Structural Studies of the ATPase from the Type II Secretion System

Yiching C. Lu

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Wilhelmus G.J. Hol, Chair

Ronald E. Stenkamp

William N. Zagotta

Program Authorized to Offer Degree:

Biochemistry

University of Washington

Abstract

Structural Studies of the ATPase from the Type II Secretion System

Yiching C. Lu

Chair of the Supervisory Committee:

Professor Wilhelmus G.J. Hol

Biochemistry

The type II secretion system (T2SS) in Gram-negative bacteria is a multi-protein machinery that spans both the inner and the outer membranes. The T2SS in *Vibrio cholerae*, secretes folded proteins including the major virulence factor, cholera toxin, from the periplasm across the outer membrane into the intestinal tract of the host. The T2SS contains 12-15 different proteins, with many of them present in multiple copies. This system can be divided into three major sub-assemblies: the outer membrane complex, the pseudopilus, and the inner membrane complex. The inner membrane complex contains the only ATPase GspE of the T2SS, and is thought to provide the energy for T2SS assembly or secretion. GspE belongs to the Type II/IV secretion ATPase family and has three domains, N1E, N2E and CTE. GspE is tightly associated with the inner membrane complex through its N1E interacting with the cytoplasmic domain of the inner membrane protein GspL (cyto-GspL). Although numerous members of the secretion ATPases have been characterized or crystallized as hexamers, the N1E-truncated GspE crystallized as a 6_1-helical filament. To this date, the oligomeric state of GspE remains unclear.

Here, we unravel the crystal structure of the full-length GspE in complex with cyto-GspL from *Vibrio vulnificus* in the presence of AMPPNP. The structure revealed a novel mutual orientation of N2E and CTE, cyto-GspL rods that extend throughout the crystal, and unexpected interactions of cyto-GspL with AMPPNP and CTE. The GspE in complex with cyto-GspL structure shows no evidence of GspE multimers, therefore, an "assistant hexamer" Hcp1 was fused to GspE for facilitating GspE hexamerization. The N1E-truncated *Vibrio cholerae* GspE fused to Hcp1 showed a ~20-fold increase in ATPase activity compared to the monomeric form, and crystallized into two different hexamers: a quasi-C6 regular hexamer and an elongated C2 hexamer. Small-angle X-ray scattering studies of the fusion supports the hexameric forms observed in the structures are present in solution, but several conformations are in equilibrium in solution. Based on the analysis of domain arrangements and comparison with homologs from the secretion ATPases, GspE shows considerable conformational flexibility and likely undergoes several different conformations during the secretion process.

Table of Contents

Chapter 1

Introduction

1.1 PROTEIN SECRETION

Bacteria have devised highly efficient methods to transport proteins from their site of synthesis to their appropriate subcellular locations or extracellular environment. Gram-negative bacteria have essentially four cellular compartments: the cytoplasm, and the cell envelope consisting of the inner membrane, the periplasm and the outer membrane. Proteins for extracellular localization must move from an interior cellular compartment to the extracellular space. This process of protein secretion across the cell envelope plays a significant role in the phenomenon of bacterial pathogenesis. The targeting of proteins in the cytoplasm to the extracellular locale requires special mechanisms to overcome a significant problem: getting primarily hydrophilic exoproteins (proteins secreted across the outer membrane) across the hydrophobic double lipid bilayer barrier separating the cell from the environment. There are seven systems, with variations thereof, which have been identified and characterized, utilized by bacteria to overcome this hydrophobic barrier. One of these is the focus of this thesis: the signal sequence-dependent Type II secretion system (T2SS).

Type II Secretion System

The T2SS uses a two-step mechanism for the extracellular export of specific sets of exoproteins. The first step of the mechanism, the translocation of the protein precursors into the periplasm, involves the general secretion (Sec) pathway or, in rare cases, the alternative twin-arginine translocation (Tat) pathway. For those exoproteins using the Sec-pathway, the protein is initially synthesized in the cytoplasm as a precursor with an N-

terminal signal sequence. Generally, a specific chaperone, SecB, holds the precursor in an unfolded state and directs it to the Sec pathway, which translocates the precursor polypeptide to the periplasm [1]. Once reaching the periplasm, the N-terminal signal sequence is cleaved off by a signal peptidase, and the protein is then folded and assembled before being secreted. Regarding the Tat-system, protein folding takes place in the cytoplasm and the folded protein is translocated across the inner membrane [2].

The second step is the Type II secretion mechanism, also known as the General Secretory Pathway (GSP), which is responsible for the translocation of exoproteins from the periplasm across the outer membrane into the extracellular space. The T2SS shares many features with other transport related processes. Several components involved in the biogenesis of type IV pilus system (T4PS) are structurally and functionally similar to the components of the T2SS. In some species, the two systems share the same prepilin peptidase, which is required for processing critical proteins of the T2SS and T4PS as will be discussed later.

It is unknown how the T2SS is able to distinguish between the proteins that will remain in the periplasm and those that will be secreted. Proteins secreted by the T2SS are translocated across the outer membrane in their folded state [3]. The periplasm is an oxidizing environment, unlike the reducing environment of the cytoplasm, which allows the formation of disulfide bonds using a specialized disulfide oxidoreductase, DsbA, and disulfide isomerase, DsbC [4]. The formation of disulfide bonds is critical to the tertiary and quaternary structure of multi-subunit toxins, such as the cholera toxin (CT) and heat-labile enterotoxin (LT), two closely related major virulence factors of *V. cholerae* and enterotoxigenic *E. coli* (ETEC), respectively. CT and LT are complex heterohexameric AB_5 toxins that are assembled in the periplasm and translocated as folded holoenzyme complexes across the outer membrane by the T2SS [5].

The T2SS was first identified in *Klebsiella oxytoca* [6]. This system has been found in both pathogenic and non-pathogenic Gram-negative bacteria. In most bacteria, a diversity

of proteins is substrates of the T2SS. For instance, *V. cholerae* translocates not only cholera toxin, but also proteases, lipases, chitinases, etc., up to ~23 different proteins [7]. Other proteins secreted through the T2SS include alkaline phosphatase and elastase in *Pseudomonas aeruginosa* [8], pullulanase in *Klebsiella pneumoniae* [6], pectinases and cellulases in *Dickeya dadantii* and *Erwinia carotovora* [9], aerolysin in *Aeromonas hydrophila* [10], and proteases and pectinases in *Xanthomonas campestris* [11], etc. Although most T2SS proteins are highly conserved across species, the T2SS in each species is specific for a particular set of exoproteins; the exoproteins from one organism can usually not be secreted through heterologous systems [12-15]. An exception is CT, which can be secreted by ETEC, and LT, which can be secreted by *V. cholerae* [16]. There is no common secretion sequence identified among the exoproteins that are exported by the T2SS. It is possible that there are tertiary structural motifs among these proteins to which the T2SS recognizes [17].

The T2SS consists of 12-16 proteins, differing between species. The nomenclature of the proteins in the literature is complex. Here we will follow the generic names Gsp, from the main terminal branch of the general secretion pathway [18], followed by a capital letter A-O or S [19-24]. This is also the nomenclature used in various *E. coli* species. In other cases the names of T2SS proteins are often species-specific. For instance, in *V. cholerae*, the T2SS proteins are called Eps plus a capital letter A-O. In *P. aeruginosa*, two T2SSs occur with the initial three letters Xcp and Hxc, followed by capital letters P-Z. A quite complete Table of the names of T2SS proteins is given in Table 1 of the review by Korotkov et al. [22].

The genes encoding T2SS proteins are often largely arranged in one transcribed polycistronic gene cluster. The protein GspO, in some cases, is encoded by a gene elsewhere on the chromosome as part of a distinct operon encoding genes involved in the T4PS, which will be briefly described later. The protein GspS is encoded separately. Critical mutations or deletions of any one of the T2SS genes generally cause a secretion defect resulting in the periplasmic accumulation of normally secreted proteins.

A potential model of the T2SS apparatus is given in Figure 1-1. The T2SS is a large multi-protein complex spanning the inner and outer membranes. The main components of the T2SS are often divided into four sub-assemblies: the outer-membrane complex, the pseudopilus, the inner-membrane platform, and the cytoplasmic ATPase. Exoproteins are thought to be pushed through the outer membrane pore comprised of 12 subunits of the secretin GspD. The pseudopilus, composed of the minor pseudopilins GspH-K and the most abundant major pseudopilin GspG, is likely to act as a piston pushing the exoprotein through the secretin [25]. The process is energy dependent with the cytoplasmic protein GspE performing ATP hydrolysis [26, 27]. The cytoplasmic ATPase GspE is associated with the T2SS assembly through interactions with GspL, a protein in the inner membrane platform that also includes three other inner membrane proteins GspC, GspF and GspM. This model does not include the GspA, GspB, GspN and GspO proteins. GspA and GspB are not found and required in many species. The role of GspN is currently unknown and not every species has a GspN identified. GspO is required for the T2SS assembly, but it is not part of the T2SS apparatus. The major individual components of the T2SS are described in more detail below.

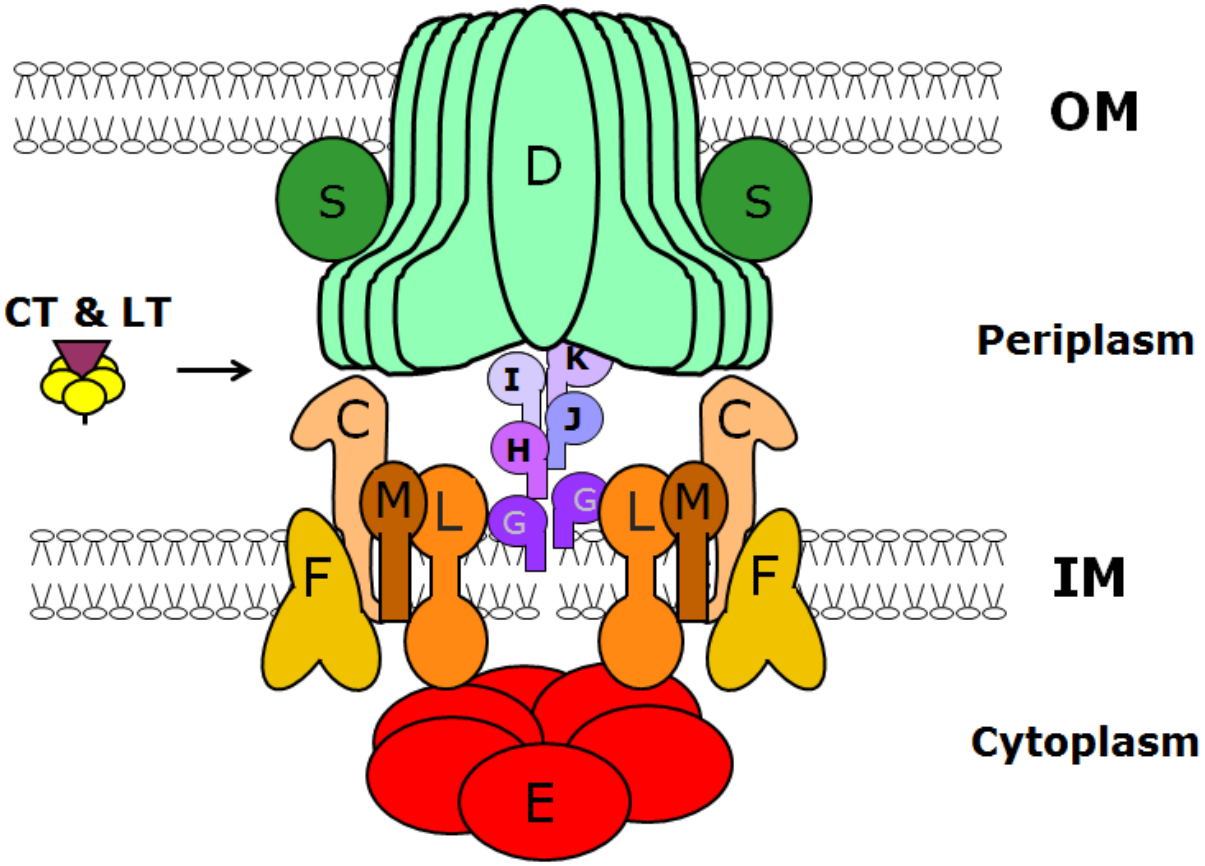


Figure 1-1: Schematic depiction of the T2SS apparatus. The proteins are labeled C to M and S, following the generic T2SS nomenclature, see also text. The T2SS is composed of the cytoplasmic ATPase (E), the inner membrane complex (C, F, L and M), the pseudopilins (G-K) and outer membrane secretin (D). The presence of pilotin (S) in the assembled complex is uncertain. These components assemble into a complex spanning the inner and outer membranes of Gram-negative bacteria. Exoproteins, like the toxins CT and LT in *V. cholerae* and ETEC, are translocated by the T2SS in a folded state from the periplasm to the extracellular milieu.

1.2 T2SS PROTEINS

The T2SS contains 12-16 different proteins, most of them as multiple copies. The stoichiometry is still largely unknown. The system is highly dynamic, so the composition is likely to vary during the functioning of the machinery. Some known functions of the main T2SS proteins are as follows. GspE will be described separately and in more detail in the next section.

1.2.1 Proteins for the Outer Membrane Complex

GspA and GspB The genes *exeA* and *exeB* have been identified and demonstrated to be required for secretion of aerolysin in *A. hydrophila* [28], but they may be dispensable in *K. oxytoca* and *D. dadantii* secretion systems [29]. These genes have not yet been described in *V. cholerae*.

GspO is a prepilin peptidase, a polytopic, inner-membrane, aspartic-acid protease. GspO homolog in *P. aeruginosa* PilD, *A. hydrophila* tapD and *V. cholerae* VcpD have been shown to be functional in cleaving the leader peptide from the major and minor pseudopilins as well as the major and minor pilins [30-33]. In *P. aeruginosa*, there is only one prepilin peptidase gene *pilD* found, and it is used for both the T2SS and T4PS [34].

GspO is a bifunctional enzyme that uses S-adenosyl methionine for methylation of the amino group formed after cleaving the leader sequence. The only known structure of this protease family is the pre-flagellin peptidase FlaK from the archaella assembly system in *Methanococcus maripaludis* [35]. FlaK contains six transmembrane helices and a cytoplasmic domain located between helix 5 and 6. However, the FlaK structure is very different from the predicted structure of PilD, the GspO homolog in *P. aeruginosa*. Predicted PilD structure contains seven transmembrane helices and a cytoplasmic domain between helix 1 and 2. Interestingly, this cytoplasmic domain has a zinc-binding tetracysteine motif, a similar Zinc-binding motif in GspE. It has been shown that zinc is necessary for methyl-

transferase activity, but not for protease function [36]. FlaK lacks this tetracysteine motif. In addition, the conserved aspartic acids of FlaK are located at the cytoplasmic side of helix 1 and 4, but the predicted PilD structure places these aspartates at the cytoplasmic side of helix 3 and 6. The low sequence identity suggests these proteins might have different system specificities and possibly different structural arrangements [36].

GspD is an outer membrane protein and belongs to a family of proteins known as secretins. The secretin family consists of outer membrane proteins functioning as a channel in the T2SS and other assemblies like the T4PS, Type III secretion system and filamentous phage [37]. GspD oligomerizes into a pore used in the translocation of exoproteins across the outer membrane (Figure 1-1). It has been suggested that GspD may be required for the localization and assembly of the T2SS [38]. Most secretins require the use of the pilotin (GspS) for proper localization of GspD in the outer membrane [25, 39].

GspD chains consist of several domains (Figure 1-2): the Sec-dependent N-terminal signal peptide that targets the protein to the periplasm, followed by four N-terminal periplasmic domains N0-N3, a conserved C-terminal domain located in the outer membrane and an S-domain recognized by the pilotin [37]. A crystal structure of the N0-N1-N2 domains in complex with a nanobody revealed the folds of these domains and a close association of the N0 and N1 domains [40]. The C-terminal domain is predicted to have a β-barrel structure in the outer membrane that is thought to form the pore for exoproteins to pass [25, 41, 42]. Cryo-electron microscopy reconstruction of the *V. cholerae* GspD showed a cylindrical channel with 12-fold symmetry, 155 Å in diameter and length of 200 Å in length [43]. The cross-section of the secretin revealed a periplasmic vestibule composed of the N0, N1 and N2 domains with an opening of 75 Å that narrows to a constriction of 55 Å approximately one-thirds of the way from the opening. This constriction site is probably composed of the N3 domain of GspD. A major characteristic of this electron microscopy reconstruction is a gate which closes the channel. A small extracellular chamber is located

near the environment surrounding the bacterium, with an extracellular gate having in this state an opening of only 10 Å. The N-terminal periplasmic N0-N1 domains of GspD from ETEC have also been co-crystallized with the HR domain of GspC. The resultant structures showed that the N0-domain of GspD is associated with the HR-domain of GspC, an inner membrane platform protein.

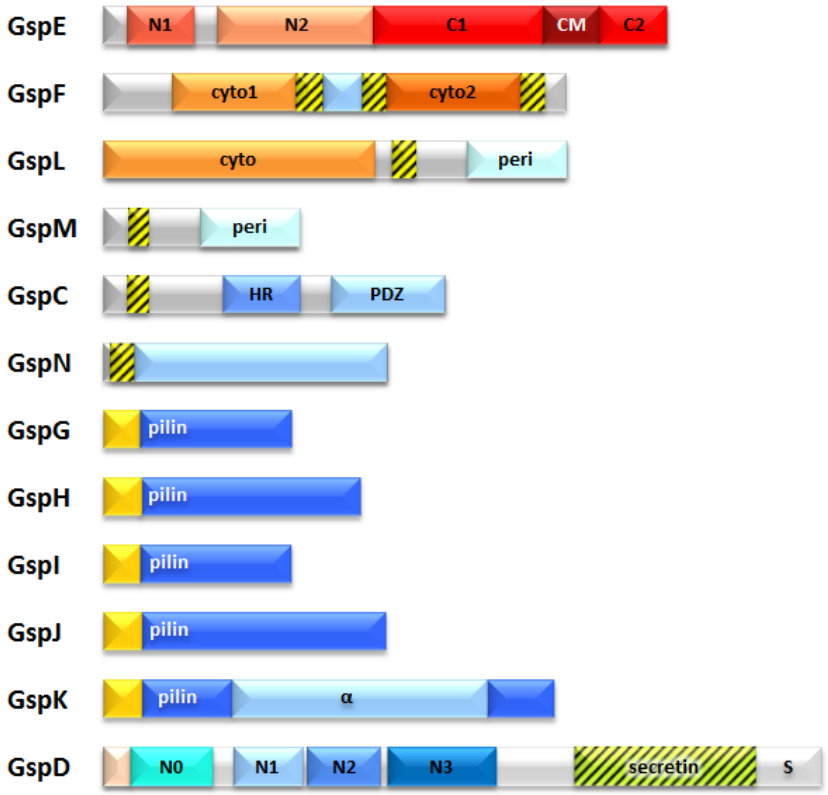


Figure 1-2: Domain bar diagram of the components of the T2SS. Domains in cytoplasm are in shades of red and orange, and blue for those in periplasm. Yellow color with black stripes indicates a predicted transmembrane domain.

