siguero, I., Ruiz, F. M., Santillana, E., Medrano, F. J., & Romero, A. (2016). The structure of VgrG1 from *Pseudomonas aeruginosa* , the needle tip of the bacterial type VI secretion system. *Acta Crystallographica Section D Structural Biology*, *72*(1), 22–33. https://doi.org/10.1107/S2059798315021142

Toesca, I. J., French, C. T., & Miller, J. F. (2014). The Type VI Secretion System Spike Protein VgrG5 Mediates Membrane Fusion during Intercellular Spread by Pseudomallei Group Burkholderia Species. *Infection and Immunity*, *82*(4), 1436–1444. https://doi.org/10.1128/IAI.01367-13

Uchida, K., Leiman, P. G., Arisaka, F., & Kanamaru, S. (2014). Structure and properties of the C-terminal β-helical domain of VgrG protein from Escherichia coli O157. *The Journal of Biochemistry*, *155*(3), 173–182. https://doi.org/10.1093/jb/mvt109

Unterweger, D., Kostiuk, B., Ötjengerdes, R., Wilton, A., Diaz-Satizabal, L., & Pukatzki, S. (2015). Chimeric adaptor proteins translocate diverse type VI secretion system effectors in *Vibrio cholerae*. *The EMBO Journal*, *34*(16), 2198–2210. https://doi.org/10.15252/embj.201591163

Wan, B., Zhang, Q., Ni, J., Li, S., Wen, D., Li, J., Xiao, H., He, P., Ou, H., Tao, J., Teng, Q., Lu, J., Wu, W., & Yao, Y.-F. (2017). Type VI secretion system contributes to Enterohemorrhagic Escherichia coli virulence by secreting catalase against host reactive oxygen species (ROS). *PLOS Pathogens*, *13*(3), e1006246. https://doi.org/10.1371/journal.ppat.1006246

Wettstadt, S., Wood, T. E., Fecht, S., & Filloux, A. (2019). Delivery of the Pseudomonas aeruginosa Phospholipase Effectors PldA and PldB in a VgrG- and H2-T6SS-Dependent Manner. *Frontiers in Microbiology*, *10*, 1718. https://doi.org/10.3389/fmicb.2019.01718

Whitney, J. C., Chou, S., Russell, A. B., Biboy, J., Gardiner, T. E., Ferrin, M. A., Brittnacher, M., Vollmer, W., & Mougous, J. D. (2013). Identification, Structure, and Function of a Novel Type VI Secretion Peptidoglycan Glycoside Hydrolase Effector-Immunity Pair.

Journal of Biological Chemistry, *288*(37), 26616–26624. https://doi.org/10.1074/jbc.M113.488320

Wood, T. E., Howard, S. A., Förster, A., Nolan, L. M., Manoli, E., Bullen, N. P., Yau, H. C. L., Hachani, A., Hayward, R. D., Whitney, J. C., Vollmer, W., Freemont, P. S., & Filloux, A. (2019). The Pseudomonas aeruginosa T6SS Delivers a Periplasmic Toxin that Disrupts Bacterial Cell Morphology. *Cell Reports*, *29*(1), 187-201.e7. https://doi.org/10.1016/j.celrep.2019.08.094

Zallot, R., Oberg, N., & Gerlt, J. A. (2019). The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry*, *58*(41), 4169–4182. https://doi.org/10.1021/acs.biochem.9b00735

Zhang, D., de Souza, R. F., Anantharaman, V., Iyer, L. M., & Aravind, L. (2012). Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biology Direct*, *7*(1), 18. https://doi.org/10.1186/1745-6150-7-18

Zhang, H., Gao, Z.-Q., Wei, Y., Xu, J.-H., & Dong, Y.-H. (2013). Insights into the Cross-Immunity Mechanism within Effector Families of Bacteria Type VI Secretion System from the Structure of StTae4-EcTai4 Complex. *PLoS ONE*, *8*(9), e73782. https://doi.org/10.1371/journal.pone.0073782

Zhang, Z., Liu, Y., Zhang, P., Wang, J., Li, D., & Li, Y. (2021). PAAR Proteins Are Versatile Clips That Enrich the Antimicrobial Weapon Arsenals of Prokaryotes. *MSystems*, *6*(6), e00953-21. https://doi.org/10.1128/mSystems.00953-21

Zondlo, N. J. (2013). Aromatic–Proline Interactions: Electronically Tunable CH/π Interactions. *Accounts of Chemical Research*, *46*(4), 1039–1049. https://doi.org/10.1021/ar300087y

Zoued, A., Brunet, Y. R., Durand, E., Aschtgen, M.-S., Logger, L., Douzi, B., Journet, L., Cambillau, C., & Cascales, E. (2014). Architecture and assembly of the Type VI secretion system. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1843*(8), 1664–1673. https://doi.org/10.1016/j.bbamcr.2014.03.018

Tables

Table 1. X-ray data collection and structure refinement for *V. xiamenensis DUF2169*.

Data collection	*V. xiamenensis DUF2169*
Wavelength (Å)	0.97918
Space group	P1
Cell dimensions	
a, b, c (Å)	53.35 56.63 69.51
α, β, γ (°)	83.27 88.72 66.72
Resolution (Å)	30.40 – 1.85 (1.90 – 1.85)
Unique reflections	53211 (1298)
CC(1/2)	0.999 (0.629)
R_{merge} (%)	0.071 (1.325)
$I/\sigma I$	13.5 (1.3)
Completeness (%)	84.0 (40.9)
Redundancy	5.9 (5.6)
Refinement	
R_{work}/R_{free} (%)	0.1815 / 0.2177
Average B-factors (Å^2)	33.55
Ramachandran plot (%)	
Favored	97.19
Allowed	2.65
Outlier	0.16

Table 2. SAXS structural parameters from Guinier fits & P(r) functions using GNOM

Data collection	*V. parahaemolyticus DUF2169* (10mg mL $^{-1}$)
Guinier Analysis	
$I(0)$ (cm^{-1})	0.63 ± 0.01
R_g (Å)	36.15 ± 0.73
qR_g max (Å^{-1})	1.32
qR_gmin (Å^{-1})	0.5
Coefficient of correlation, R^2	0.91
P(r) analysis	
$I(0)$ (cm^{-1})	0.53 ± 0.003
R_g (Å)	29.12 ± 0.19
d_{max}	91.0
q range (Å^{-1})	0.0137 - 0.0366
x^2 (total estimate from GNOM)	1.621 (0.84)
Corrected Porod volume (V_p) (Å^{-3}) (ratio V_p / calculated M)	60500 (1.2)
V (Å^{-3}), M (kDa) using Fischer method (ratio M to expected)	50500, 41.0 (1.06)
M using Bayesian (95% Confidence Interval)	44.7 (39.8 – 47.1)