GspS is a small lipoprotein, also called the pilotin. Pilotins play an important role in the assembly process of the outer membrane channel GspD, by binding to the C-terminal end of GspD [25, 44]. This interaction helps the proper localization of GspD in the outer membrane and also protects GspD from proteolysis. Although many of the secretins require a pilotin for proper localization, not all T2SS have a pilotin characterized. It has been shown that some T2SS secretins use a lipid anchor for the localization of the secretin [45]. Others, such as *A. hydrophila*, used GspA and GspB for secretin assembly [46]. This may explain why some GspD homologs require pilotins for proper localization while others do not. It is also not yet sure if the pilotin is permanently associated with the T2SS after secretin assembly. In addition, the T2SS in *Vibrio* species possesses a functionally similar but structurally unique pilotin AspS, compared with other bacteria.

1.2.2 The Pseudopilins

The periplasmic pseudopilus consists of five proteins GspG, H, I, J and K. These proteins are called pseudopilins because of sequence homology with pilins of the Type IV pilus in particular regarding the N-terminal residues [47, 48]. They are thought to form a pilus-like structure that can act like a piston, pushing exoproteins out of the secretin (Figure 1-1) [20, 22, 23, 25]. It has also been proposed that the pseudopilins may act as a rotary ratchet for protein secretion [24, 49]. The most abundant protein in the pseudopilus is GspG (the major pseudopilin), while GspH-K are termed as minor pseudopilins due to a lower abundance than GspG [33, 50]. GspG can form a visible pilus-like structure when overexpressed, a unique characteristic that is not observed for other pseudopilins [51].

All pseudopilins contain an N-terminal signal peptide that localizes the protein for translocation across the inner membrane by the Sec pathway [52], followed by a long N-terminal hydrophobic α-helix tail embedded in the inner membrane before pilus assembly, a variable region and a C-terminal β-sheet globular head. As mentioned above, a prepilin peptidase (called GspO) cleaves the signal peptide and methylates the newly formed N-terminal amino group, a step necessary for pilus assembly [53]. Structural studies of the T2SS pesudopilins and comparisons of the T4PS pilins suggest that the pseudopilus is a right-handed thin helical filament stabilized by calcium ions in GspG and GspK [54-56]. An invariant glutamic acid residue (present in all pseudopilins, except GspK) is hypothesized to be involved in an interaction with the N-terminal amino group of the next subunit in the structure [57-59]. A trimeric complex formed by GspK, GspI and GspJ in a right-handed helical arrangement is believed to be the tip of the pseudopilus, with GspK located at the top due to its large C-terminal alpha-helical domain that is absent in other pseudopilins [57, 58, 60]. The observations that ATPase GspE is required for polymerization of the pseudopilus and GspG has been shown to interact with GspL, provide a link between the ATPase and the pseudopilus, since GspE interacts with GspL [51, 61].

1.2.3 The Inner Membrane Platform Proteins

GspC is a bitopic inner membrane protein with a short cytoplasmic sequence and a periplasmic region consists of the HR domain and a PDZ domain that in some species is replaced by a coiled-coil domain. GspC has been shown to interact and protect the inner membrane GspL-GspM complex from degradation [62]. The complex of the HR domain and the periplasmic domains N0-N1-N2 of GspD was co-crystallized, showing how the HR domain and the N0 domain link the inner membrane platform to the outer membrane pore of the T2SS [63]. It has been proposed that either 6 or 12 copies of GspC are needed to maintain the interaction within the inner membrane platform and with the outer membrane complex [63]. The second periplasmic subdomain, the PDZ or coiled-coil domain, may be important for exoprotein selectivity in the T2SS [62, 64], based on the observation that the PDZ domain is involved in exoprotein recognition [64].

GspF is predicted to have three transmembrane domains with two cytoplasmic domains. The first cytoplasmic domain has crystallized with two subunits per asymmetric unit, though it is not sure if their association represents a dimer occurring in the assembled T2SS [65]. The first 55 residues were not observed in the crystal structure, but are thought to interact with other T2SS inner membrane proteins. The first and the second cytoplasmic domains share 28-39% sequence identity, and it is predicted that the second cytoplasmic domain has a similar structure as the first cytoplasmic domain. Co-purification and yeast two-hybrid experiments have shown that GspF can interact with a complex of GspE and GspL [52, 66, 67]. It is possible that the interaction requires both cytoplasmic domains of GspF.

GspL is a bitopic membrane protein with a large cytoplasmic domain and a small periplasmic domain [68, 69], and it may dimerize in the membrane [70]. The structures of the cytoplasmic domain of GspL (cyto-GspL) alone and in complex with the first N-terminal domain of GspE have been reported [71]. The periplasmic domain of GspL forms a dimer in

solution and also crystallized as a dimer with a rather extensive interface. Surprisingly, its closest structural homolog is the periplasmic domain of another T2SS inner membrane protein GspM, though there is no sequence homology and the dimer arrangements are entirely different [72]. In addition to interact with the cytoplasmic GspE, GspL has also been shown to interact with other inner membrane proteins GspC, GspF and GspM and pseudopilin GspG. However, their stoichiometry remains to be determined.

GspM is a bitopic membrane protein with a short cytoplasmic sequence and a small periplasimic domain. Evidence has been provided that GspM forms a stable complex with GspL through its transmembrane and periplasmic domains [73]. It is still unclear how these two proteins form a complex and what their ratio is in the GspM-GspL complex. As mentioned previously, the periplasmic domain of GspM is structurally homologous to the periplasmic domain of GspL [69]. In this structure, two peri-GspM domains form a hydrophobic cleft with a mysterious electron density at the interface. This cleft may be important for peptide or ligand binding.

1.2.4 The ATPase

GspE is the only protein in the T2SS being characterized as an ATPase. GspE is required for the function of the T2SS. It belongs to a large family of ATPases, known as the Type II/IV secretion ATPases [74]. They often form homo-hexameric rings and display enzymatic or conformational changes in response to the presence of phospholipids or detergents. They are cytosolic proteins, but they can be found in both the bacterial cytoplasm and membrane fractions. While members of this ATPase family can form hexamers, the truncated form of *V. cholerae* GspE, without its first N-terminal (N1) domain, was not crystallized as a hexamer, but helices with 6_1-symmetry [75]. Full-length *V. cholerae* GspE was purified as monomers with no detectable catalytic activity [27]. Surprisingly, a nearly undetectable amount of protein corresponding to the size of hexamers exhibited high ATPase activity [27].

Furthermore, ATPase activity can be stimulated by the cytoplasmic domain of the inner membrane protein GspL with the addition of acidic phospholipid, cardiolipin [76]. When GspE is purified from *V. cholerae*, it is found in both the soluble and membrane fractions [26]. In fact, GspE is a cytosolic protein that associates with the inner membrane platform of the T2SS through the interaction of its N1 domain with the cytoplasmic domain of GspL [71], and forms a heterodimer with a 1:1 ratio in the crystal structure (Figure 1-3).

Figure 1-3: Crystal structure of the complex formed by the cytoplasmic domain of GspL (orange) with the N1 domain of GspE (red) [71].

1.3 THE GspE FAMILY OF PROTEINS

The secretion ATPases also include the T4PS assembly ATPase (PilB/F) and retraction ATPase (PilT/U), the archaella (formerly known as archaeal flagella) assembly system (AAS) ATPase (FlaI), and one of the three type IV secretion system (T4SS) ATPases, VirB11. The

domain architecture of this often very distantly related set of proteins is shown in Figure 1-4. Alignments of these protein sequences indicate that the T2SS ATPases display the highest homology to the T4aPS (subclass of T4PS) assembly ATPases, and thus, these two groups of ATPases can be further viewed as a GspE/PilB subfamily [74]. The lowest homology of the T2SS GspE in this family is with the T4SS ATPase VirB11 family.

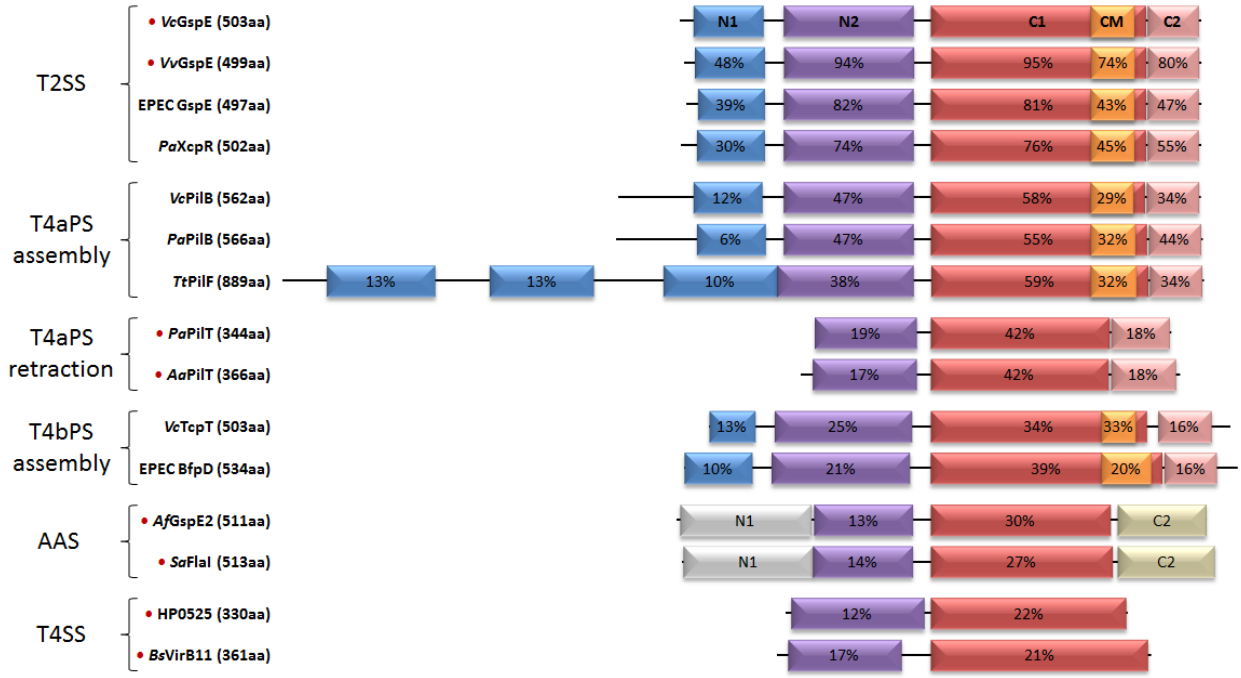


Figure 1-4: Domain bar diagram of representative proteins in the Type II/IV secretion ATPase family. The red circle (•) indicates proteins with crystal structures. The percentage sequence identity compared to *Vc*GspE is given for the domain homologous to GspE. N1 domain in blue, N2 domain in purple, C1 domain in red, CM domain in orange and C2 domain in pink. For domains that are not homologous in structures, the domain bar color is different from *Vc*GspE. *Vc: Vibrio cholerae, Vv: Vibrio vulnificus*, EPEC: enteropathogenic *Escherichia coli*, *Pa: Pseudomonas. aeruginosa, Tt: Thermus thermophiles, Aa: Aquifex aeolicus, Af: Archaeoglobus fulgidus, Sa: Sulfolobus acidocaldarius,* HP: *Helicobacter pylori, Bs: Brucella suis.*

1.3.1 Structural Characteristics

Members of this ATPase family share many structural characteristics (Figure 1-4). Each ATPase in the secretion ATPase family has at least two domains, known as the N-terminal domain (N2 in GspE) and the C-terminal domain (C1 in GspE), separated by a flexible linker. These two domains can form a double-stacked hexameric ring, and nucleotides are bound at the interface between the two domains. The C1 domain contains a RecA-like core with universal features shared among ATPases from other families. There is a highly conserved mononucleotide binding motif, known as the Walker A box. This motif is a sequence of GXXXXGKS/T, where X is any residue [77]. ATP-binding proteins also contain a feature called the Walker B box. In secretion ATPases, this motif is conserved, but has an atypical Walker-B sequence: RXXPDhhhhGEI/MRD, where h is any hydrophobic residue. These two ATP-binding motifs are highly conserved in other ATPases, such as the enormous AAA+ (ATPases associated with various cellular activities) family, which includes chaperones, proteases, and helicases. The C1 domain has the highest sequence homology among the secretion ATPases.

The 50-60 residues between the regions of the two Walker boxes are highly conserved within the entire secretion ATPase family. There is a short motif (TXEDPXE), known as the "aspartate box," in this family. This motif contains one absolutely conserved aspartic acid residue. This aspartic acid residue contributes to nucleotide binding by interacting with the Mg^{2+} ion bound to ATP. Mutational analysis of the GspE homolog, where the conserved aspartic acid residue in the aspartate box was mutated to an asparagine residue, shows that the aspartic acid residue is important for ATPase function [78].

Although the N2 domain has a lower level of sequence homology than the C1 domain, crystal structures of the secretion ATPase family show that they are structurally similar and residues from the N2 domains also directly contribute to ATPase activity [79-81]. In addition to the N2 and C1 domains, GspE chains also have two other domains: N1

and CM domains. The N1 domain is known for the interaction with the cytoplasmic domain of GspL, which may confer species specificity.

The feature that distinguishes the GspE/PilB subfamily from the other secretion ATPases is the presence, in almost all species, of the C-terminal metal-binding (CM) domain, also known as the Zinc-binding domain. This domain has two CXXC sequences, connected by ~30 residues, which are absolutely conserved in the GspE/PilB subfamily of the secretion ATPases, with one exception: the *Xanthomonas campestris* XpsE homolog. These cysteine pairs form a tetracysteine motif that is involved in coordinating a zinc metal [75]. The T4aPS assembly ATPases, but not the retraction ATPases, possess this tetracysteine motif as well as the T4bPS (subclass of T4PS) ATPases TcpT and BfpD. The high degree of conservation of this domain within this subfamily of secretion ATPases indicates that they are functionally important. These cysteine residues are essential to extracellular protein secretion [82], but the precise role of the Zn-binding domain remains a mystery [27].

1.3.2 Hexamers

Hexameric crystal structures are reported for many of the secretion ATPases, and these structures have revealed that nucleotide binding changes the conformation of the ATPases (Figure 1-5). The apo form of HP0525, a member of the T4SS VirB11 subfamily, shows a widening asymmetrical ring of the N-terminal domain relative to the C-terminal domain, and the ring closes with C_3 symmetry while bound to ADP and ATPγS [83, 84]. Archaeal secretion ATPases from *Archaeoglobus fulgidus* GspE2 (AfGspE2) and the homolog from *Sulfolobus acidocaldarius* FlaI (SaFlaI) exhibit large domain movements when nucleotide binds [81, 85]. Three different crystal structures of *Aquifex aeolicus* T4aPS retraction ATPase PilT (AaPilT) bound to ADP or ATP are reported, two with C_6 symmetry and the other with quasi-C_2 symmetry [79]. The structure of this quasi-C_2 symmetry showed approximately three conformational states within the hexamer, from a wide-open to

a closed configuration of N2-C1 domains. Two structures with C_2 symmetry from *P. aeruginosa* T4aPS PilT (PaPilT) were obtained in both the presence and absence of AMPPCP [86]. The C_2 hexamer with AMPPCP again show three different nucleotide-binding states bound to the N2 and C1 domains.

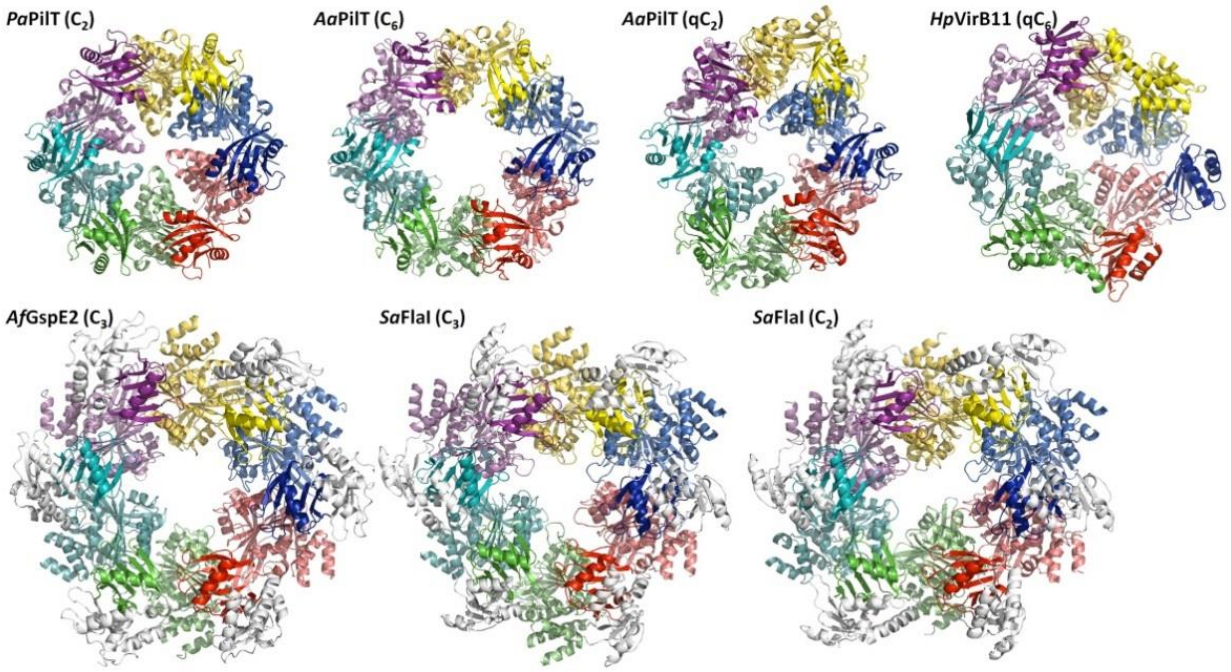


Figure 1-5: Hexameric crystal structures of members of the Type II/IV secretion ATPase family. N2 domains are colored in a darker shade of color than C-terminal domains of the same subunit, while all N1 domains are in grey. View is approximately along the symmetry axis. The symmetry of the structures is given in parenthesis, and q denotes quasi.