Table 3. Plasmids used in this study

Plasmids	Description
pEXG2::*ΔPA0097*	Creation of *ΔPA0097*(*ΔDUF2169*) mutants in *P. aeruginosa* PAO1
pEXG2::*ΔPA0098*	Creation of *ΔPA0098*(*ΔPRK06147*) mutants in *P. aeruginosa* PAO1
pEXG2::*Δtse7-tsi7*	Creation of *Δtse7-tsi7* mutants in *P. aeruginosa* PAO1
pEXG2::*ΔPA0097min*	Creation of *ΔPA0097min* (*ΔDUF2169min*) mutants in *P. aeruginosa* PAO1
pPSV39-CV::PA0097-VSV-G	Complementation of PA0097 in *P. aeruginosa* PAO1, encodes a C-terminal VSV-G tag
pPSV39-CV::PA0098	Complementation of PA0098 in *P. aeruginosa* PAO1.
pET29b::SAMN04488136_12145-His	Expression of *V. xiamenensis* DUF2169 with C-terminal His6 tag, synthesized by Genscript.
pET29b::VP1398-His6	Expression of *V. parahaemolyticus* DUF2169 with C-terminal His6
pETDuet1::HisVP1415DUF4150::empty	Expression of *Vibrio parahaemolyticus* PAAR-like domain with N-terminal His6 tag. VP1415 residues 19 - 144
pET29b::VP1398-VSV-G	Expression of *V. parahaemolyticus* DUF2169 with C-terminal VSV-G tag, for co-expression with VP1415 PAAR-like domain
pET29b::PA0097-VSV-G	Expression of *P. aeruginosa* PA01 DUF2169 with C-terminal VSV-G tag, for co-expression with PA0099 PAAR-like domain and VgrG1b.
pET29b::PA0097-His	Expression of *P. aeruginosa* PA01 DUF2169 with C-terminal His tag, for co-expression with PA0099 PAAR-like domain
pETDuet-1::empty::FLAG-PA0099(1-186)	Expression of *P. aeruginosa* PA01 PA0099 PAAR-like with N-terminal FLAG tag, for co-expression with DUF2169.
pETDuet-1::His-PA0099 ::PA0100	Co-expression of *P. aeruginosa* PA01 full length PA0099 (*tse7*) and PA0100 (*tsi7*). PA0099 contains a N-terminal His tag
pETDuet-1::His-PA0099$_{CTD}$ ::PA0100	Co-expression of *P. aeruginosa* PA01 PA0099 (*tse7*) toxin domain and PA0100 (*tsi7*). PA0099$_{CTD}$ encodes residues L188 to C-terminus stop codon and contains a N-terminal His tag.
pETDuet-1::His-PA0095(576-Cterm):: PA0099(1-186)	Co-expression of *P. aeruginosa* PA01 VgrG1b$_{C\text{-}term}$ and PA0099 PAAR-like domain (residues 1-186). VgrG1b$_{C\text{-}term}$ encodes residues 576 to C-terminus stop codon and contains a N-terminal His tag.
pETDuet-1::His-PA0095::FLAG-PA0099(1-186)	Co-expression of *P. aeruginosa* PA01 full length VgrG1b and PA0099 PAAR-like domain (residues 1-186). VgrG1b contains a N-terminal His tag.
pETDuet-1::His-PA0095(1-566)::FLAG-PA0099(1-186)	Co-expression of *P. aeruginosa* PA01 VgrG1b$_{\Delta C\text{-}term}$ and PA0099 PAAR-like domain (residues 1-186). VgrG1b contains a N-terminal His tag.
pETDuet-1::His-PA0095(576-Cterm):: empty	Expression of *P. aeruginosa* PA01 VgrG1b$_{C\text{-}term}$. VgrG1b$_{C\text{-}term}$ encodes residues 576 to C-terminus stop codon and contains a N-terminal His tag.
pETDuet-1::His-PA0095::empty	Expression of *P. aeruginosa* PA01 full length VgrG1b. VgrG1b contains a N-terminal His tag.
pETDuet-1::His-PA0095(1-566)::empty	Expression of *P. aeruginosa* PA01 VgrG1b$_{\Delta C\text{-}term}$. VgrG1b contains a N-terminal His tag.

Figures

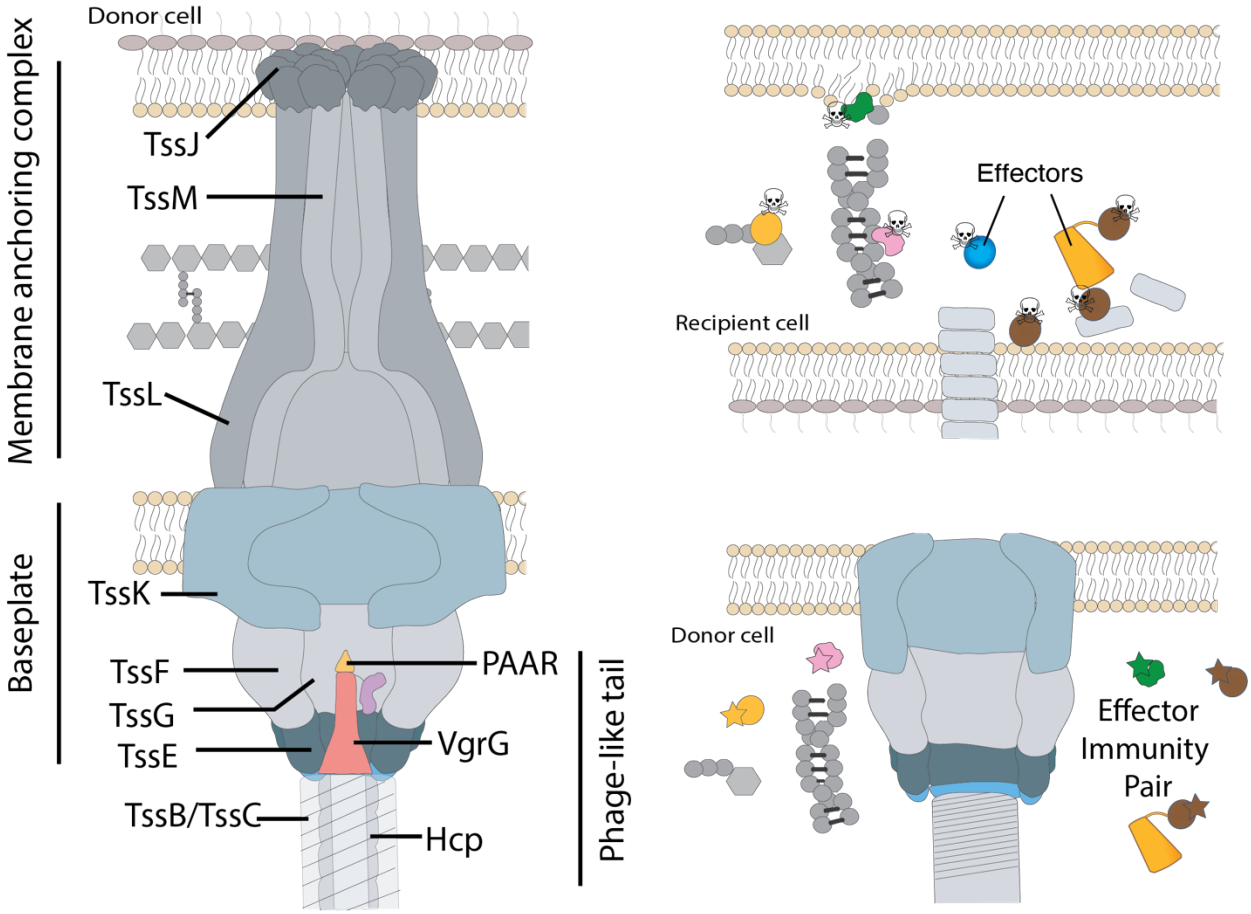

Figure 1. Schematic of the type VI secretion system and its proposed mechanism of effector export. Cartoon depiction of the major sub-complex that forms the type six secretion system on the left. Sheath contraction resulting in the ejection of the effectors into the target cell (right). The host cells are protected from self-intoxication through production of immunity proteins.