All of these structures revealed not only conformational differences among the same subunits, but also interactions with nucleotides contributed by two conserved arginine residues from the neighboring subunits, known as the arginine fingers. The arginine fingers, also a feature in the AAA+ ATPase family that forms hexamers [87], are required for ATP hydrolysis [80]. Therefore, a ring conformation may be required for performing ATP

hydrolysis. This may explain why secretion ATPases function as multimers, usually hexamers.

In contrast to the above hexameric secretion ATPase structures, the truncated form of *V. cholerae* GspE, lacking the N1 domain, did not crystallize as a hexamer. Both structures, in the presence and absence of AMPPNP, had similar conformations and revealed a helical filament with 6_1-symmetry (Figure 1-6). The helical filament is formed by the interface between the C1 domain of one subunit and the N2' domain of a neighboring subunit. This is the largest interface observed in these crystal structures. Surprisingly, the same C1•N2' interface is also observed in all of the above secretion ATPase structures. In fact, these hexameric secretion ATPase structures are formed by maintaining this essentially same interface between the adjacent subunits. This C1•N2' pair of domains of adjacent subunits has been called the "building block" [85] and "construction unit" of this family of proteins in Chapter 3.

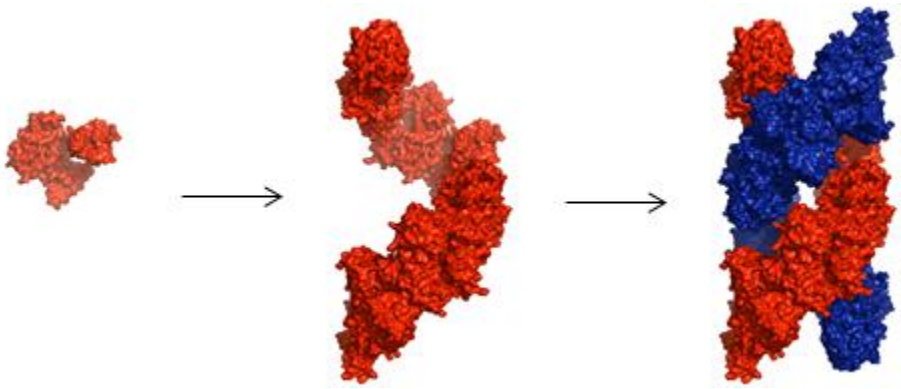


Figure 1-6: Quaternary structure of *V. cholerae* GspE lacking the N1 domain. One monomer in the crystal generates a helical filament (red), by a crystallographic 6_1 axis. An anti-parallel filament (blue) is generated by a crystallographic 2-fold axis perpendicular to the 6_1 axis.

All translocation systems require energy to actively transport proteins. Although energy is certainly required for the transport of secreted proteins across the outer membrane by the T2SS, it is also possible that the ATPase of the T2SS is not involved in energizing the secretion process, but involved in energy transduction to assist the assembly of the secretion apparatus [88, 89]. It has been reported that proton motive force may be the energy source for exoprotein translocation across the outer membrane, instead of the hydrolysis of ATP [90]. However, there is only limited work discussed in the literature regarding the evidence for a role of the proton motive force in T2SS secretion or assembly [24, 91]. How GspE uses ATP hydrolysis to power the T2SS is still unclear.

1.4 T4PS – the closest relative of the T2SS.

There are many similarities between the T4PS and T2SS. Several of the proteins in the T4PS are structurally similar to those of the T2SS. For example, both the pseudopilins of the T2SS and the pilins of the T4PS are synthesized with similar N-terminal signal sequences, which are recognized and cleaved by the GspO peptidase that is required for both protein secretion and pilus biogenesis. There are two subclasses of T4PS: T4aPS and T4bPS. The most common distinguishing characteristics separating T4a and T4b pili are the size of the major pilin subunit, the length of the prepilin peptidase cleavage sequence, and the operon organization [92].

A striking difference between the T2SS and T4PS protein components involves the secretion ATPases. As noted previously, the T4aPSs in general have more than one ATPase, unlike the T2SS. The T4aPS has an ATPase dedicated to assembly (called PilB/PilF – names and capital letter differ per species, the T4PS nomenclature is very confusing – see e.g. [93]. Hence we use PilB), which is the closest homolog to GspE. The assembly ATPase also has a CM domain, but no crystal structure is available so far. The cryoEM reconstruction of full-length *Thermus thermophiles* T4aPS PilF (TtPilF) shows a hexameric ring formation, as expected as a basic feature of the assembly ATPases. TtPilF has two additional N1 domains

(Figure 1-4), a unique feature that is not conserved in the PilB/F family. The three N1 domains form a disk and connect to the hexameric ring through a stalk region between the N0 and N1 domains. In addition to the assembly ATPase, there is in the T4aPS at least one ATPase dedicated to retraction (PilT), as discussed previously. Most systems have at least another accessory ATPase (PilU), while some also have a fourth, named PilT2. The role of these ATPases is less clear and seems to vary between systems [94]. So far, only one ATPase in a T4bPS is found, such as the assembly ATPases TcpT in *V. cholerae* and BfpD in enteropathogenic *E. coli* (EPEC). These ATPases also form hexameric rings.

Another feature that sets the T4PS apart from the T2SS concerns the proteins homologous to the T2SS inner membrane protein T4aPS, the structural equivalent of GspL is separated into two proteins: a cytosolic protein PilM and an inner membrane protein PilN (Figure 1-7). PilM is structurally equivalent to the cytoplasmic domain of GspL, while PilN is homologous to the transmembrane and periplasmic domains of GspL. While GspL has not been shown to be able to bind nucleotides, remarkably, PilM can bind ATP due to the presence of an additional domain that is missing in GspL. However, whether PilM has ATPase activity is not yet been detected [95]. T4bPS homologs of GspL appear to be strikingly different from the T4aPS PilM-PilN duo. The T4bPS BfpC protein is an inner membrane with both cytoplasmic and periplasmic domains like GspL. However, the cytoplasmic domain of BfpC is substantially smaller than GspL.

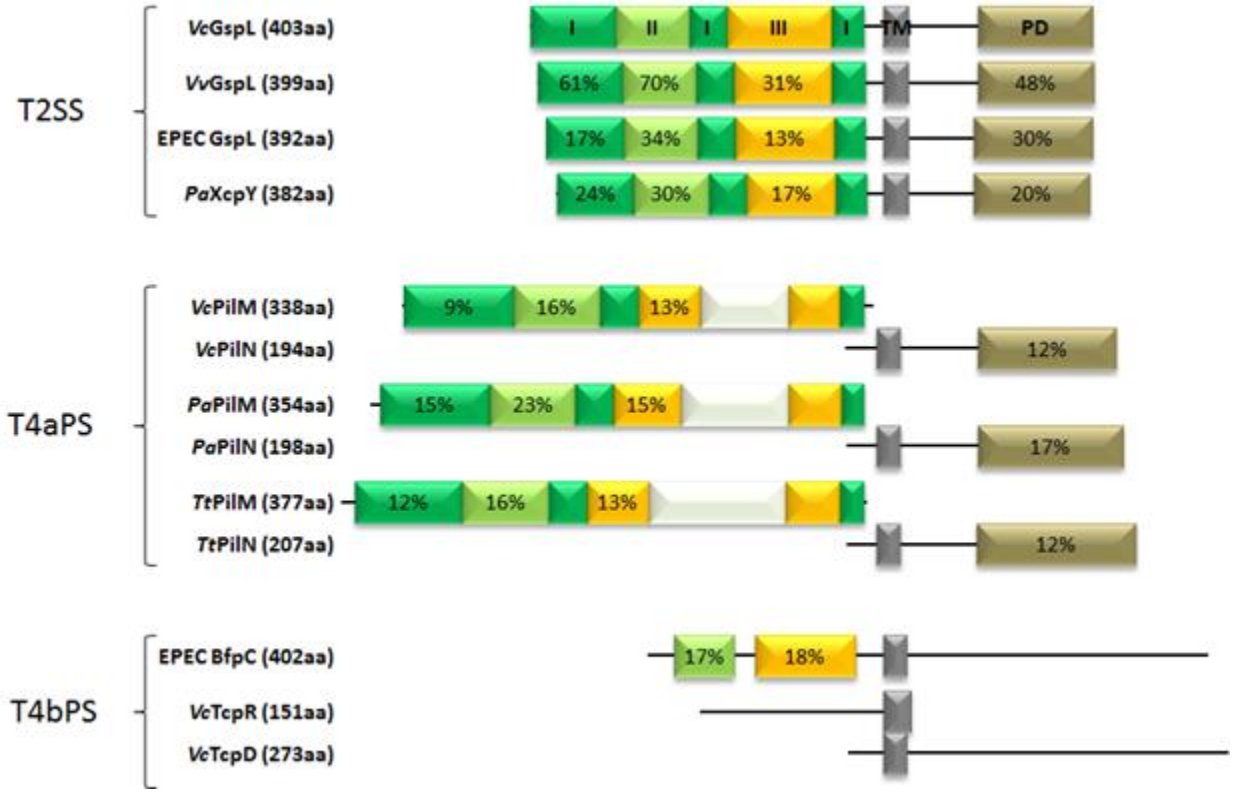


Figure 1-7: Domain bar diagram of proteins or protein duos homologous to GspL. The percentage sequence identity of each domain compared to VcGspL is given. Domain I in green, domain II in lime, domain III in yellow, transmembrane region in grey and periplasmic domain in brown. White bar represents a domain missing in GspL. Lines indicate regions without crystal structures or any structural prediction in the literature. *Vc: Vibrio cholerae, Vv: Vibrio vulnificus*, EPEC: enteropathogenic *Escherichia coli*, *Pa: Pseudomonas. aeruginosa, Tt: Thermus thermophiles.*

1.5 SELECTED HUMAN PATHOGENS CONTAINING ONE OR MORE T2SS

Vibrio cholerae is a Gram-negative rod-shaped bacterium that uses the T2SS to secrete several proteins, including cholera toxin (CT), proteases, lipase, neuraminidase and chitinase [7]. CT is the major virulence factor of the diarrheal disease cholera, a disease causing profuse watery diarrhea by inducing efflux of water and electrolytes in the intestinal epithelial cells. About 5 million people are infected with cholera diseases every year [96]. It can cause death if patients are not treated with oral rehydration salt. CT is an AB_5 multimeric protein consisting of a catalytic subunit A and a pentameric-ring subunit B. The B pentamer is responsible for the binding and secretion of the toxin [3]. *V. cholerae* contains at least four proteins in addition to GspE that belong to the secretion ATPase family: TcpT, PilB, PilT and PilU, all of which are involved in the biogenesis and function of the T4PS found on the bacterial cell surface. Sequence comparisons of the four proteins show that PilT and PilU are very similar to each other, while GspE, PilB and TcpT are more closely related, although all four are similar to some degree.

Enterotoxigenic *E coli* (ETEC) is the causative agent of children's and traveler's diarrhea, which are cholera-like diseases. It is estimated that at least 400,000 ETEC-related deaths in children under the age of five occur each year according to WHO. Several secreted proteins have been reported to influence ETEC pathogenesis. The primary virulence factor is the heat-labile enterotoxin (LT), which causes the diarrhea associated with ETEC infection. LT is structurally and functionally similar to CT and shares ~80% sequence identity for both the A and B chain.

Pseudomonas aeruginosa is a Gram-negative opportunistic pathogen of particular importance for cystic fibrosis patients. This pathogen secretes multiple virulence factors into the extracellular milieu. Many of these virulence factors, such as alkaline phosphatase, exotoxin A, phospholipase C, lipase and elastase are secreted by the type II secretion

system. There are two T2SS identified in *P. aeruginosa*: the Hxc T2SS, which secretes an alkaline phosphatase under phosphate-limited conditions, and the predominant Xcp T2SS encoded by the *xcp* genes (*xcpR-xcpZ* and *pilD)*, which is responsible for all other T2SS dependent substrates [97].

1.6 OVERVIEW

The main theme of this thesis is to determine structures of the GspE in an attempt to confirm the hypothesis that GspE form multimers, more likely hexamers within the T2SS assembly. GspE has been studied for decades, but there was at the start of this study only a truncated GspE structure available, which gives only partial picture without the context of the whole T2SS architecture. Chapter 2 describes the full-length structure of *Vibrio vulnificus* GspE in complex with the cytoplasmic domain of GspL (cyto-GspL). This complex structure shows that GspE subunits are very flexible and able to adopt a conformation entirely different from the *V. cholerae* helical GspE structure. In contrast, the C1•N2' interface observed in the helical structure and other secretion ATPase structures is again maintained in this crystal. The interaction between *V. vulnificus* N1-GspE and cyto-GspL is very similar to the reported *V. cholera*e structures. Remarkably, cyto-GspL•cyto-GspL interfaces are nearly identical to those observed in the two *V. cholera*e cyto-GspL•cyto-GspL and N1-GspE•cyto-GspL structures. GspL has been implicated in inducing GspE multimerization, possibly hexamerization. However, the structure of full-length GspE in complex with cyto-GspL does not show any evidence of GspE multimers.

In order to accomplish the goal of obtaining GspE multimers, an assistant hexamer Hcp1 was used as a template for GspE hexamerization by fusing GspE to Hcp1. The linking strategy had been used in other studies of protein-protein interactions [98], and in this work, it proves to be successful in not only preparing GspE hexamers in solution with high activity but also in structures. Two very different hexamers of *V. cholerae* GspE without N1 domain are crystallized, one with quasi-C_6 symmetry and another with C_2 symmetry. When

comparing the two GspE hexamers with other secretion ATPase structures, they display two entirely different ring structures. The linkers between the N2 and C1 domains allow various conformations of the same subunit, while maintaining the same rigid C1•N2' construction unit as observed in the reported secretion ATPases structures. This architecture may shed light on how secretion ATPases multimerize and couple ATP binding or hydrolysis, which result in conformational changes within the ATPase multimers, to the T2SS mechanism of action. This work has shown that GspE can form hexamers, However, nothing is exactly certain about the GspE oligomerization states in the T2SS assembly. Much remains to be elucidated regarding its full-length conformation and oligomeric state within the T2SS apparatus.

1.7 REFERENCE

1. Kudva, R., et al., *Protein translocation across the inner membrane of Gram-negative bacteria: the Sec and Tat dependent protein transport pathways.* Research in Microbiology, 2013. 164(6): p. 505-34.
2. Voulhoux, R., et al., *Involvement of the twin-arginine translocation system in protein secretion via the type II pathway.* The EMBO journal, 2001. 20(23): p. 6735-41.
3. Hirst, T.R., et al., *Mechanism of toxin secretion by Vibrio cholerae investigated in strains harboring plasmids that encode heat-labile enterotoxins of Escherichia coli.* Proceedings of the National Academy of Sciences of the United States of America, 1984. 81(24): p. 7752-6.
4. Collet, J.F. and J.C. Bardwell, *Oxidative protein folding in bacteria.* Molecular microbiology, 2002. 44(1): p. 1-8.
5. Hirst, T.R. and J. Holmgren, *Conformation of protein secreted across bacterial outer membranes: a study of enterotoxin translocation from Vibrio cholerae.* Proceedings of the National Academy of Sciences of the United States of America, 1987. 84(21): p. 7418-22.
6. d'Enfert, C., A. Ryter, and A.P. Pugsley, *Cloning and expression in Escherichia coli of the Klebsiella pneumoniae genes for production, surface localization and secretion of the lipoprotein pullulanase.* The EMBO journal, 1987. 6(11): p. 3531-8.
7. Sikora, A.E., et al., *Proteomic analysis of the Vibrio cholerae type II secretome reveals new proteins, including three related serine proteases.* The Journal of biological chemistry, 2011. 286(19): p. 16555-66.
8. Lazdunski, A., et al., *Secretion of extracellular proteins by Pseudomonas aeruginosa.* Biochimie, 1990. 72(2-3): p. 147-56.
9. Murata, H., et al., *Characterization of transposon insertion out- mutants of Erwinia carotovora subsp. carotovora defective in enzyme export and of a DNA segment that complements out mutations in E. carotovora subsp. carotovora, E. carotovora subsp. atroseptica, and Erwinia chrysanthemi.* Journal of bacteriology, 1990. 172(6): p. 2970-8.

10. Howard, S.P. and J.T. Buckley, *Protein export by a gram-negative bacterium: production of aerolysin by Aeromonas hydrophila.* Journal of bacteriology, 1985. 161(3): p. 1118-24.
11. Dums, F., J.M. Dow, and M.J. Daniels, *Structural characterization of protein secretion genes of the bacterial phytopathogen Xanthomonas campestris pathovar campestris: relatedness to secretion systems of other gram-negative bacteria.* Molecular & general genetics : MGG, 1991. 229(3): p. 357-64.
12. Py, B., et al., *Cellulase EGZ of Erwinia chrysanthemi: structural organization and importance of His98 and Glu133 residues for catalysis.* Protein engineering, 1991. 4(3): p. 325-33.
13. de Groot, A., A. Filloux, and J. Tommassen, *Conservation of xcp genes, involved in the two-step protein secretion process, in different Pseudomonas species and other gram-negative bacteria.* Molecular & general genetics : MGG, 1991. 229(2): p. 278-84.
14. Lindeberg, M., G.P. Salmond, and A. Collmer, *Complementation of deletion mutations in a cloned functional cluster of Erwinia chrysanthemi out genes with Erwinia carotovora out homologues reveals OutC and OutD as candidate gatekeepers of species-specific secretion of proteins via the type II pathway.* Molecular microbiology, 1996. 20(1): p. 175-90.
15. Lee, V.T. and O. Schneewind, *Protein secretion and the pathogenesis of bacterial infections.* Genes & development, 2001. 15(14): p. 1725-52.
16. Neill, R.J., B.E. Ivins, and R.K. Holmes, *Synthesis and secretion of the plasmid-coded heat-labile enterotoxin of Escherichia coli in Vibrio cholerae.* Science, 1983. 221(4607): p. 289-91.
17. Connell, T.D., et al., *Initial studies of the structural signal for extracellular transport of cholera toxin and other proteins recognized by Vibrio cholerae.* Infection and immunity, 1995. 63(10): p. 4091-8.
18. Pugsley, A.P., et al., *Recent progress and future directions in studies of the main terminal branch of the general secretory pathway in Gram-negative bacteria--a review.* Gene, 1997. 192(1): p. 13-9.
19. Cianciotto, N.P., *Type II Secretion and Legionella Virulence.* Current topics in microbiology and immunology, 2013. 376: p. 81-102.
20. Douzi, B., A. Filloux, and R. Voulhoux, *On the path to uncover the bacterial type II secretion system.* Philosophical Transactions of the Royal Society B-Biological Sciences, 2012. 367(1592): p. 1059-1072.
21. Howard, S.P., *Assembly of the type II secretion system.* Research in Microbiology, 2013. 164(6): p. 535-544.
22. Korotkov, K.V., M. Sandkvist, and W.G. Hol, *The type II secretion system: biogenesis, molecular architecture and mechanism.* Nature reviews. Microbiology, 2012. 10(5): p. 336-51.
23. McLaughlin, L.S., R.J.F. Haft, and K.T. Forest, *Structural insights into the Type II secretion nanomachine.* Current Opinion in Structural Biology, 2012. 22(2): p. 208-216.
24. Nivaskumar, M. and O. Francetic, *Type II secretion system: A magic beanstalk or a protein escalator.* Biochim Biophys Acta, 2014.
25. Shevchik, V.E., J. Robert-Baudouy, and G. Condemine, *Specific interaction between OutD, an Erwinia chrysanthemi outer membrane protein of the general secretory pathway, and secreted proteins.* The EMBO journal, 1997. 16(11): p. 3007-16.
26. Sandkvist, M., et al., *Interaction between the autokinase EpsE and EpsL in the cytoplasmic membrane is required for extracellular secretion in Vibrio cholerae.* The EMBO journal, 1995. 14(8): p. 1664-73.
27. Camberg, J.L. and M. Sandkvist, *Molecular analysis of the Vibrio cholerae type II secretion ATPase EpsE.* Journal of bacteriology, 2005. 187(1): p. 249-56.