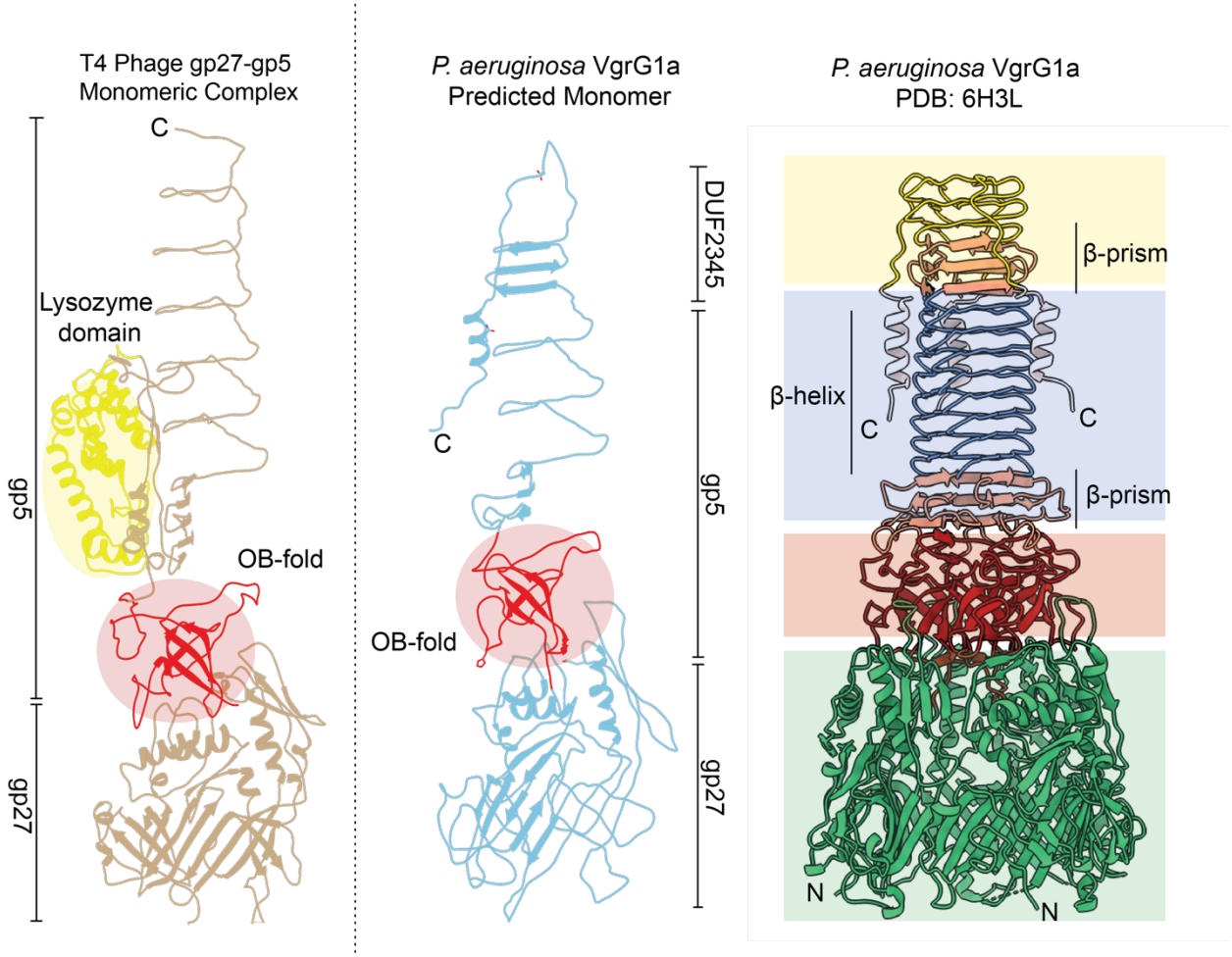

Figure 2. Domains and structure of VgrG. Monomeric comparison of domains within VgrG and phage gp27-gp5 complex. OB-fold is highlighted in red, phage gp5 lysozyme domain in yellow. Phage gp27-gp5 monomeric complex derived from the trimeric complex (PDB: 1K28). Trimeric VgrG1a (PDB: 6H3L) highlighting the differences between β-helix and β-prism regions.

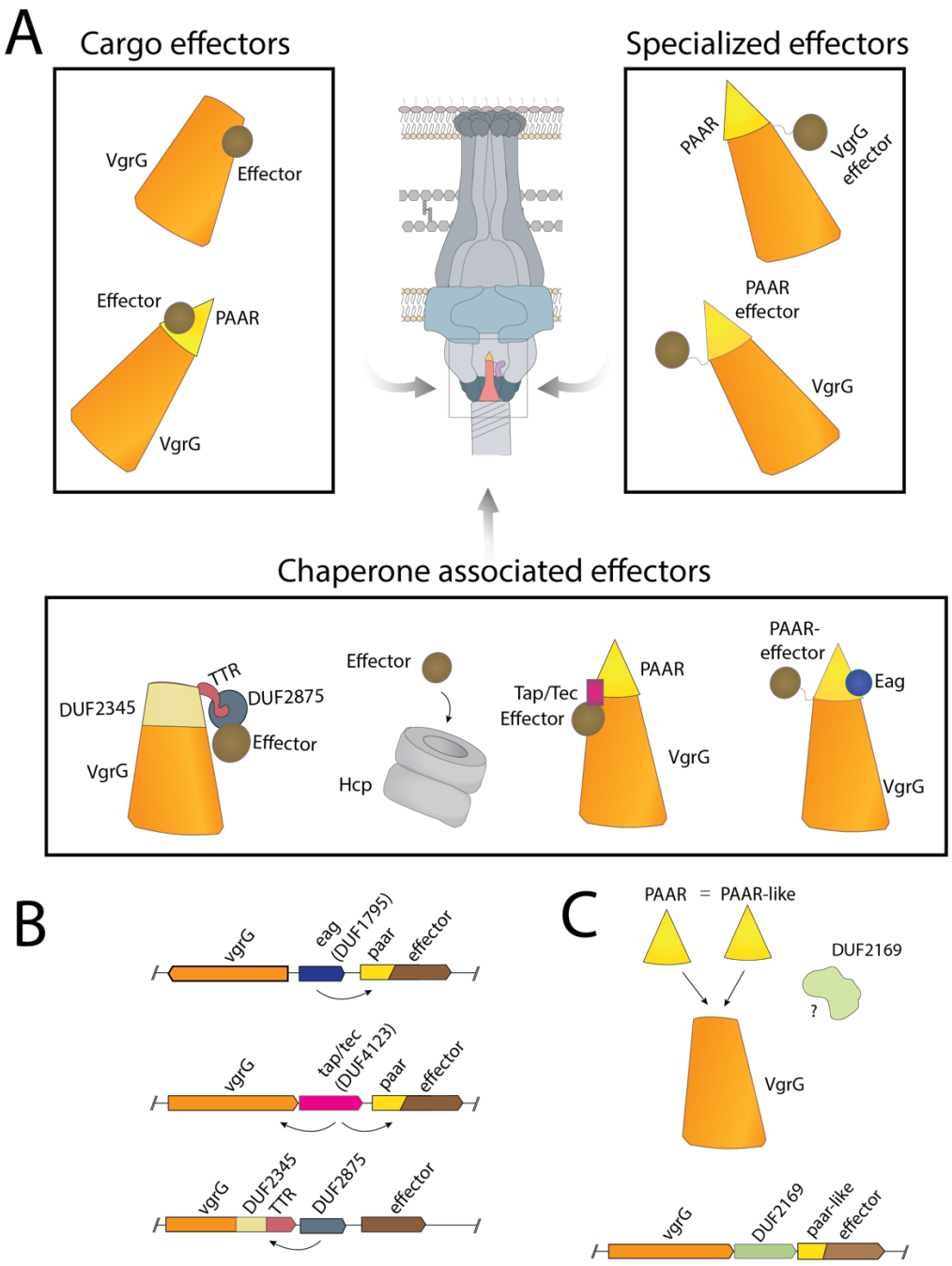

Figure 3. Schematic of effector recruitment to T6SS phage-tail tail apparatus. A) Effectors are attached to T6SS either through covalently termed "cargo effectors" or non-covalently termed "specialized effectors". Additionally, some effectors require additional chaperone proteins which are required for effector stabilization prior to secretion. B) Genomic arrangements associated with T6SS-associated chaperone proteins. C) Pictorial summary of my hypothesis. Schematic of PAAR and PAAR-like interactions with VgrG while DUF2169 interactions are unknown. This depicts the underlying questions this work aims to answer, understanding the role of DUF2169 proteins in context of these interactions.

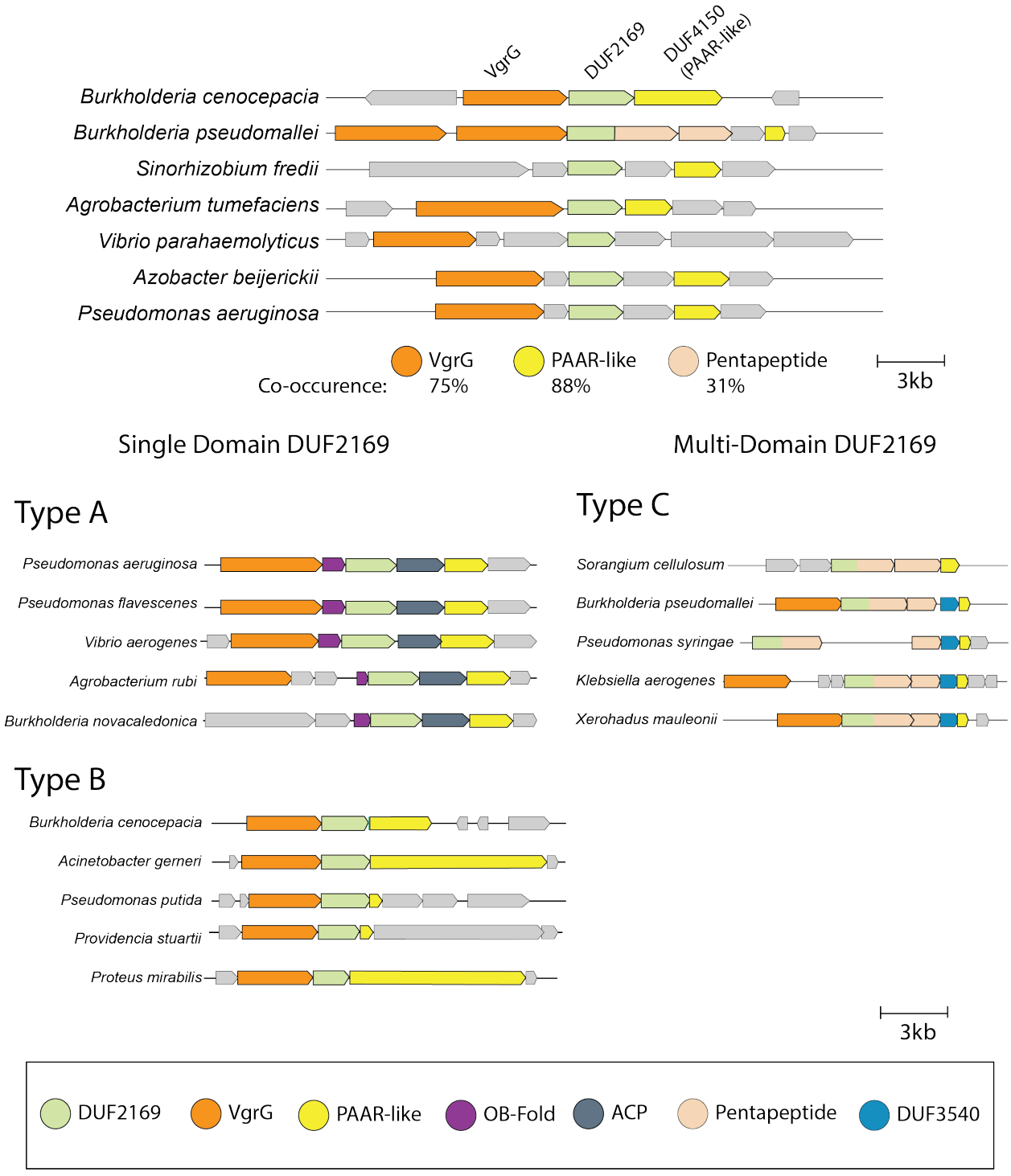

Figure 4. Genomic Neighborhood Network (GNN) summary of DUF2169 encoding genes. DUF2169 encoding genes can be categorized within 3 gene syntenies. Type A and B encode single domain DUF2169 genes and Type encodes multidomain DUF2169 with C-terminal pentapeptide repeat protein (PRP) extension. Different genes encoding VgrG, DUF4150 (PAAR-like), OB-fold, Acyl-carrier protein (ACP), PRP and DUF3540 are found to co-occur with DUF2169.

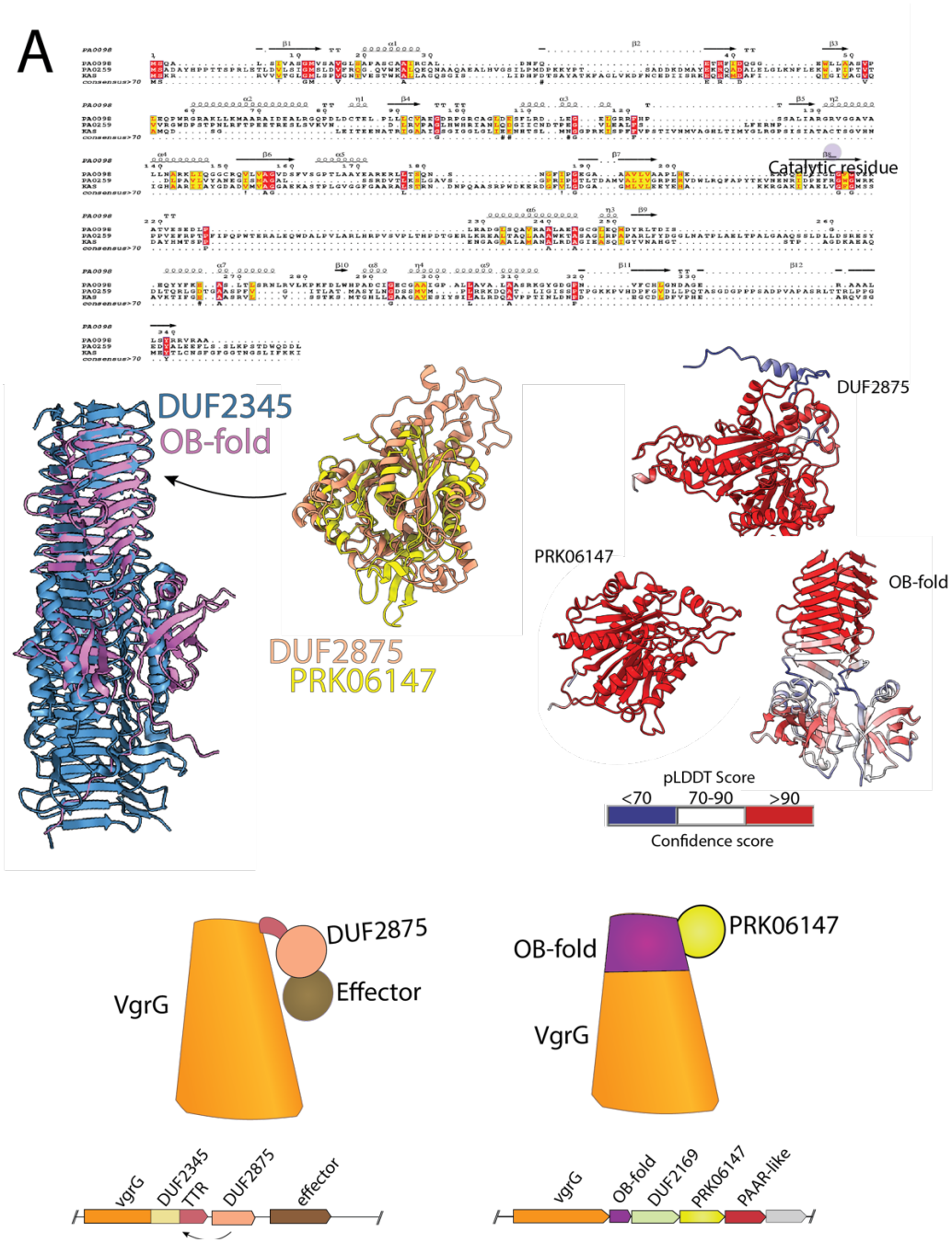

Figure 5. PRK06147 and OB-fold domains share similarities with characterized DUF2875 chaperone and VgrG extension. A) Amino acid sequence alignment of PA0098 (PRK06147), PA0259 (DUF2875) and KAS II (Acyl-carrier protein). The secondary structure predicted by AF2 of PA0098 shown above the alignment. KAS II catalytic cysteine residue is highlighted in purple (position 132) B) Structural alignment of AF2 predicted OB-fold (PA0096) domain with VgrG2b DUF2345 (PDB: 6SK0 and AF2 predicted DUF2875 (PA0259) with predicted PRK06147 (PA0098). Confidence score (pLDDT) of all predicted structures is shown on the right. C) Schematic depicting the potential interaction of PRK06147 with OB-fold compared to the similarities of DUF2875 interactions with TTR and DUF2345 domain of VgrG2b as described in Berni et al. 2019.