28. Howard, S.P., et al., *A TonB-like protein and a novel membrane protein containing an ATP-binding cassette function together in exotoxin secretion.* Molecular microbiology, 1996. 22(4): p. 595-604.
29. Condemine, G., et al., *Some of the out genes involved in the secretion of pectate lyases in Erwinia chrysanthemi are regulated by kdgR.* Molecular microbiology, 1992. 6(21): p. 3199-211.
30. Strom, M.S., D.N. Nunn, and S. Lory, *A single bifunctional enzyme, PilD, catalyzes cleavage and N-methylation of proteins belonging to the type IV pilin family.* Proceedings of the National Academy of Sciences of the United States of America, 1993. 90(6): p. 2404-8.
31. Pepe, C.M., M.W. Eklund, and M.S. Strom, *Cloning of an Aeromonas hydrophila type IV pilus biogenesis gene cluster: complementation of pilus assembly functions and characterization of a type IV leader peptidase/N-methyltransferase required for extracellular protein secretion.* Molecular microbiology, 1996. 19(4): p. 857-69.
32. Marsh, J.W. and R.K. Taylor, *Identification of the Vibrio cholerae type 4 prepilin peptidase required for cholera toxin secretion and pilus formation.* Molecular microbiology, 1998. 29(6): p. 1481-92.
33. Nunn, D.N. and S. Lory, *Cleavage, methylation, and localization of the Pseudomonas aeruginosa export proteins XcpT, -U, -V, and -W.* Journal of bacteriology, 1993. 175(14): p. 4375-82.
34. Strom, M.S., D. Nunn, and S. Lory, *Multiple roles of the pilus biogenesis protein pilD: involvement of pilD in excretion of enzymes from Pseudomonas aeruginosa.* Journal of bacteriology, 1991. 173(3): p. 1175-80.
35. Hu, J., et al., *The crystal structure of GXGD membrane protease FlaK.* Nature, 2011. 475(7357): p. 528-31.
36. Aly, K.A., et al., *Cell-free production of integral membrane aspartic acid proteases reveals zinc-dependent methyltransferase activity of the Pseudomonas aeruginosa prepilin peptidase PilD.* MicrobiologyOpen, 2013. 2(1): p. 94-104.
37. Korotkov, K.V., T. Gonen, and W.G. Hol, *Secretins: dynamic channels for protein transport across membranes.* Trends in biochemical sciences, 2011. 36(8): p. 433-43.
38. Lybarger, S.R., et al., *Docking and assembly of the type II secretion complex of Vibrio cholerae.* Journal of bacteriology, 2009. 191(9): p. 3149-61.
39. Hardie, K.R., S. Lory, and A.P. Pugsley, *Insertion of an outer membrane protein in Escherichia coli requires a chaperone-like protein.* The EMBO journal, 1996. 15(5): p. 978-88.
40. Korotkov, K.V., et al., *Crystal structure of the N-terminal domain of the secretin GspD from ETEC determined with the assistance of a nanobody.* Structure, 2009. 17(2): p. 255-65.
41. Chami, M., et al., *Structural insights into the secretin PulD and its trypsin-resistant core.* The Journal of biological chemistry, 2005. 280(45): p. 37732-41.
42. Guilvout, I., et al., *Genetic dissection of the outer membrane secretin PulD: are there distinct domains for multimerization and secretion specificity?* Journal of bacteriology, 1999. 181(23): p. 7212-20.
43. Reichow, S.L., et al., *Structure of the cholera toxin secretion channel in its closed state.* Nature structural & molecular biology, 2010. 17(10): p. 1226-32.
44. Hardie, K.R., et al., *The secretin-specific, chaperone-like protein of the general secretory pathway: separation of proteolytic protection and piloting functions.* Molecular microbiology, 1996. 22(5): p. 967-76.
45. Viarre, V., et al., *HxcQ liposecretin is self-piloted to the outer membrane by its N-terminal lipid anchor.* The Journal of biological chemistry, 2009. 284(49): p. 33815-23.

46. Strozen, T.G., et al., *Involvement of the GspAB complex in assembly of the type II secretion system secretin of Aeromonas and Vibrio species.* Journal of bacteriology, 2011. 193(9): p. 2322-31.
47. Nunn, D.N. and S. Lory, *Components of the protein-excretion apparatus of Pseudomonas aeruginosa are processed by the type IV prepilin peptidase.* Proceedings of the National Academy of Sciences of the United States of America, 1992. 89(1): p. 47-51.
48. Pugsley, A.P., *The complete general secretory pathway in gram-negative bacteria.* Microbiological reviews, 1993. 57(1): p. 50-108.
49. Nivaskumar, M., et al., *Distinct Docking and Stabilization Steps of the Pseudopilus Conformational Transition Path Suggest Rotational Assembly of Type IV Pilus-like Fibers.* Structure, 2014. 22(5): p. 685-96.
50. Kohler, R., et al., *Structure and assembly of the pseudopilin PulG.* Mol Microbiol, 2004. 54(3): p. 647-64.
51. Durand, E., et al., *XcpX controls biogenesis of the Pseudomonas aeruginosa XcpT-containing pseudopilus.* The Journal of biological chemistry, 2005. 280(36): p. 31378-89.
52. Arts, J., et al., *Interaction domains in the Pseudomonas aeruginosa type II secretory apparatus component XcpS (GspF).* Microbiology, 2007. 153(Pt 5): p. 1582-92.
53. LaPointe, C.F. and R.K. Taylor, *The type 4 prepilin peptidases comprise a novel family of aspartic acid proteases.* The Journal of biological chemistry, 2000. 275(2): p. 1502-10.
54. Yanez, M.E., et al., *The crystal structure of a binary complex of two pseudopilins: EpsI and EpsJ from the type 2 secretion system of Vibrio vulnificus.* Journal of molecular biology, 2008. 375(2): p. 471-86.
55. Lam, A.Y., et al., *Nanobody-aided structure determination of the EpsI:EpsJ pseudopilin heterodimer from Vibrio vulnificus.* J Struct Biol, 2009. 166(1): p. 8-15.
56. Korotkov, K.V., et al., *Calcium is essential for the major pseudopilin in the type 2 secretion system.* J Biol Chem, 2009. 284(38): p. 25466-70.
57. Korotkov, K.V. and W.G. Hol, *Structure of the GspK-GspI-GspJ complex from the enterotoxigenic Escherichia coli type 2 secretion system.* Nature structural & molecular biology, 2008. 15(5): p. 462-8.
58. Campos, M., et al., *Detailed structural and assembly model of the type II secretion pilus from sparse data.* Proceedings of the National Academy of Sciences of the United States of America, 2010. 107(29): p. 13081-6.
59. Craig, L., et al., *Type IV pilus structure by cryo-electron microscopy and crystallography: implications for pilus assembly and functions.* Molecular cell, 2006. 23(5): p. 651-62.
60. Douzi, B., et al., *The XcpV/GspI pseudopilin has a central role in the assembly of a quaternary complex within the T2SS pseudopilus.* The Journal of biological chemistry, 2009. 284(50): p. 34580-9.
61. Gray, M.D., et al., *In vivo cross-linking of EpsG to EpsL suggests a role for EpsL as an ATPase-pseudopilin coupling protein in the Type II secretion system of Vibrio cholerae.* Molecular microbiology, 2011. 79(3): p. 786-98.
62. Gerard-Vincent, M., et al., *Identification of XcpP domains that confer functionality and specificity to the Pseudomonas aeruginosa type II secretion apparatus.* Molecular microbiology, 2002. 44(6): p. 1651-65.
63. Korotkov, K.V., et al., *Structural and functional studies on the interaction of GspC and GspD in the type II secretion system.* PLoS Pathog, 2011. 7(9): p. e1002228.
64. Bouley, J., G. Condemine, and V.E. Shevchik, *The PDZ domain of OutC and the N-terminal region of OutD determine the secretion specificity of the type II out pathway of Erwinia chrysanthemi.* Journal of molecular biology, 2001. 308(2): p. 205-19.

65. Abendroth, J., et al., *The three-dimensional structure of the cytoplasmic domains of EpsF from the type 2 secretion system of Vibrio cholerae.* J Struct Biol, 2009. 166(3): p. 303-15.
66. Py, B., L. Loiseau, and F. Barras, *An inner membrane platform in the type II secretion machinery of Gram-negative bacteria.* EMBO Rep, 2001. 2(3): p. 244-8.
67. Robert, V., A. Filloux, and G.P. Michel, *Subcomplexes from the Xcp secretion system of Pseudomonas aeruginosa.* FEMS microbiology letters, 2005. 252(1): p. 43-50.
68. Abendroth, J., et al., *The structure of the cytoplasmic domain of EpsL, an inner membrane component of the type II secretion system of Vibrio cholerae: an unusual member of the actin-like ATPase superfamily.* Journal of molecular biology, 2004. 344(3): p. 619-33.
69. Abendroth, J., A.C. Kreger, and W.G. Hol, *The dimer formed by the periplasmic domain of EpsL from the Type 2 Secretion System of Vibrio parahaemolyticus.* Journal of structural biology, 2009. 168(2): p. 313-22.
70. Sandkvist, M., et al., *Direct interaction of the EpsL and EpsM proteins of the general secretion apparatus in Vibrio cholerae.* J Bacteriol, 1999. 181(10): p. 3129-35.
71. Abendroth, J., et al., *The X-ray structure of the type II secretion system complex formed by the N-terminal domain of EpsE and the cytoplasmic domain of EpsL of Vibrio cholerae.* Journal of molecular biology, 2005. 348(4): p. 845-55.
72. Abendroth, J., et al., *The crystal structure of the periplasmic domain of the type II secretion system protein EpsM from Vibrio cholerae: the simplest version of the ferredoxin fold.* Journal of molecular biology, 2004. 338(3): p. 585-96.
73. Sandkvist, M., et al., *Two regions of EpsL involved in species-specific protein-protein interactions with EpsE and EpsM of the general secretion pathway in Vibrio cholerae.* Journal of bacteriology, 2000. 182(3): p. 742-8.
74. Planet, P.J., et al., *Phylogeny of genes for secretion NTPases: identification of the widespread tadA subfamily and development of a diagnostic key for gene classification.* Proceedings of the National Academy of Sciences of the United States of America, 2001. 98(5): p. 2503-8.
75. Robien, M.A., et al., *Crystal structure of the extracellular protein secretion NTPase EpsE of Vibrio cholerae.* Journal of molecular biology, 2003. 333(3): p. 657-74.
76. Camberg, J.L., et al., *Synergistic stimulation of EpsE ATP hydrolysis by EpsL and acidic phospholipids.* The EMBO journal, 2007. 26(1): p. 19-27.
77. Rivas, S., et al., *TrwD, a protein encoded by the IncW plasmid R388, displays an ATP hydrolase activity essential for bacterial conjugation.* The Journal of biological chemistry, 1997. 272(41): p. 25583-90.
78. Possot, O. and A.P. Pugsley, *Molecular characterization of PulE, a protein required for pullulanase secretion.* Molecular microbiology, 1994. 12(2): p. 287-99.
79. Satyshur, K.A., et al., *Crystal structures of the pilus retraction motor PilT suggest large domain movements and subunit cooperation drive motility.* Structure, 2007. 15(3): p. 363-76.
80. Patrick, M., et al., *Oligomerization of EpsE coordinates residues from multiple subunits to facilitate ATPase activity.* The Journal of biological chemistry, 2011. 286(12): p. 10378-86.
81. Yamagata, A. and J.A. Tainer, *Hexameric structures of the archaeal secretion ATPase GspE and implications for a universal secretion mechanism.* The EMBO journal, 2007. 26(3): p. 878-90.
82. Possot, O.M. and A.P. Pugsley, *The conserved tetracysteine motif in the general secretory pathway component PulE is required for efficient pullulanase secretion.* Gene, 1997. 192(1): p. 45-50.
83. Savvides, S.N., et al., *VirB11 ATPases are dynamic hexameric assemblies: new insights into bacterial type IV secretion.* The EMBO journal, 2003. 22(9): p. 1969-80.

84. Yeo, H.J., et al., *Crystal structure of the hexameric traffic ATPase of the Helicobacter pylori type IV secretion system.* Molecular Cell, 2000. 6(6): p. 1461-72.
85. Reindl, S., et al., *Insights into FlaI Functions in Archaeal Motor Assembly and Motility from Structures, Conformations, and Genetics.* Molecular Cell, 2013. 49(6): p. 1069-82.
86. Misic, A.M., K.A. Satyshur, and K.T. Forest, *P. aeruginosa PilT structures with and without nucleotide reveal a dynamic type IV pilus retraction motor.* Journal of Molecular Biology, 2010. 400(5): p. 1011-21.
87. Ye, J., et al., *RecA-like motor ATPases--lessons from structures.* Biochimica et biophysica acta, 2004. 1659(1): p. 1-18.
88. Possot, O., et al., *Pullulanase secretion in Escherichia coli K-12 requires a cytoplasmic protein and a putative polytopic cytoplasmic membrane protein.* Molecular microbiology, 1992. 6(1): p. 95-105.
89. Possot, O.M., L. Letellier, and A.P. Pugsley, *Energy requirement for pullulanase secretion by the main terminal branch of the general secretory pathway.* Molecular microbiology, 1997. 24(3): p. 457-64.
90. Wong, K.R. and J.T. Buckley, *Proton motive force involved in protein transport across the outer membrane of Aeromonas salmonicida.* Science, 1989. 246(4930): p. 654-6.
91. Filloux, A., *The underlying mechanisms of type II protein secretion.* Biochimica et biophysica acta, 2004. 1694(1-3): p. 163-79.
92. Craig, L. and J. Li, *Type IV pilli: paradoxes in form and function.* Current Opinion in Structural Biology, 2008. 18(2): p. 267-277.
93. Ayers, M., P.L. Howell, and L.L. Burrows, *Architecture of the type II secretion and type IV pilus machineries.* Future microbiology, 2010. 5(8): p. 1203-18.
94. Brown, D.R., et al., *Systematic functional analysis reveals that a set of seven genes is involved in fine-tuning of the multiple functions mediated by type IV pili in Neisseria meningitidis.* Infection and immunity, 2010. 78(7): p. 3053-63.
95. Karuppiah, V. and J.P. Derrick, *Structure of the PilM-PilN inner membrane type IV pilus biogenesis complex from Thermus thermophilus.* J Biol Chem, 2011. 286(27): p. 24434-42.
96. WHO. *World Health Organization, Media Centre, Cholera, Fact Sheet No. 107* February 2014; Available from: http://www.who.int/mediacentre/factsheets/fs107/en/.
97. Ball, G., et al., *A novel type II secretion system in Pseudomonas aeruginosa.* Molecular microbiology, 2002. 43(2): p. 475-85.
98. Reddy Chichili, V.P., V. Kumar, and J. Sivaraman, *Linkers in the structural biology of protein–protein interactions.* Protein Science, 2013. 22(2): p. 153-167.

Chapter 2

Crystal Structure of the Full-Length ATPase GspE from the *Vibrio vulnificus* Type II Secretion System in Complex with the Cytoplasmic Domain of GspL

This chapter is submitted for publication.

SUMMARY

The type II secretion system (T2SS) is present in many Gram-negative bacteria and is responsible for secreting a large number of folded proteins, including major virulence factors, across the outer membrane. The T2SS consists of 11-15 different proteins most of which are present in multiple copies in the assembled secretion machinery. The ATPase GspE, essential for the functioning of the T2SS, contains three domains (N1E, N2E and CTE) of which the N1E domain is associated with the cytoplasmic domain of the inner membrane protein GspL.

Here we describe and analyze the structure of the GspE•cyto-GspL complex from *Vibrio vulnificus* in the presence of an ATP analog, AMPPNP. There are three such ~83 kDa complexes per asymmetric unit with essentially the same structure. The N2E and CTE domains of a single *V. vulnificus* GspE subunit adopt a mutual orientation that has not been seen before in any of the previous GspE structures, neither in structures of related ATPases from other secretion systems. This underlines the tremendous conformational flexibility of the T2SS secretion ATPase.

Cyto-GspL interacts not only with the N1E domain, but also with the CTE domain and is even in contact with AMPPNP. Moreover, the cyto-GspL domains engage in two types of mutual interactions, resulting in two essentially identical, but crystallographically independent, "cyto-GspL rods" that run throughout the crystal. Very similar rods are present in previous crystals of cyto-GspL and of the N1E•cyto-GspL complex. This arrangement,

now seen four times in three entirely different crystal forms, involves contacts between highly conserved residues suggesting a role in the biogenesis or the secretion mechanism or both of the T2SS.

2.1 INTRODUCTION

Secretion of proteins into the extra-cellular milieu is important for many pathogenic and non-pathogenic bacteria, which have developed an impressive variety of often complex multi-protein assemblies to perform this task. One of these protein secretion machineries is the sophisticated type II secretion system (T2SS) that spans the inner and outer membranes from many Gram-negative bacteria [1]. The T2SS is highly relevant for the pathogenicity of several major pathogens since it translocates major virulence factors in a folded form from the periplasm to the extracellular milieu. Examples of bacterial human pathogens where the T2SS plays an important role include:

(i) *Vibrio cholerae,* where the T2SS secretes the heterohexameric AB_5 cholera toxin (CT) and ~20 other proteins [2, 3]. *V. cholerae* is responsible for estimated 100,000-120,000 deaths per year, mainly in low-income countries and disaster areas (http://www.who.int/mediacentre/factsheets/fs107/en/);

(ii) Enterotoxigenic *E. coli* (ETEC), where the T2SS translocates heat-labile enterotoxin (LT), a close structural and functional homolog of CT [4]. ETEC are an extremely important cause of diarrhea in the developing world [5, 6], and also are the most common cause of travelers' diarrhea [7];

(iii) Enterohaemorrhagic *E. coli* (EHEC), which can cause severe foodborne disease, and even life-threatening renal failure in children and elderly. The T2SS deletion mutant of EHEC shows defects in colonization. In addition, the zinc metalloprotease StcE and the metal binding protein YodA, which are crucial for EHEC adherence to host cells, are secreted by the T2SS encoded on plasmid pO158 [8-10];

(iv) Enteropathogenic *E. coli* (EPEC), is one of the most important pathogens affecting children worldwide with the infection resulting in persistent diarrhea [11]. The T2SS is required for EPEC virulence [12];

(v) *Pseudomonas aeruginosa*, an opportunistic pathogen of major importance in cystic fibrosis patients, contains two distinct T2SS machineries [13, 14];

(vi) The intracellular pathogen *Legionella pneumophila*, the causative agent of Legionnaire's disease, is dependent on the T2SS [15, 16].