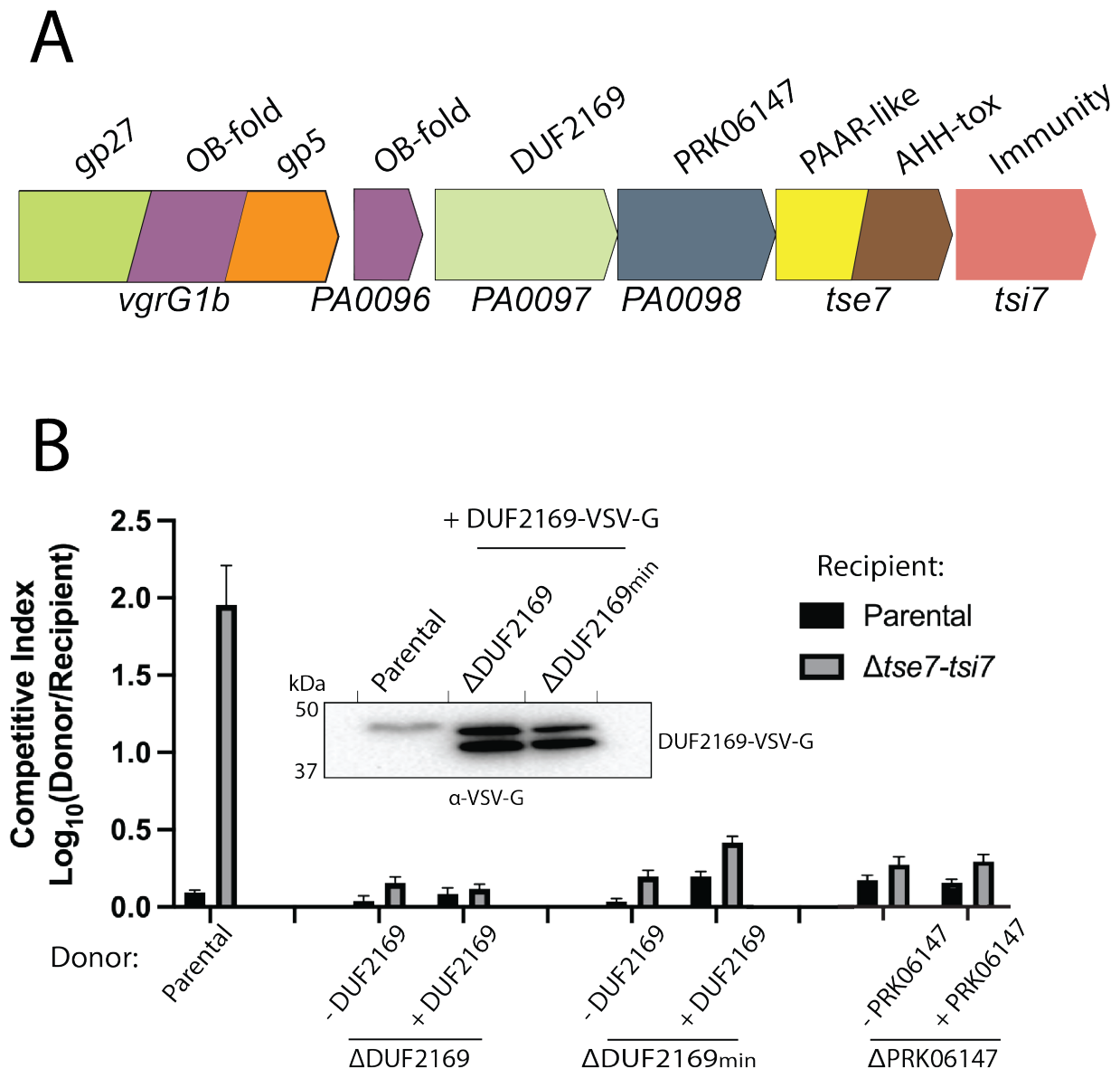

Figure 6. Interbacterial competitions of DUF2169 and PRK06147 gene deletions using *P. aeruginosa* PAO1 TAC cluster. A) Summary of protein domains encoded within *P. aeruginosa* TAC cluster. B) Interbacterial competition results for DUF2169 and PRK06147 deletions vs parental strain (PA01 Δ*retS*) or strain lacking *tse7* (Δ*retS* Δ*tse7-tsi7*). All competitions were conducted using *P. aeruginosa* PA01 with indicated gene deletion derived from the same parental strain (PA01 Δ*retS*). Western blot analysis of VSV-G bead pulldown using lysate from PAO1 Δ*retS* Δ*DUF2169* or PAO1 Δ*DUF2169min* expressing full-length DUF2169-VSV-G compared to PAO1 Δ*retS* (parental). Complementation of genes were done using pPSV39 vector (pPSV39::DUF2169-VSV-G or pPSV39::PRK06147).

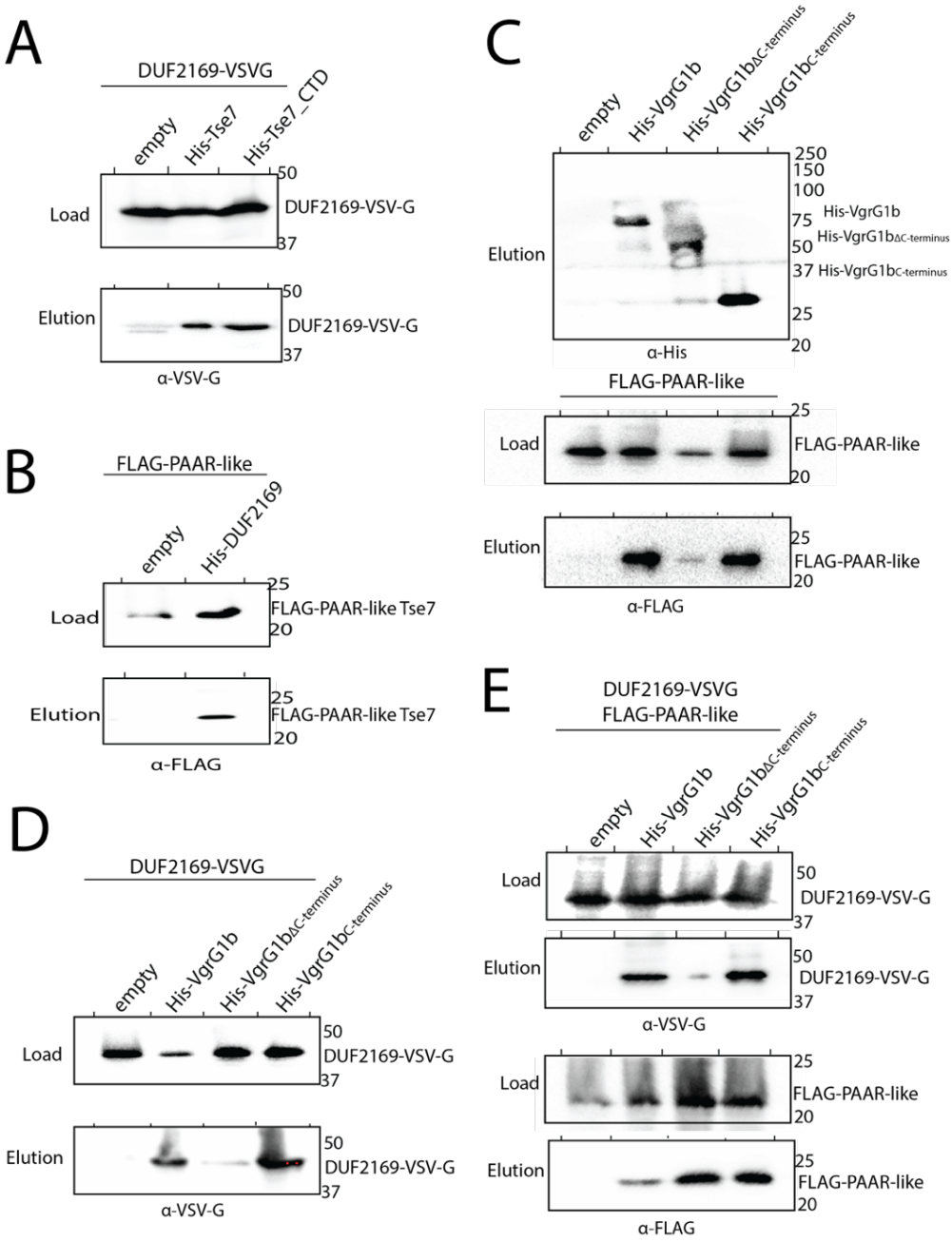

Figure 7. Identifying potential DUF2169 interactions by heterologous expression in *E. coli*. *P. aeruginosa* PAO1 TAC cluster (PA0095-PA0101) was used a model for identifying protein interaction network. Indicated domain encoding genes were cloned into either pET29b or pETDuet-1 and expressed in *E. coli* BL21 Codon Plus. The load shows western blot detection of whole cell lysate using indicated antibodies for corresponding proteins. Elution shows western blot detection results after lysate was purified using nickel affinity chromatography. A) Co-expression pull down of DUF2169-VSV-G using His-Tse7 or His-Tse7 C-terminal toxin domain (CTD) as bait. B) Pull down of FLAG-PAAR-like domain of Tse7 using His-DUF2169. C) Pull down of FLAG-PAAR-like domain of Tse7 using 3 different His-VgrG1b constructs. D) Pull down of DUF2169-VSV-G using the same 3 different His-VgrG1b constructs. F) Pull-down assay with DUF2169-VSV-G, FLAG-PAAR-like domain of Tse7 using His-VgrG1b constructs.

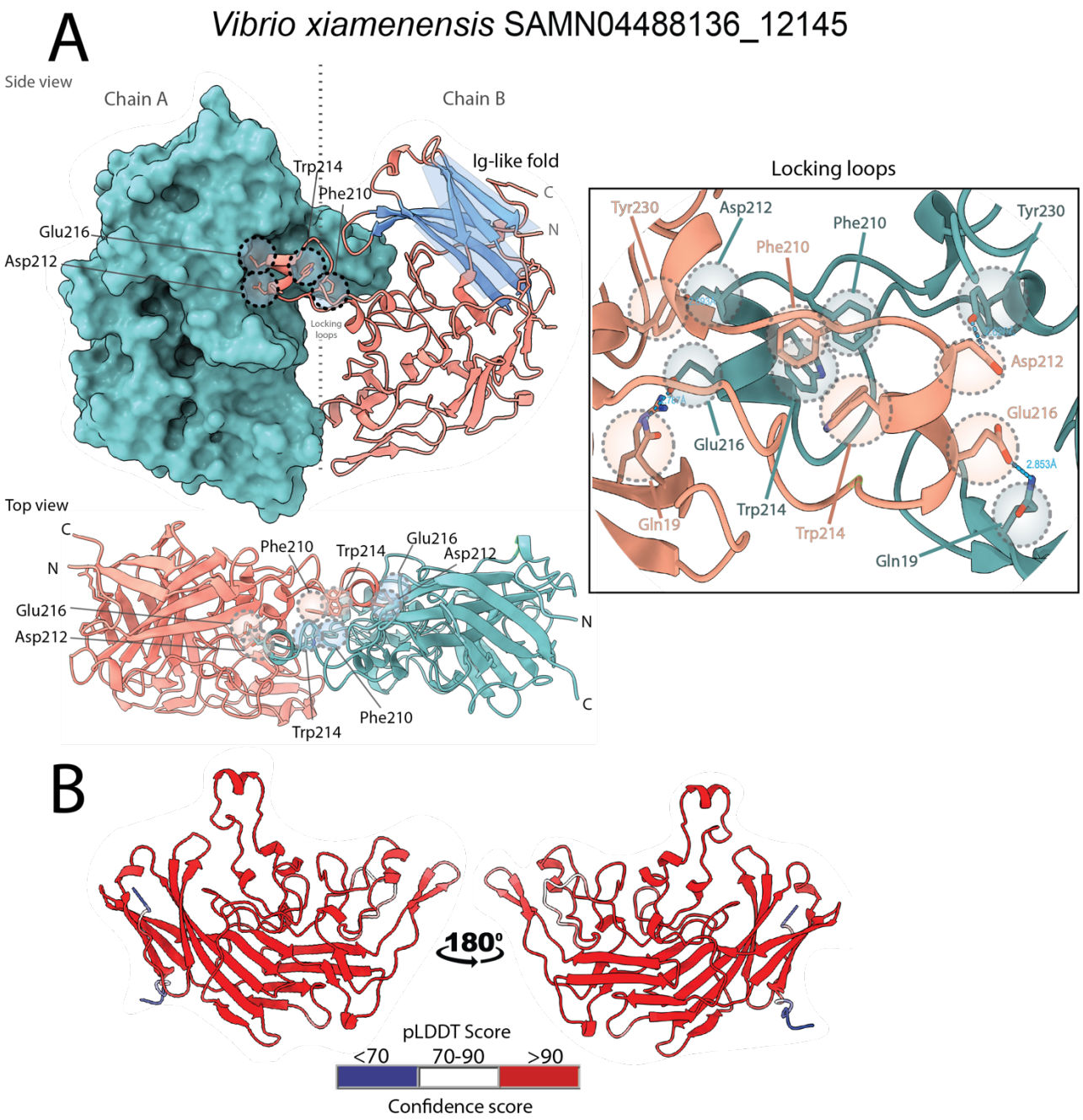

Figure 8. X-ray structure of *Vibrio xiamenensis* DUF2169 protein. A) Highlighting structural features of DUF2169. *V. xiamenensis* DUF2169 shows a homodimer with domain exchange consisting to one loop from each monomer (locking loops). Surface and ribbon representation is shown from side view, the conserved Ig-fold is highlighted in blue on ribbon representation. Interacting residues within this domain exchange from each monomer are highlighted from side and top view of the model. B) Confidence Score of AF2 model used for molecular replacement.