The T2SS is made up from ~11-15 proteins, most of these present in multiple copies in the assembled secretion complex. As a result of numerous biochemical and structural studies, and from analogies to related systems, a generally accepted picture has emerged with the T2SS thought to consist of three subassemblies: the Inner Membrane Platform, the dynamic Pseudopilus, and the channel-forming Outer Membrane Complex [1, 17-23]. The Inner Membrane Platform [24] is composed of the T2SS membrane proteins GspC, GspF, GspL, GspM and, in some species, GspN. The ATPase GspE resides in the cytoplasm interacting with the cytoplasmic domain of GspL [25-27] and with GspF [24, 28]. The stoichiometry of the Inner Membrane Complex, the nexus of the T2SS since it interacts with all other subassemblies, is still a mystery. The Pseudopilus contains five different pseudopilins: GspK, GspI, GspJ, GspH and GspG. The tip is formed by a GspK•GspI•GspJ heterotrimer [29], most likely linked by one or a few GspH subunits [30, 31] to a helical filament made up of multiple copies of a calcium-requiring GspG [31-35]. The Outer Membrane Complex is composed of a dodecamer of GspD subunits which form a gated channel of ~880 kDa [36, 37]. Intriguingly, the T2SS is possibly only fully assembled transiently, perhaps triggered by the presence of exoproteins in the periplasm [19, 38].

Over the decades, an increasing number of bacterial multi-protein machineries spanning the inner and outer membrane of Gram-negative bacteria have been uncovered. The system closest related to the T2SS is the type IV pilus system (T4PS) [39]. There are at least two types of T4PS exist: the Type 4a Pilus system (T4aPS) and the Type 4b Pilus

system (T4bPS). The best studied T4aPS differs from the T4bPS in several ways, including a different major pilin subunit [40] and a different protein and domain organization of the homolog of the T2SS inner membrane protein GspL (Figure 2-1B). T4PS variants perform a diversity of functions in a wide range of species [40-42]. More distantly related systems are the bacterial transformation system and the archaellum assembly system (AAS) [1, 43, 44]. The critical functions of these systems in bacterial survival and pathogenicity increase the importance of our understanding the T2SS.

Among the many protein components of the T2SS, the secretion ATPase GspE plays an essential role, and is thought to be responsible for providing energy for the protein translocation process [27, 45-47]. GspE has several other, species-specific, names and is e.g. in *Vibrio* species called EpsE [27]. Here we will use the generic nomenclature, i.e. GspE. GspE is a protein of ~500 residues folding into three major domains, the N-terminal domains N1E and N2E, and the C-terminal domain CTE. The CTE can be subdivided into the subdomains C1E, CME and C2E (Figure 2-1A) [48, 49], where the CME is the critical zinc-binding domain [47, 50]. In a few species an additional domain occurs prior to N1E [51]. The amino acid sequences of *V. vulnificus* and *V. cholerae* GspE share 48, 94 and 90% identity for the N1E, N2E and CTE, respectively (Figure 2-2).

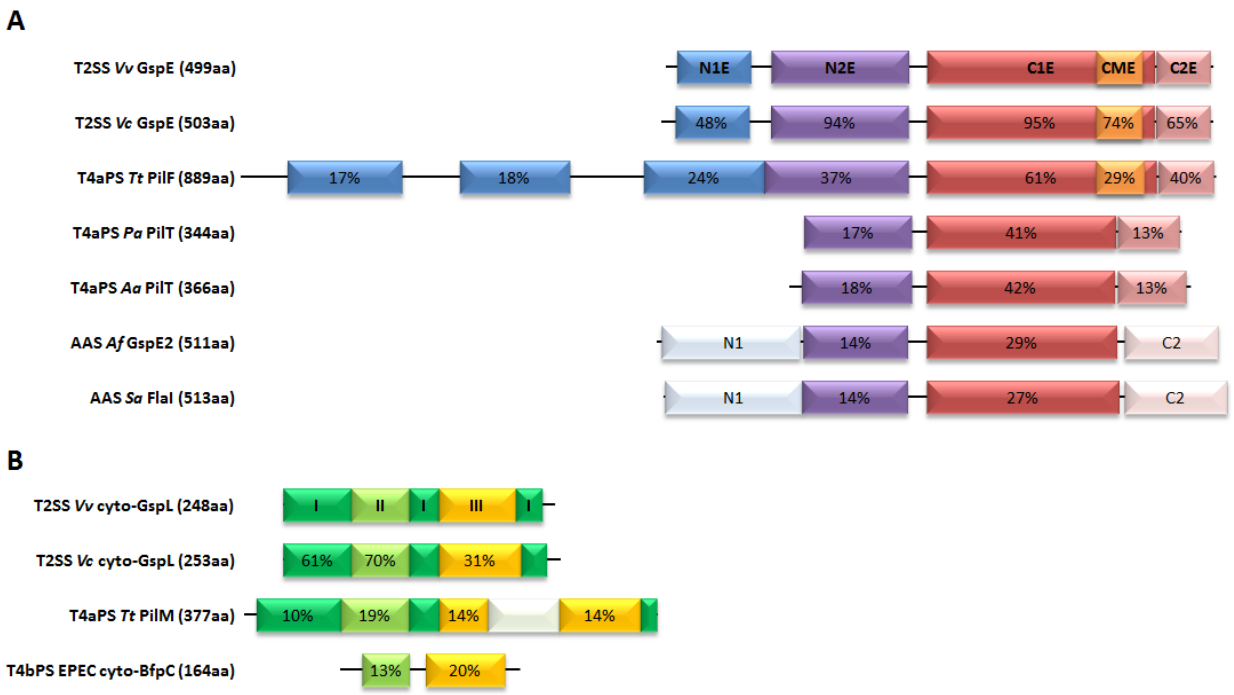


Figure 2-1: Domain bar diagrams of GspE and cyto-GspL homologs. For domains that are homologous in structure to GspE, the percentage sequence identity is given compared to *Vv* GspE. (A) GspE and homologous ATPases with electron microscopy reconstructions or crystal structures. *Vv* GspE: *Vibrio vulnificus* GspE. *Vc* GspE: *Vibrio cholerae* GspE. *Tt* PilF: *Thermus thermophilus* assembly ATPase PilF. *Pa* PilT: *Pseudomonas aeruginosa* retraction ATPase PilT. *Aa* PilT: *Aquifex aeolicus* retraction ATPase PilT. *Af* GspE2: one of three related ATPases in *Archaeoglobus fulgidus*; *Sa* FlaI: *Sulfolobus acidocaldarius* FlaI. (B) cyto-GspL and homologs with known crystal structures. The white bar in T4aPS *Tt* PilM indicates sub-domain 2B that is missing in T2SS cyto-GspL. *Vv* cyto-GspL: *Vibrio vulnificus* cytoplasmic domain of GspL. *Vc* cyto-GspL: *Vibrio cholerae* cytoplasmic domain of GspL. *Tt* PilM: *Thermus thermophilus* PilM. EPEC cyto-BfpC: enteropathogenic *Escherichia coli* cytoplasmic domain of BfpC.

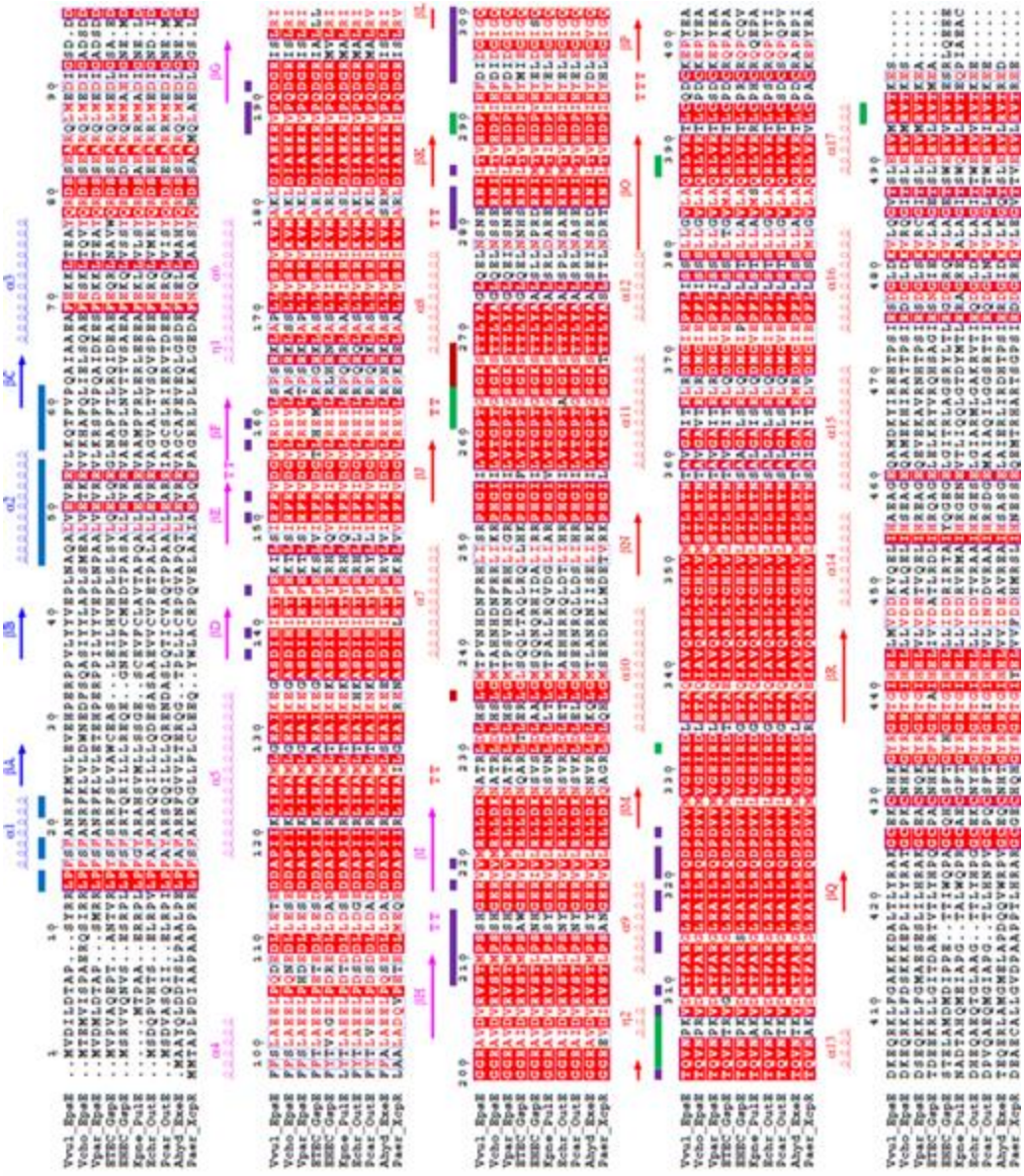


Figure 2-2: Family sequence alignment of GspE. Representative aligned sequences from a broad family of T2SS GspE. Shown are *Vibrio vulnificus, Vibrio cholerae, Vibrio parahaemolyticus,* enterotoxigenic *Escherichia coli,* enterohaemorrhagic *Escherichia coli, Klebsiella pneumonia, Erwinia chrysanthemi, Pectobacterium carotovorum, Aeromonas hydrophila, Pseudomonas aeruginosa* GspE homologs. The secondary structure elements of *Vibrio vulnificus* GspE are annotated at the top with N1E in blue, N2E in pink and CTE in red. The bar above the alignment indicates key residues in the interactions of N1E•cytoL (blue), cytoL•CTE (green), CTE•N2E' (purple), and AMPPNP•CTE (red).

Here we report the first crystal structure containing a full-length T2SS secretion ATPase, while previous structures of the T2SS GspE missed the N1E. The initial *V. cholerae* $^{\Delta N1E}$GspE structure contained an arrangement of molecules with 6_1 helical symmetry [49]. Solution studies have provided evidence that GspE tends to form multimers, most likely hexamers [25, 46]. Recently, crystal structures of two different hexamers of *V. cholerae* $^{\Delta N1E}$GspE have been obtained by using an "assistant hexamer", Hcp1 (Chapter 3), which served to induce multimer formation of the fused $^{\Delta N1E}$GspE chains. One of these *V. cholerae* $^{\Delta N1E}$GspE hexamers adopts an arrangement with quite regular, quasi C_6, symmetry, another hexamer is elongated exhibiting C_2 symmetry (Chapter 3) . These hexamers reveal considerable variability in the orientation of the N2E versus the CTE. In contrast, the association of a CTE and a N2E' from a neighboring subunit is remarkably similar in both hexamers of $^{\Delta N1E}$GspE-Hcp1 fusion proteins as well as in the helical *V. cholerae* $^{\Delta N1E}$GspE structure. This CTE•N2E' "construction unit" has also been observed in ATPase hexamers from related systems such as in the retraction ATPase PilT from the *Pseudomonas aeruginosa* and *Aquifex aeolicus* T4aPS [52, 53], and in the ATPases from the AAS, *Archaeoglobus fulgidus* GspE2 and *Sulfolobus acidocaldarius* FlaI [54, 55]. These latter ATPases lack the CME, and contain either no N1E at all, or an N1E with a different fold from the T2SS N1Es (Figure 2-1A). Hence the T2SS, T4PS and AAS ATPases share two common core domains, the N2E and CTE. These domains display major variations in length and number of subdomains, and are often distantly related in sequence. While T4PS and AAS ATPases form hexamers readily, the T2SS ATPase has so far been captured only as a stable hexamer when fused to Hcp1 as assistant hexamer (Chapter 3).

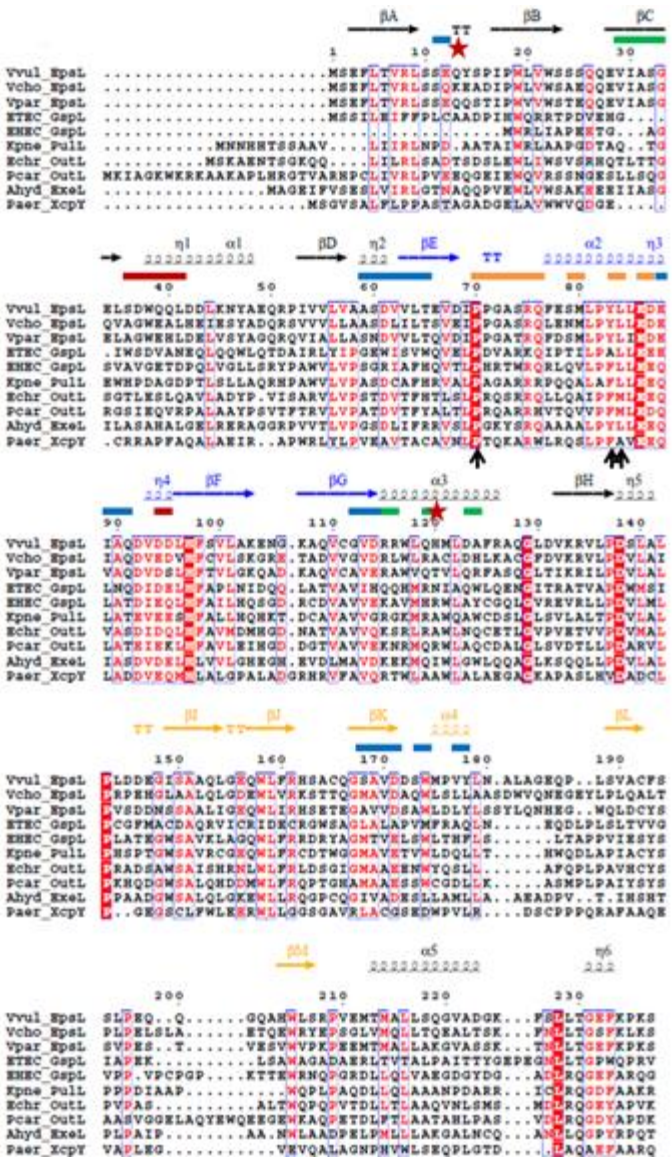


Figure 2-3: Family sequence alignment of the T2SS cyto-GspL from representative species. Shown are *Vibrio vulnificus, Vibrio cholerae, Vibrio parahaemolyticus,* enterotoxigenic *Escherichia coli,* enterohaemorrhagic *Escherichia coli, Klebsiella pneumonia, Erwinia chrysanthemi, Pectobacterium carotovorum, Aeromonas hydrophila, Pseudomonas aeruginosa* cyto-GspL homologs. The secondary structure elements of *Vibrio vulnificus* cyto-GspL are annotated at the top, with subdomain I in black, II in blue and III in yellow. The bar above the alignment indicates key residues in the interactions of β-interface (brown), α-interface (orange), AMPPNP•CTE (red star), cytoL•CTE (green) and N1E•cytoL (blue). Black arrows indicate complete or highly conserved residues in the α-interface.

Another important T2SS protein is GspL, which in *Vibrio* species is also called EpsL [27], and has additional, species-specific, names (Figure 2-3). We use here the generic name GspL. GspL is a bitopic inner membrane protein that plays a central role in T2SS function since it interacts with several other T2SS proteins, including (1) the inner membrane platform protein GspM [56]; (2) the major pseudopilin GspG [57]; and (3) GspE [26, 27]. The cytoplasmic domain of GspL (cyto-GspL) is responsible for the interactions with the first domain of GspE [58] and consists of three subdomains with similarities to proteins belonging to the actin-like ATPase superfamily [59].

In the related T4aPS, the "homolog" of GspL is actually two separate proteins, PilM and PilN, corresponding to the cytoplasmic and periplasmic domains of GspL, respectively. PilM and PilN from *Thermus thermophilus* interact with each other in the cytoplasm via the N-terminus of PilN. PilM does not hydrolyze but binds ATP using a subdomain which is absent in *V. cholerae* cyto-GspL [26, 59, 60]. In the T4bPS responsible for the biogenesis of the bundle-forming pilus of EPEC, the homolog of GspL is a single protein, BfpC. The N-terminal domain of BfpC has only two of the three sub-domains in cyto-GspL [61]. Hence, there are distinct differences between the homologs of GspL in the T2SS, T4aPS and T4bPS, in particular regarding the cytoplasmic subdomains (Figure 2-1B).