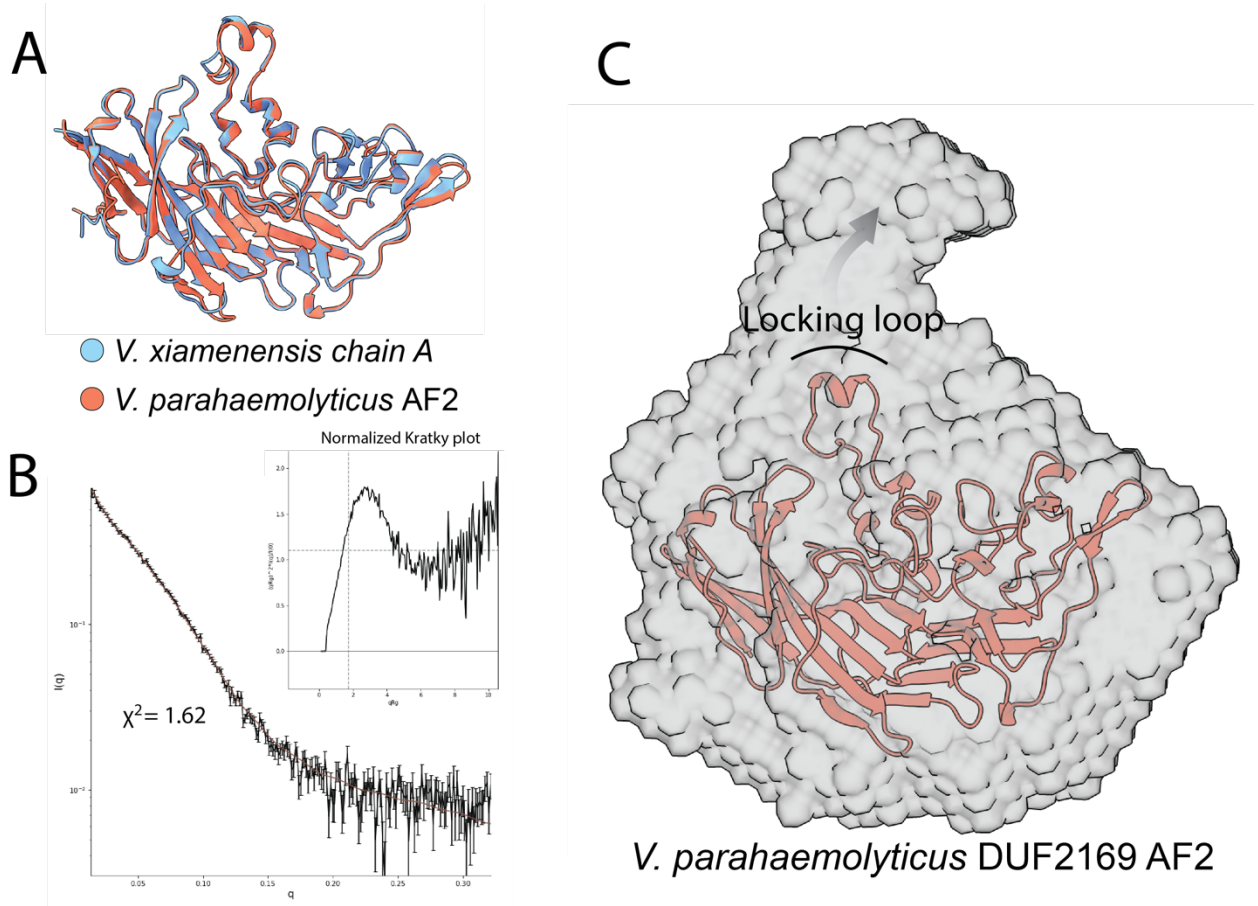

Figure 9. SAXS molecular envelope of *V. parahaemolyticus* DUF2169. A) Structural alignment of AF2 DUF2169 prediction from *V. parahaemolyticus* (VP1398) in orange and chain A of the X-ray data derived DUF2169 structure from *V. xiamenensis* (blue) are nearly identical with a C-alpha RMSD of 0.69Å. B) Experiment SAXS scattering plot, Kratky plot shows a partially unfolded protein in solution. This further supports the dynamic locking loop model. C) SAXS-derived *ab initio* envelope and AF2 predicted structural model of *V. parahaemolyticus* DUF2169.

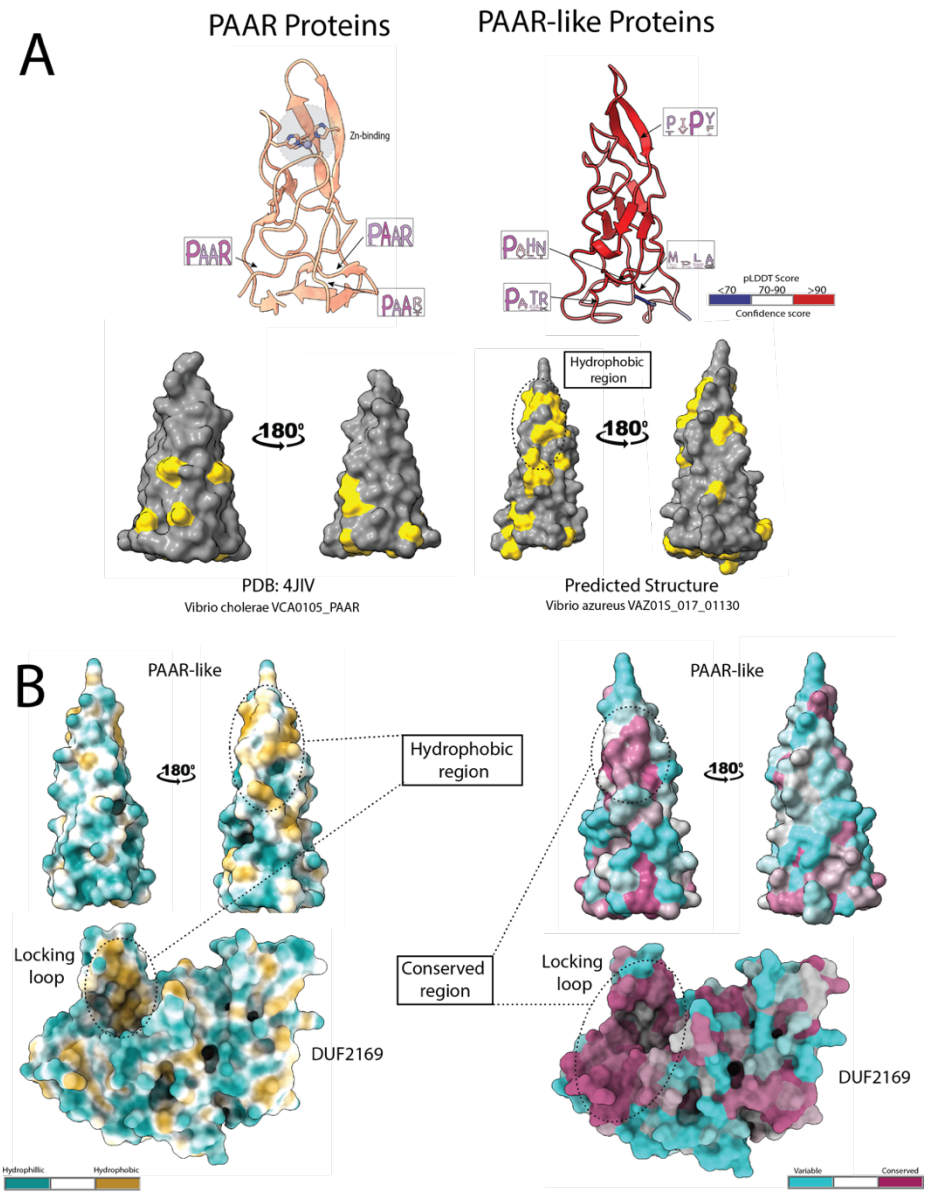

Figure 10. Structural differences between PAAR and PAAR-like protein family. A) Structural representatives for PAAR protein family using *V. cholerae* protein VCA0105 (PDB: 4JIV) and PAAR-like protein family AF2 predicted *V. azureus* VAZ01S_017_01130 PAAR-like domain (residues 1-137). Conserved PAAR motif residues and residues corresponding to the same region on PAAR-like shows that PAAR-like domain lacks the PAAR motif. Instead, PAAR-like contain a conserved P(I/V)P(Y/F) motif near the vertex of the conical structure. Surface exposed hydrophobic are highlighted in gold on the representative from each protein family. PAAR-like shows region of surface exposed hydrophobic residues which are not present in PAAR. Sequence logos were obtained from pfam seed alignment for each family. B) Conserved residues (right) and hydrophobic residue (left) surface model representation of DUF2169 (bottom) and PAAR-like domain (top) of *V. parahaemolyticus*. Highlighted regions of conserved P(I/V)P(Y/F) motif which corresponds to the hydrophobic region in PAAR-like. Conserved Ig-fold and locking loops of DUF2169 which creates a hydrophobic groove as a potential binding pocket for PAAR-like proteins.