Two structures of *V. cholerae* cyto-GspL have been reported previously. The crystals containing *V. cholerae* cyto-GspL [59] revealed a three-subdomain architecture and also interactions between neighboring cyto-GspL subunits. The crystals containing *V. cholerae* cyto-GspL in complex with the N1E of *V. cholerae* GspE [26] showed the N1E•cyto-GspL heterodimer, revealing the interactions between the ATPase and inner membrane protein. Interestingly, in the crystals of this heterodimer, interactions were observed between neighboring cyto-GspL subunits, which are similar to interactions seen in the crystals of cyto-GspL by itself [59].

The reluctance of T2SS ATPases by themselves to form hexamers in solution [25, 47-49], combined with evidence that cyto-GspL stimulates the ATPase activity of GspE [46] and

indirect evidence for hexamer formation in solution [25, 45], encouraged us to undertake a study of a full-length T2SS GspE in complex with the cytoplasmic domain of GspL, to investigate whether or not the presence of cyto-GspL would induce hexamer formation of T2SS ATPases. We report here the 2.7 Å resolution crystal structure of the *V. vulnificus* GspE•cyto-GspL complex with bound AMPPNP. The three copies of this complex per asymmetric unit are very similar, but no hexameric arrangement of GspE is present in the crystals. Entirely unexpected was the close interaction of cyto-GspL with the CTE of GspE and the bound AMPPNP. A novel N2E-*vs*-CTE orientation, which differs considerably from that in previously solved structures, was observed, expanding the range over which the N2E of T2SS ATPases can move and rotate with respect to the CTE. Amazingly, the current crystals of the *V. vulnificus* GspE•cyto-GspL complex contains essentially the same two types of interfaces between cyto-GspL domains as present in the crystals of *V. cholerae* *c*yto-GspL and of the *V. cholerae* N1E•cyto-GspL heterodimer. In one of the interfaces a highly conserved hydrophobic cluster of residues is responsible for cyto-GspL•cyto-GspL contacts. The possible significance of these contacts for the T2SS is discussed.

2.2 METHODS

2.2.1 Protein Expression and Purification

The DNA fragments corresponding to GspE and cyto-GspL were PCR amplified from genomic DNA of *V. vulnificus* YJ016 and cloned into a modified pET-21d vector (EMD Millipore) to create a bicistronic operon. The two *V. vulnificus* proteins, comprising residues 1-499 of GspE and 1-237 of GspL, with the latter carrying a C-terminal His_6-tag, were expressed in BL21(DE3) *Escherichia coli* cells and purified via Ni^{2+} immobilized affinity chromatography using standard methods. The final purification included size-exclusion chromatography using Superdex 200 column (GE Healthcare) in a 20 mM HEPES, pH7.5, 200 mM NaCl buffer.

2.2.2 Crystallization and Data Collection

The *V. vulnificus* GspE•cyto-GspL complex was crystallized in the presence of 1 mM AMPPNP and 2 mM $MgCl_2$. Crystals were grown using the sitting drop vapor diffusion method by mixing 1 µL of protein solution and 1 µL of reservoir solution at 294 K. The reservoir solution was 0.2 M Na malonate pH 7.0, 18% PEG 3350. Crystals were cryoprotected using reservoir solution supplemented with 20% ethylene glycol. Data were collected on SSRL beam line BL12-2. Data were integrated and scaled using XDS and XSCALE [62]. Data collection statistics are provided in Table 1.

2.2.3 Crystal Structure Determination

The 2.7 Å resolution structure was solved by molecular replacement with the program Phaser [63] using the structures of *V. cholera*e $^{\Delta N1E}$GspE (PDB 1P9R) and the *V. cholera*e N1E•cyto-GspL complex (PDB 2BH1) as search models [26, 49]. The crystals of the *V. vulnificus* GspE•cyto-GspL complex belong to space group C2 and contain three copies of the ~83 kDa complex per asymmetric unit. The model was iteratively built using Coot [64] and refined with noncrystallographic symmetry restraints using REFMAC [65], resulting eventually in a R_{work} of 25.5% and a R_{free} of 28.9% with good geometry (Table 1). The electron density for almost all residues, including the 14-residue linker between the N2E and CTE in each of the three *V. vulnificus* GspE subunits, is well defined. Electron density is absent, however, for the linker residues (residues 79-95) between the N1E and N2E. Density is also missing for the first eleven and last two N- and C-terminal residues of GspE, and for the last two residues of cyto-GspL.

2.2.4 Structure Analysis and Figure Preparation

Structures were superimposed and compared to other structures using SSM [66] in Coot [64]. Interfaces between domains were analyzed with PISA [67]. Figures were

prepared using Pymol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.) and ESPript [68].

Table 1. Data collection and refinement statistics

Space Group	C2
Unit cell dimensions	226.4, 133.9, 93.5, 90, 91.4, 90
Resolution range (Å)	43.5, 2.69 (2.76-2.69)[a]
Unique reflections	72192 (5110)
Average redundancy	4.8
Completeness (%)	98.0 (93.8)
Rsym	7.4% (76.9%)
<I/sigI>	10.4 (1.51)
Rwork/Rfree	25.5/28.9 (35.8/42.1)
Rms bond length (Å)	0.007
Rms bond angle (Deg)	0.963
Ramachandran outliers[b]	5 (0.26%)
Ramachandran favored	1896 (97.58%)
Number of residues per AU	2238 (three ^{full}E•cyto-GspL complexes)
Number of water molecules	21
Average B factors (Å^2):	
Protein	59.4
AMPPNP	38.3
Zn	55.8

[a] Values in parenthesis correspond to the highest-resolution shell.
[b] Calculated using the Molprobity server [69].

2.3 RESULTS

2.3.1 Three *V. vulnificus* GspE•cyto-GspL Complexes

The crystals of GspE•cyto-GspL contain three copies of full-length *V. vulnificus* GspE and three copies of cyto-GspL domains per asymmetric unit with a total molecular weight of ~250 kDa. The current 2.7 Å resolution structure is, to the best of our knowledge, the first structure containing a full-length ATPase from a T2SS. The overall B factors of the three complexes in the asymmetric unit are similar and range from 58 to 61 $Å^2$. The average B-factor of the three cyto-GspL domains is ~53 $Å^2$, of the N1E domains ~78 $Å^2$, of the N2E domains ~63 $Å^2$, and of the CTE ~59 $Å^2$, indicating that N1E is the most flexible domain per complex. The distances between the last visible residue of N1E and the first visible residue of the nearest three N2E domains in the crystal varies from 33 to 41 Å. This means that, given the ~17 residue-linker length, it is not possible to establish which N1E is actually connected to which N2E in the crystals. Since the three complexes are very similar, we describe below complex 1 (Figure 2-4).

Key results are: (i) a new N2E-vs-CTE orientation angle, considerably larger than seen before; (ii) the cyto-GspL domain unexpectedly interacting with the CTE and with AMPPNP; and (iii) the presence of a linear arrangement of cyto-GspL domains observed now in three entirely different cyto-GspL-containing crystals. In these linear arrangements, two types of interfaces alternate. In one of these interfaces a conserved hydrophobic cluster of residues contacting each other across a twofold axis. The consistent occurrence of these linear arrangements in different crystal forms leads to the suggestion of a pre-assembly complex of inner membrane and GspE proteins.

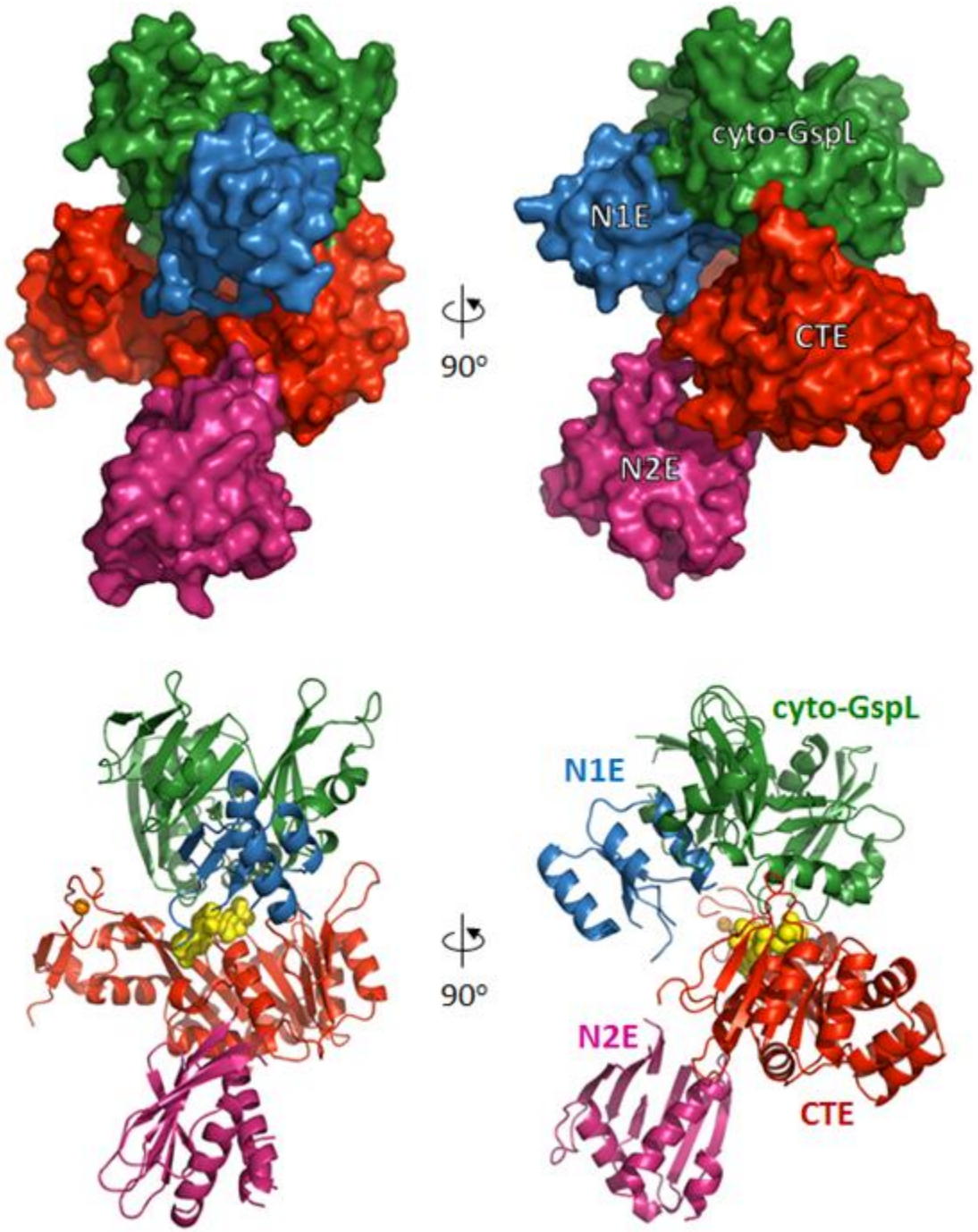


Figure 2-4: Overall structure of a *V. vulnificus* GspE•cyto-GspL complex. Complex #1 of N1E•cyto-GspL•CTE•N2E in the asymmetric unit of the GspE•cyto-GspL crystal. Domains are shown with N1E in blue, cyto-GspL in green, CTE in red, and N2E in pink. AMPPNP atoms as yellow spheres. Zn ion as orange sphere. For nomenclature of the domains, see Figure 2-1.

2.3.2 The *V. vulnificus* GspE Subunit

Superposition of *V. vulnificus* N2E in the current structure onto N2E from the *V. cholerae* $^{\Delta N1E}$GspE structure [49] yields an r.m.s.d. value of 0.49 Å , and the superposition of *V. vulnificus* CTE and *V. cholerae* CTE in these two structures an r.m.s.d. value of 0.61 Å. These low r.m.s.d values reflect the 94 and 90 % amino acid sequence identities for these two domains from these two species. While the individual domains in these two structures superimpose very well, the orientation of the N2E with respect to the CTE differs, however, by not less than 171 degrees (Figure 2-5). The two different orientations are observed in spite of the fact that in each instance the CTE binds Mg•AMPPNP. This difference in relative N2E-*vs*-CTE orientation in two structures is accompanied by a completely different conformation of the 14 residues in the linker connecting the N2E and CTE. The N2E-*vs*-CTE orientation seen in the current structure is also different from the four unique N2E-*vs*-CTE orientations observed in the recently published *V. cholerae* $^{\Delta N1E}$GspE hexamers (Chapter 3). Clearly, the flexible linker between the domains allows the N2E a large degree of freedom to move with respect to the CTE.

The most prominent interface in the current crystals occurs between the CTE of one *V. vulnificus* GspE chain and the N2E' of an adjacent GspE chain, with a buried solvent accessible surface area of ~1913 Å^2. This interface is similar in size to that observed in helical and hexameric arrangements for *V. cholerae* $^{\Delta N1E}$GspE [48, 49] (PDB identifiers 1P9W, 4KSR, 4KSS), and CTE•N2E' have been called construction units of hexameric secretion ATPases (Chapter 3). Structural comparisons of the *V. vulnificus* CTE•N2E' construction unit with that from the helical *V. cholerae* $^{\Delta N1E}$GspE arrangement, and the four constructions units from the recently reported hexamers in *V. cholerae* $^{\Delta N1E}$GspE-Hcp1 fusions, yield r.m.s.d. values between 0.50 and 0.73 Å for ~335 residues, showing that these construction units are essentially the same despite entirely different crystal environments.

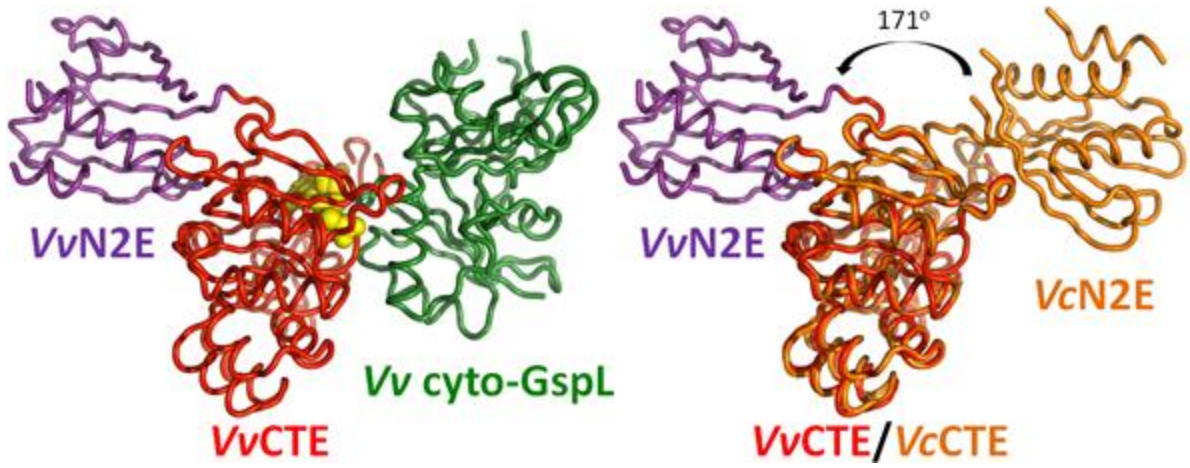


Figure 2-5: The variability of N2E-*vs*-CTE orientation in GspE. Left: Overall structure of *V. vulnificus* N2E-CTE (N2E purple, CTE red) in complex with *V. vulnificus* cyto-GspL (green), shown as cartoon with bound AMPPNP (yellow spheres). Right: Comparison of *V. vulnificus* and *V. cholerae* N2E-CTE "di-domains" in the current *V. vulnificus* GspE•cyto-GspL structure and *V. cholerae* $^{\Delta N1E}$GspE [49]. Superimposed are the CTEs from *V. vulnificus* (red) and *V. cholerae* GspE (orange). There is a dramatic change of ~171 degrees in N2E-*vs*-CTE orientation in these two cases – compare the *V. vulnificus* N2E (purple) and the *V. cholerae* N2E (orange) domains. Comparison with the left panel also shows that *V. vulnificus* cyto-GspL in the current structure occupies approximately the same position as N2E in the *V. cholerae* N2E-CTE didomain structure [49].

This CTE•N2E' interface has also been consistently observed in the hexameric structures of the T2SS GspE homologs *Pseudomonas aeruginosa* PilT and *Aquifex aeolicus* PilT from the T4PS, and *S. acidocaldarius* FlaI and *A. fulgidus* GspE2 from the AAS [52-55]. The mutual orientation of the N2E' and CTE in seven structures of these distantly related ATPases from a variety of secretion and assembly systems, with mutual amino acid sequence identities in the N2E domains as low as 14 %, differs by as little as 11 degrees.

2.3.3 The Structure of the *V. vulnificus* cyto-GspL Domain

The current *V. vulnificus* cyto-GspL structure superimposes onto that of *V. cholerae* cyto-GspL [59] with an r.m.s.d. of 1.4 Å, and onto cyto-GspL from the *V. cholerae* N1E•cyto-GspL complex [26] with an r.m.s.d. of 1.2 Å with amino acid sequence identities of 61, 70 and 31 % for subdomains I, II and III, respectively. Clearly, the three-subdomain structure of cyto-GspL is essentially the same in these three crystal structures. While 1 mM AMPPNP was present in the crystallization solution, we observed no density representing this nucleotide analog bound to *V. vulnificus* cyto-GspL at a position equivalent to that of the distant homolog PilM of the *T. thermophilus* T4aPS, in agreement with the fact that subdomain 2B of PilM, involved in nucleotide binding by PilM, is absent in GspL proteins from the T2SS [26, 59, 60] (Figure 2-1B).

2.3.4 The Interaction between *V. vulnificus* N1E and cyto-GspL

The second largest interface in the current crystals occurs between *V. vulnificus* N1E and cyto-GspL domains, and buries ~1823 $Å^2$. Each N1E in the crystal makes only contacts with a cyto-GspL domain, mainly by placing its α2 helix in the cleft between domains II and III of cyto-GspL (Figure 2-6). The interface in the *V. cholerae* N1E•cyto-GspL interaction [26] buries ~1758 $Å^2$, a number close to that in the current *V. vulnificus* N1E•cyto-GspL structure. Superposition of these two *Vibrio* N1E•cyto-GspL complexes yields an r.m.s.d. of 1.4 Å while many of the interface residues are hydrophobic and highly conserved, as can be

seen from the amino acid sequence alignment of a diverse set of cyto-GspL sequences (Figure 2-3).

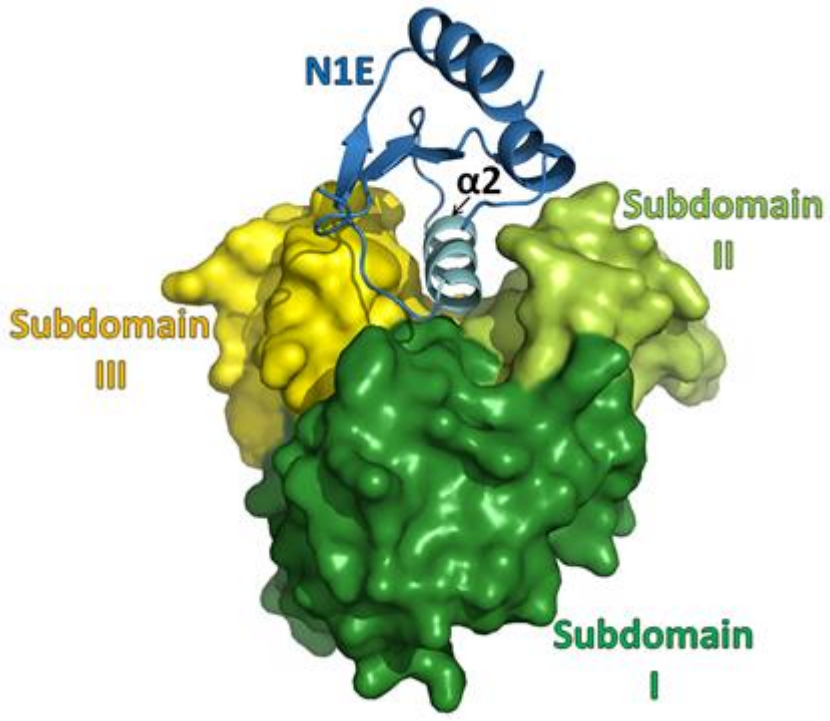


Figure 2-6: Structure of the *V. vulnificus* N1E•cyto-GspL Heterodimer. The N1E of GspE shown in blue, subdomain I of cyto-GspL in green, subdomain II of cyto-GspL in lime, subdomain III of cyto-GspL in yellow. Note how all three GspL subdomains are involved in contacting N1E with helix α2 of N1E, a major component of the interface. When compared with the *V. cholerae* N1E•cyto-GspL heterodimer [26] the r.m.s.d. is 1.4 Å with a difference of ~11 degrees in N1E-*vs*-cyto-GspL orientation, and ~ 48 and 53 % amino acid sequence identity for N1E and cyto-GspL, respectively (see also Figure 2-2 and 2-3).

2.3.5 AMPPNP Binding by *V. vulnificus* CTE and cyto-GspL

Density for AMPPNP is clearly present in the active site of all three *V. vulnificus* CTE domains in the current structure (Figure 2-7). The Mg^{2+} ion interacts with oxygens from the β- and γ-phosphates of AMPPNP. AMPPNP is bound to *V. vulnificus* CTE by the side chains of residues Leu235 and Arg437 which interact with the adenine (Figure 2-8A). Residues from the highly conserved P-loop Gly265-Lys266-Ser267-Thr268 make extensive contacts with the phosphate moiety of the nucleotide. The oxygen atoms of the α-phosphate form hydrogen bonds with the backbone amide groups of Ser267 and Thr268, and the side-chain oxygen of Thr268. The oxygen atoms of the β-phosphate form hydrogen bonds with the backbone amide groups of Gly265, Lys266 and Ser267, and with the side chains of Ser264, Lys266 and Ser267. The side chain amino group of Lys266 forms hydrogen bonds with the oxygen atoms of the γ-phosphate.

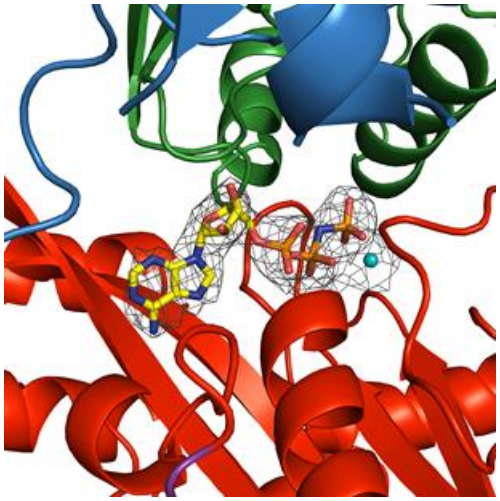


Figure 2-7: Electron density of nucleotide in the *V. vulnificus* GspE•cyto-GspL structure. A (2Fobs-Fcalc) difference electron density at the 1.5 sigma level is shown. Proteins are shown in ribbon with CTE in red, N1E in blue, cyto-GspL in green. AMPPNP is shown in sticks with carbon in yellow, nitrogen in blue, oxygen in red, phosphorous in orange. Mg is shown in cyan sphere. The phases were calculated without including nucleotide coordinates.

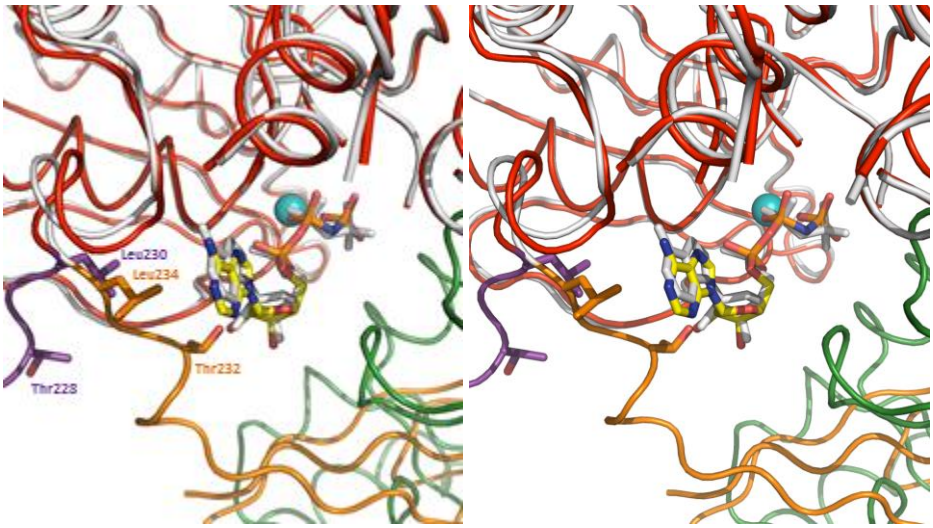
The interactions of CTE residues with AMPPNP are similar in the *V. vulnificus* and *V. cholerae* secretion ATPases (Figure 2-8A). Whereas residues 239-497 of the CTE domains in the *V. vulnificus* GspE and *V. cholerae* $^{\Delta N1E}$GspE structures [49] adopt almost the same conformation, the loop formed by residues 225-238 connecting the N2E and CTE adopts a very different conformation in the current structure compared to the helical *V. cholerae* $^{\Delta N1E}$GspE structure, as mentioned before. As a result of this conformational change, residues Thr228 and Leu230 of the *V. vulnificus* GspE structure are not in contact with the ribose and adenine moieties, respectively, while the equivalent Thr232 and Leu234 in the *V. cholerae* GspE structure are (Figure 2-8A).

Surprisingly, the cyto-GspL domain is also in contact with AMPPNP, specifically: (1) the $N^{\varepsilon 2}$ of Gln13 of *V. vulnificus* cyto-GspL makes hydrogen bonds with an oxygen of the γ-phosphoryl group; (2) the $O^{\varepsilon 1}$ of Gln13 interacts with the NH of AMPPNP bridging the β- and γ-phosphorous atoms; (3) the $N^{\varepsilon 2}$ of His120 forms an hydrogen bond with an oxygen of the γ-phosphoryl group; and (4) the side chain of Tyr14 contacts the ribose through hydrophobic interactions (Figure 2-8B). This unanticipated result will be discussed below.

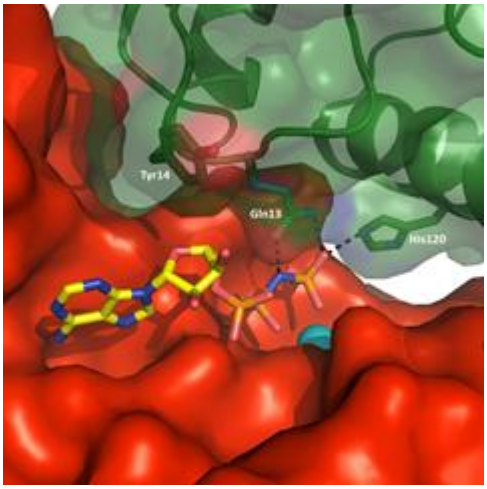
2.3.6 The Interactions between *V. vulnificus* CTE and cyto-GspL

A most unexpected finding in our current structure is that *V. vulnificus* cyto-GspL interacts with the CTE domain while burying ~1349 $Å^2$ solvent accessible surface in the interface. Three residues from cyto-GspL involved in the contacts are engaged in salt bridges: Asp37, Arg116 and Asp123 are within 2.9-3.6 Å from the side chains of residues Arg494, Asp289 and Arg332 of the CTE, respectively. Additional contacts occur between side chains of Gln39 and Arg116 of cyto-GspL form hydrogen bonds with main chain atoms of CTE residues Val303 and Leu388 (Figure 2-8C). Residues Ser36, Trp38, Leu41 and Ala124 of cyto-GspL and residues Pro261, Thr262, Gly263, Ser264 and Val 495 of CTE contribute van der Waals interactions to the interface.

A


Leu230
Leu234
Thr232
Thr228

B


Tyr14
Gln13
His120

C

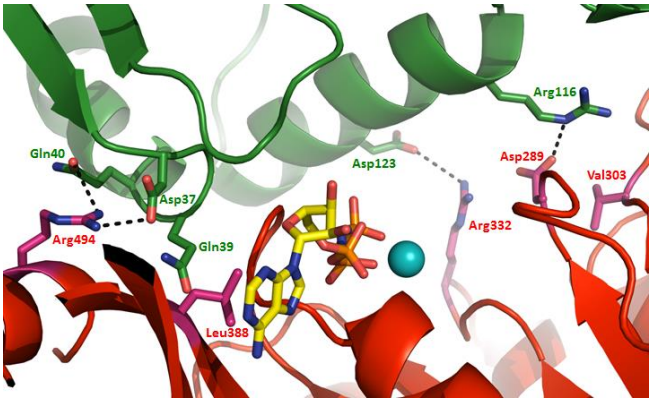

Arg116
Gln40
Asp123
Asp289
Val303
Asp37
Arg494
Gln39
Arg332
Leu388

Figure 2-8: Nucleotide binding by *V. vulnificus* GspE•cyto-GspL and *V. cholerae* GspE. (A) Residues that interact with AMPPNP in the *V. cholerae* GspE•AMPPNP complex but do not interact with AMPPNP in the V. vulnificus GspE•cyto-GspL•AMPPNP, due to a major change in conformation of the N2E-CTE linker (225-238), are shown in sticks. *V. vulnificus*: CTE red, N2E-CTE linker purple, cyto-GspL green, AMPPNP as sticks with yellow carbons, and Mg in cyan. *V. cholerae*: N2E and N2E-CTE linker orange, CTE white, and AMPPNP as white sticks (PDB 1P9W). Some CTE residues are removed for clarity. (B) *V. vulnificus* cyto-GspL•AMPPNP interactions. Cyto-GspL residues Gln13 and His120 interact with the PNP of AMPPNP. Specifically, the side chains of Gln13 and His120 form hydrogen bonds with the oxygen atoms of γ phosphorous atom with distances of ~2.5-2.8 Å. The side-chain oxygen of Gln13 forms a hydrogen bond with a distance of ~2.7 Å with the nitrogen linking the β and γ phosphor atoms. Tyr14 makes hydrophobic contacts with the ribose of AMPPNP. (C) Interactions between *V. vulnificus* Cyto-GspL and CTE. Key residues contributing to the cyto-GspL and CTE interactions are shown in sticks. The three Arg-Asp salt bridges are indicated with dashed lines. Residues making hydrogen bonds with main chain atoms of another subunit are also labeled.

The position of cyto-GspL with respect to the CTE is quite remarkable. When comparing the cyto-GspL•CTE complex in the current structure with that of the N2E-CTE arrangement in the helical *V. cholerae* $^{\Delta N1E}$GspE structure [49], it appears that in the current structure the cyto-GspL chain occupies approximately the position of the N2E in the helical structure. The N2E has swung away to allow cyto-GspL to approach the AMPPNP and the active site of the CTE (Figure 2-5).

2.3.7 Interactions between cyto-GspL Domains: Similar "Rods" in Multiple Crystals

The *V. vulnificus* GspE•cyto-GspL structure provides an interesting opportunity to look at mutual interactions between cyto-GspL domains. In the current crystals, there are three independent Cyto-GspL domains per asymmetric unit. Two of these domains, #1 and #2, interact with each other mainly involving residues from the βC strand of subdomain I. This contact is hereafter called the "β-interface" (Figure 2-9A). These two domains are related by a pseudo-twofold axis. Each of these domains also interacts, at the opposite side, with a domain from a neighboring asymmetric unit engaging primarily residues from helix α2 and the loop connecting strand βE and helix α2 of subdomain II. This interface between cyto-GspL domains #1 and #2' is hereafter called the "α-interface" (Figure 2-9B). Also these pairs of domains are related by a pseudo-twofold axis. The third crystallographically independent cyto-GspL domain, #3, uses crystallographic twofold axes to form contacts with neighboring cyto-GspL domains with a α-interface on one side and a β-interface on the other side of the domain, essentially the same arrangement as between domains #1 and #2', and between #1 and #2, respectively (Figure 2-9A).

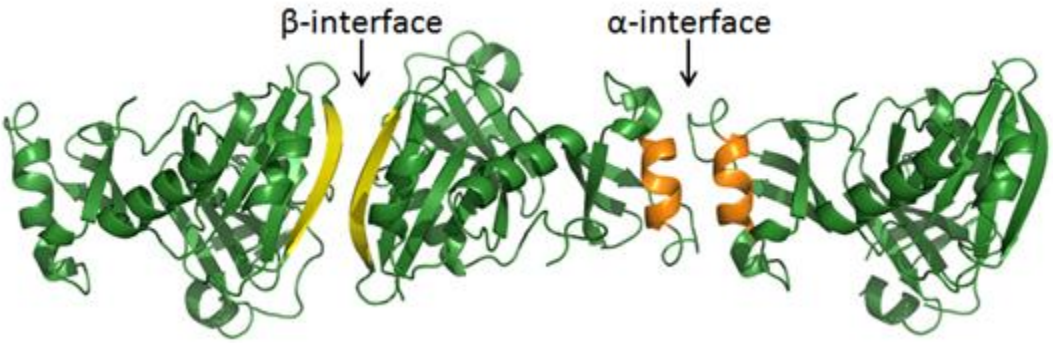
The α-interface (Figure 2-9B) buries ~1008 Å^2 solvent accessible surface area in the #1•#2 contact and ~835 Å^2 solvent accessible surface area in the #3•#3' contact. The #1•#2 and #3•#3' pairs of cyto-GspL domains are very similar and superimpose with an r.m.s.d. of ~0.6 Å. The α-interface has a hydrophobic center with Pro70 and Tyr83 engaged in hydrophobic interactions with Pro70, Gln76, Ser79, Met80, Tyr83 and Leu84 of the

neighboring subunit. Due to the twofold axis relating the two domains in this interface, this set of contacts occurs twice between this pair of cyto-GspL domains. Hydrogen bonds and electrostatic interactions involving Arg75, Gln76 and Asp87, occur on both sides of the hydrophobic center of the α-interface.

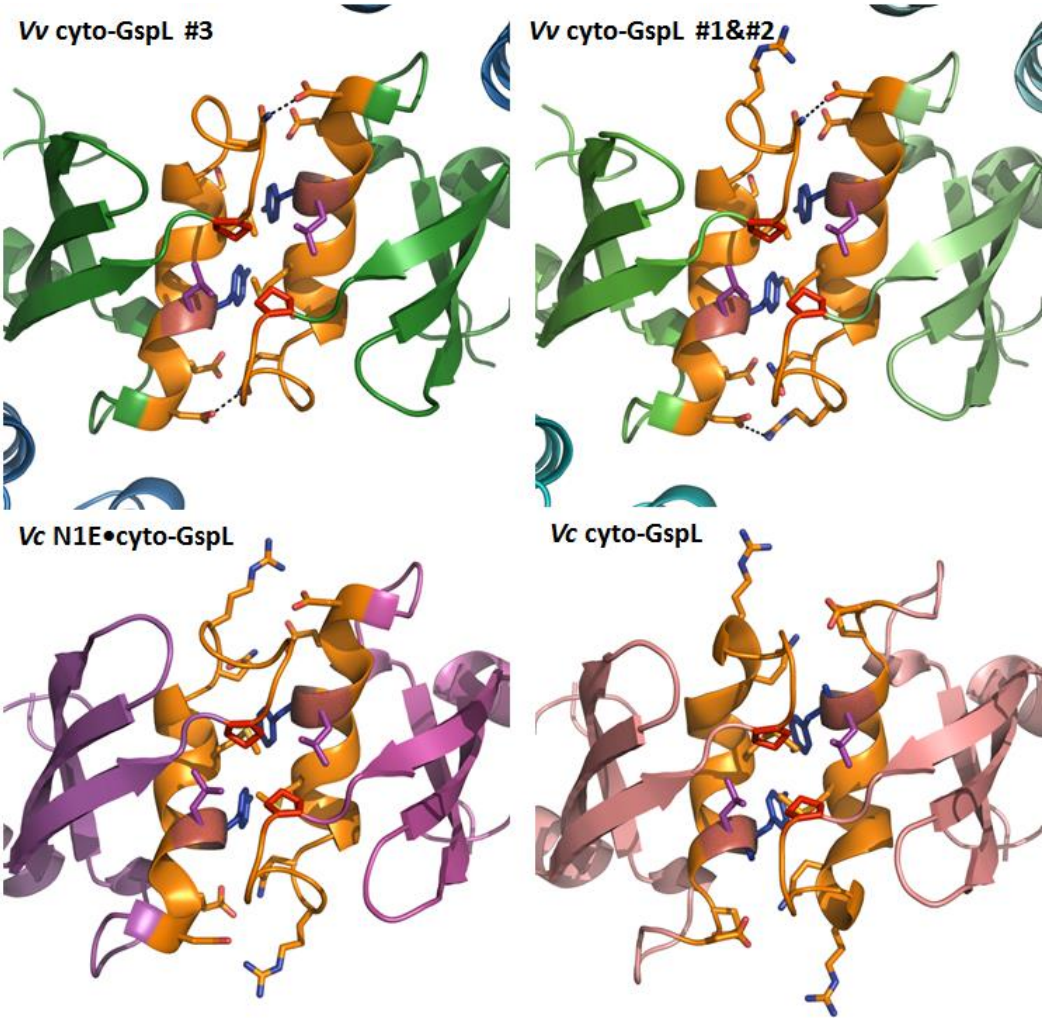
The β-interface (Figure 2-9A) buries ~530 $Å^2$ solvent accessible surface area in the #1•#2 contact and ~529 $Å^2$ in the #3•#3'' contact, where #3'' is related to #3 by a crystallographic twofold. The area buried in the β-interface is therefore almost half of that buried in the α-interface. The #1•#2 and #3•#3'' pairs of cyto-GspL domains superimpose with an r.m.s.d. of ~0.6 Å. The domains forming the β-interface are related by a pseudo-twofold axis. The major contacts in this interface are the hydrophobic interactions between residues Val29 and Glu34 and the two main-chain hydrogen bonds from Ser32, involving two anti-parallel βC strands, one from each domain.