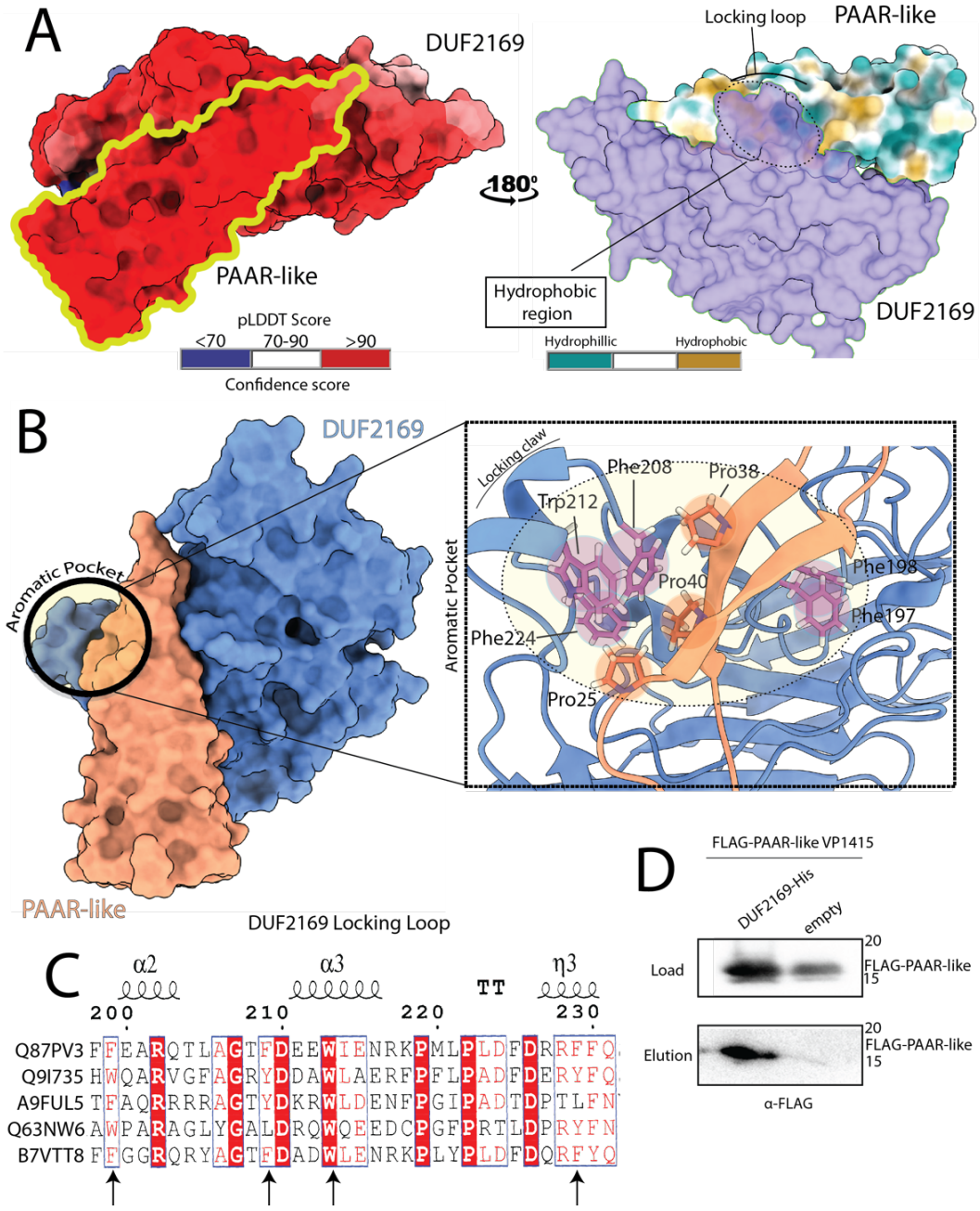

Figure 11. Predicted structural model of DUF2169 binding PAAR-like. A) High confidence (mean pLDDT score > 90) DUF2169 and PAAR-like complex prediction using AF2. The surface model of DUF2169 and hydrophobic surface of PAAR-like shows the P(I/V)P(Y/F) motif interaction with the locking loops of DUF2169. B) PAAR-like P(I/V)P(Y/F) and DUF2169 locking loops interface shows an aromatic pocket. Several key residues such as Phe198, Phe197, Phe208, Trp212, Phe224 on DUF2169 and Pro38, Pro40 and Pro25 form aromatic-proline interactions. C) Sequence alignment of DUF2169 show these key residues are conserved on DUF2169. Sequences obtained from pfam domain seed alignment. Sequences are indicated by the accession code D) Heterologous expression of *V. parahaemolyticus* DUF2169 (VP1398) and PAAR-like domain (residue 19 -144) of VP1415. Western blot detection results of pulldown of PAAR-like domain using His-tag DUF2169.

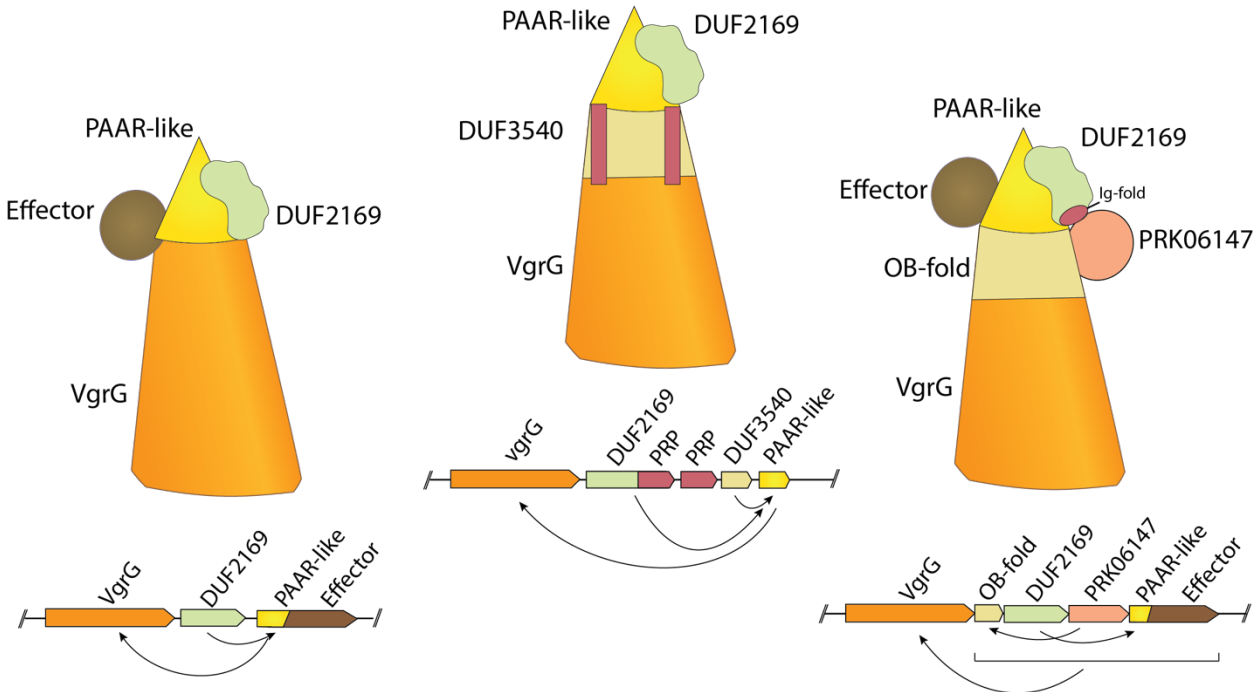

Figure 12. Pictorial summary of the hypothesized interactions between the domains associated with DUF2169. Type A and B interactions are shown on the right and the left, respectively. The middle depicts the DUF2169 type C synteny domain interactions. Arrows indicate potential domain interaction as hypothesized in discussion.