Both the α-interface and β-interface, seen twice in the current crystals of *V. vulnificus* GspE•cyto-GspL, occur in two previous crystal structures. One of these crystals contains *V. cholerae* cyto-GspL [59] and the other contains the *V. cholerae* N1E•cyto-GspL heterodimer [26]. Hence, we have four crystallographically independent views of both the α-interface and the β-interface, in crystals grown under different conditions, in different crystal forms, and with very different protein content (Table 2). The interdomain angles between the various pairs of *Vibrio* cyto-GspL domains forming α- and β-interfaces in these crystals vary only by up to ~19 degrees (Table 2). In all four instances of the α-interface, a completely conserved Pro70 and two highly conserved hydrophobic residues Tyr83 and Leu84 (Figure 2-3) are forming the hydrophobic center of this interface. In addition, in three of the four structures (two *V. vulnificus* GspE•cyto-GspL and *V. cholerae* N1E•cyto-GspL) Arg75, Gln76 and Asp87 engage in hydrogen bonds or electrostatic interactions with twofold related side chains.

A


β-interface
α-interface

B


Vv cyto-GspL #3
Vv cyto-GspL #1&#2
Vc N1E•cyto-GspL
Vc cyto-GspL

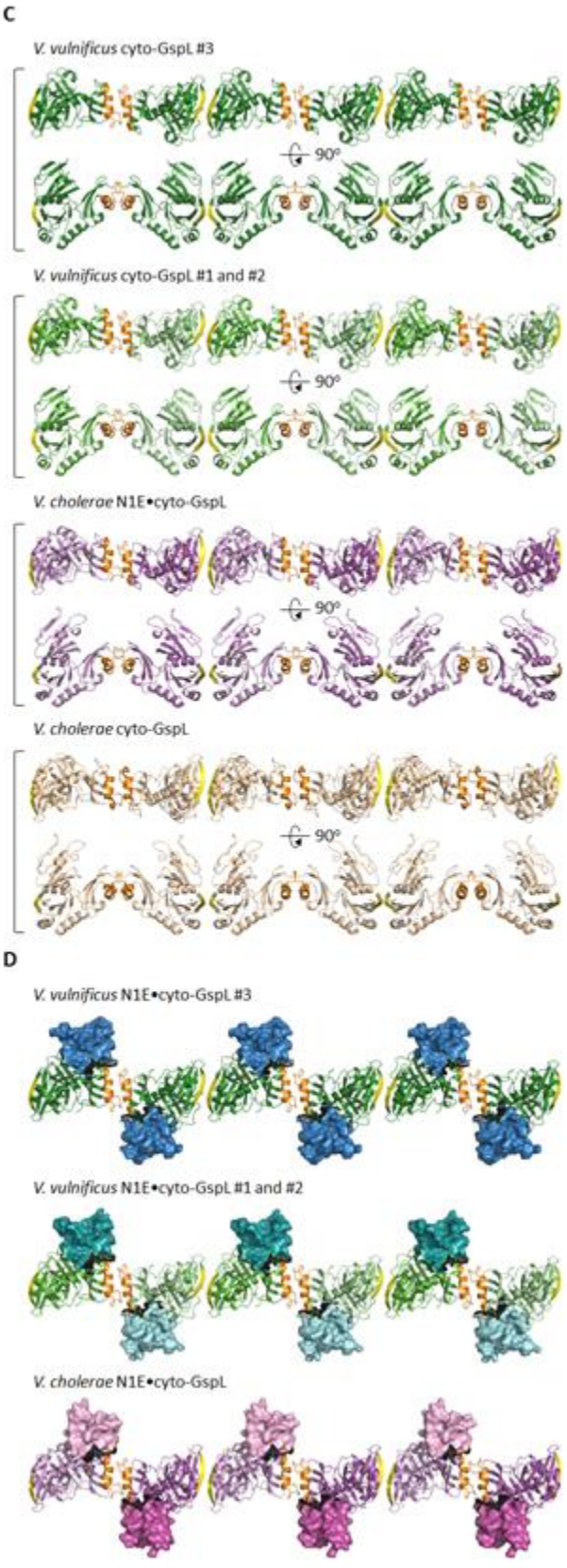

C
V. vulnificus cyto-GspL #3
90°
V. vulnificus cyto-GspL #1 and #2
90°
V. cholerae N1E•cyto-GspL
90°
V. cholerae cyto-GspL
90°
D
V. vulnificus N1E•cyto-GspL #3
V. vulnificus N1E•cyto-GspL #1 and #2
V. cholerae N1E•cyto-GspL

Figure 2-9: Linear arrangements of *Vibrio* cyto-GspL domains in multiple crystals. The α-interface and β-interfaces are colored orange and yellow, respectively. (A) The α- and β-interfaces among cyto-GspL domains in the current *V. vulnificus* GspE•cyto-GspL crystals. The interfaces between cyto-GspL domain #3 and crystallographically related domains are depicted. The interfaces between cyto-GspL domains #1 and #2 are essentially the same (see text). The α-interface is formed mainly by side chain contacts between residues of α2 helices (orange) and the loops between strand βE and helix α2 from two domains. The domains are related by a twofold axis approximately parallel to the direction of view. (See text for further description of the contacts). In the β-interface, main chain hydrogen bonds between antiparallel βC strands (yellow) are the main contacts between two subunits, which are also related by a twofold approximately parallel to the direction of view. (B) Close-ups of four similar cyto-GspL α-interfaces in three different crystal forms. Key residues are shown in sticks. Hydrogen bonds and electrostatic interactions are indicated with dashed lines. Note the completely conserved Pro70 (red) in all interfaces, twice in contact with the highly conserved Tyr83 (blue) and Leu84 (purple). Left upper: The α-interface between two neighboring cyto-GspL domains #3 in the current *V. vulnificus* GspE•cyto-GspL crystals. Right upper: The α-interface between neighboring cyto-GspL domains #1 and #2 in the current *V. vulnificus* GspE•cyto-GspL crystals. Left lower: The α-interface between neighboring cyto-GspL domains in the *V. cholerae* N1E•cyto-GspL crystals (PDB 2BH1) [26]. Right lower: The α-interface between neighboring cyto-GspL domains in the *V. cholerae* cyto-GspL crystals (PDB 1YF5) [59]. (C) Four similar linear arrangements of cyto-GspL domains in three different crystal forms. From top to bottom: cyto-GspL rods from *V. vulnificus* GspE•cyto-GspL complex #3; *V. vulnificus* GspE•cyto-GspL complex#1 and #2; *V. cholerae* N1E•cyto-GspL; and *V. cholerae* cyto-GspL. (D) Three linear arrangements of cyto-GspL domains with associated N1E domains in two different crystal structures. From top to bottom: *V. vulnificus* GspE•cyto-GspL complex #3; *V. vulnificus* GspE•cyto-GspL complex #1 and #2; and *V. cholerae* N1E•cyto-GspL.

TABLE 2. Characteristics of structures containing linear cyto-GspL arrangements

Structure		*Vv* #3	*Vv* #1 and #2	*Vc* N1E•cyto-GspL	*Vc* cyto-GspL
Crystallization Condition		0.2M Na Malonate pH7.0 18% PEG3350 1 mM AMPPNP 2 mM $MgCl_2$		0.2M $CaOAc_2$ 0.1M Bis-Tris-HCl pH 5.75 15-20% PEG3350	0.15M $CaOAc_2$ 0.1M Tris-HCl pH 8 11-15% PEG6000
Space Group		C2	C2	P2	C2
α-interface	BSA ($Å^2$)	878	960	1150	1102
	r.m.s.d.	-	0.10	1.23	1.43
	Kappa	-	4.1	19.1	16.0
β-interface	BSA ($Å^2$)	521	528	679	766
	r.m.s.d.	-	0.17	1.23	1.43
	Kappa	-	4.1	19.1	16.0

Vv #3: cyto-GspL of *V. vulnificus* GspE•cyto-GspL #3 complex; *Vv* #1 and #2: cyto-GspL of *V. vulnificus* GspE•cyto-GspL #1and #2 complex; *Vc* N1E•cyto-GspL: cyto-GspL of *V. cholerae* N1E•cyto-GspL (PDB 2BH1) [26]; *Vc* cyto-GspL: *V .cholerae* cyto-GspL (PDB 1YF5) [59].

R.m.s.d. and rotation angle (kappa) calculations are based on a superposition of cyto-GspL from three different structures to the cyto-GspL of *V. vulnificus* GspE•cyto-GspL #3 complex. The resultant superposition operation is applied to the entire cyto-GspL•cyto-GspL pairs. The second cyto-GspL in the pairs is subsequently superimposed on the cyto-GspL related to *Vv* #3 by a crystallographic twofold. The r.m.s.d. and interdomain angle of this second superposition are given in the Table.

The similarity of the interactions between cyto-GspL domain contacts in the three crystals extends far beyond that of nearest neighbors. In the three cyto-GspL-containing crystals with known structure now available, it appears that the same arrangement of adjacent subunits of cyto-GspL domains with alternating α- and β-interfaces form "cyto-GspL rods", spanning the entire crystal. The angular deviations seen in a pairwise comparison of α and β interfaces (Table 2) are such that essentially the same straight rods occur four times in these crystal forms (Figure 2-9C). Possible implications of this tendency to form such linear arrangements are discussed below.

2.4 DISCUSSION

2.4.1 The *V. vulnificus* GspE•cyto-GspL Complex

The 2.7 Å resolution crystal structure of the complex of *V. vulnificus* GspE and cyto-GspL, reveals many features of and contacts between domains and subdomains of these two key proteins of the T2SS. In particular, the cyto-GspL domain is involved in several surprising interactions. This leads to a hypothesis of a pre-assembly complex of this Inner Membrane Platform as described in the final section below.

2.4.2 The *V. vulnificus* GspE Subunit

The GspE subunit appears to be remarkably flexible, even when in complex with its partner, the cytoplasmic domain of GspL. Not only is the linker between the N1E and N2E invisible in the crystals, the current N2E-vs-CTE orientation is very different from the orientations in helical *V. cholerae* $^{\Delta N1E}$GspE [49] and in the two recently determined *V. cholerae* GspE hexamers (Chapter). In the current structure, the N2E and CTE domains of the GspE subunit are oriented with a novel N2E-*vs*-CTE orientation without contacts of the N2E with AMPPNP (Figure 2-4). The variability of conformations that the full-length GspE adopts during the biogenesis and action of the T2SS remains to be determined, but the

currently available structures of the T2SS ATPase indicate that this key T2SS protein can adopt a tremendous range of N2E-*vs*-CTE orientations.

In contrast to the surprising variation in the N2E-vs-CTE orientations observed within the same subunit, the CTE•N2E' construction unit, bringing two domains from different GspE subunits in contact, is very similar to that observed in the helical [49] and hexameric *V. cholerae* $^{\Delta N1E}$GspE structures, with r.m.s.d. values of 0.7 and 0.5 Å, respectively (Chapter 3). As previously discussed [48, 54], this same construction unit also occurs in secretion ATPases from the T4aPS [52, 53] and in the more distantly related homologs from the AAS [54, 55]. This provides further support for a general significance of the CTE•N2E' interactions in a wide variety of secretion and related systems, probably an early invention during the evolution of these systems.

Studies on purified GspE in solution indicates the presence of transient multimers, probably hexamers, as judged by e. g. ATPase activity in the size-exclusion chromatography fractions corresponding to hexamers [47]. It is apparently not easy to capture hexamers of GspE in structural studies, even in the presence of cyto-GspL as evident from the results in the current paper. So far, only fusion to the "assistant hexamer" Hcp1 has resulted in GspE hexamers (Chapter 3). In contrast, several GspE homologs do not display a similar reluctance to form hexamers [52-55]. Also, the assembly ATPase PilF from the *Thermus thermophilus* T4aPS forms, in the presence and absence of AMPPNP, hexamers as observed by negative-stained electron microscopy [70]. However, this T4aPS assembly ATPase has two additional N-terminal domains compared to T2SS assembly ATPases (Figure 2-1A), and these domains could possibly promote hexamer formation. The reason that the T2SS assembly ATPase is so reluctant to form hexamers remains a puzzle. Possibly contacts with additional inner membrane T2SS proteins, including GspF [24, 28], are required to obtain the putative hexameric arrangement of full-length GspE.

2.4.3 The *V. vulnificus* CTE•cyto-GspL Contacts and AMPPNP Binding

The quite extensive interface and direct contact of cyto-GspL with the CTE and with AMPPNP is a most surprising result of the current structure. Therefore, we looked into the conservation of the residues engaged in the CTE•cyto-GspL interface. Out of 14 cyto-GspL and 16 CTE interface residues, 2 and 11 residues are conserved, respectively (Figure 2-2 and 2-3). Seven of the conserved CTE residues are located in the highly conserved Walker A and Walker B boxes, which are functionally important for the ATP binding and catalysis in the secretion ATPases [27, 71, 72]. None of the residues of cyto-GspL engaged in the three Arg-Asp salt bridges across the *V. vulnificus* CTE•cyto-GspL interface (Figure 2-8C) are conserved in any other species, neither are the two *V. vulnificus* cyto-GspL residues, Gln13 and His120, that contact AMPPNP. In view of this lack of conservation, it is unlikely that these cyto-GspL•CTE interactions and the cyto-GspL•AMPPNP interactions are of physiological importance.

2.4.4 The Interaction between *V. vulnificus* N1E and cyto-GspL

The interactions between N1E and cyto-GspL are very similar in the T2SS from two *Vibrio* species, *V. vulnificus* and *V. cholerae,* with the mutual orientation of the two domains differing by less than 11 degrees (Figure 2-5) [26, 59]. Several N1E residues in the N1E•cyto-GspL interface are highly conserved. In fact, four (Leu14, Pro15, Glu50 and Arg53) out of the six completely conserved residues of N1E are located in this interface (Figure 2-2). The N1E•cyto-GspL interaction is therefore most likely of major importance in the T2SS of many species in order to link the ATPase GspE to the Inner Membrane Platform.

2.4.5 Linear Arrangements of cyto-GspL Domains – Possible Implications for T2SS Assembly

Most intriguingly, crystals of *V. cholerae* cyto-GspL, *V. cholerae* N1E•cyto-GspL and the current *V. vulnificus* GspE•cyto-GspL all contain essentially the same linear arrangement

of cyto-GspL domains consisting of straight rods with alternating α- and β-interfaces between adjacent cyto-GspL domains (Table 2; Figure 2-9C and D). While the β-interface is not very extensive, the α-interface buries almost twice as much surface area and has a hydrophobic center. Moreover, this hydrophobic interface center contains a completely conserved Pro70 residue, which is in contact with a highly conserved Tyr and Leu of helix α2 from an adjacent chain (Figure 2-3). These hydrophobic interactions occur around a twofold axis in the center of the α-interface (Figure 2-9A). Hydrophilic and electrostatic interactions between Arg75, Gln76 and Asp87, which are conserved hydrophilic residues across GspL family members (Figure 2-3), surround the hydrophobic interactions in this interface.

This linear arrangement of cyto-GspL domains has now been observed four times in three different crystals in the absence and presence of other protein chains (Table 2). This suggests that it is a favorable arrangement of cyto-GspL domains, which may also occur *in vivo*. This linear array might play a role during the biogenesis of the T2SS along the following lines:

- In the cell, rods of cyto-GspL domains resembling those observed in the crystals (Figure 2-9C), might be a transient arrangement in order to obtain a high local concentration of GspL subunits for rapid subsequent assembly of the Inner Membrane Platform. In this regard, it is of interest that periplasmic domains of *V. parahaemolyticus* GspL (peri-GspL) form dimers in solution as well as in crystals [73]. In the crystal lattice, this dimer interface buries 1487 $Å^2$ solvent accessible surface and involves a number of conserved residues. For convenience, we will call this the "p-interface" of GspL. Purified full-length GspL forms dimers [56]. It is conceivable that full-length GspL dimerizes with simultaneous inter-subunit interactions between the periplasmic and cytoplasmic domains. Full-length GspL dimers with an α-interface in the cytoplasm and the p-interface in the periplasm might be quite stable due to a total buried surface area of ~2400 $Å^2$ (Figure 2-10A). Multiple GspL dimers with α- and p-interfaces on opposite sides of the inner

membrane can then form linear arrangements using the β-interface as points of contact between cytoplasmic domains (Figure 10B).

- Since one cyto-GspL domain interacts with one N1E domain [25, 26, 60] (Figure 2-6), it is likely that $(GspE•GspL)_2$ hetero-tetramers form with each N1E domain tightly bound to a cyto-GspL domain. In this way, multiple GspE domains form a linear pre-assembly complex of $(GspE•GspL)_2$ hetero-tetramers (Figure 2-10C). Given the flexibility of the N1E-N2E linker, as evidenced by the lack of density for these linker residues in our current crystals, the N2E and CTE domains are at this stage highly flexible with respect to the N1E.
- Specific interaction of Inner Membrane Platform proteins with either the Outer Membrane Complex [19], or with exoproteins to be secreted [38], might be a trigger in assembling a T2SS (Figure 2-10D). For rapid assembly of the T2SS Inner Membrane Platform, rods consisting of $(GspE•GspL)_2$ hetero-tetramers in the inner membrane might provide a ready pool of building blocks for assembly.
- The C_2 symmetry of $(GspL•N1E)_2$ hetero-tetramers needs to be reconciled with the approximate C_6 symmetry of hexamers formed by the N2E and CTE domains. A key role might be played by the N1E-N2E linker that consists of ~19 residues. The "top" part of the $(GspL•GspE)_6$ multimer comprising GspL•N1E domains may adopt approximate C_3 symmetry, relating three $(GspL•N1E)_2$ heterodimers with each C_2 symmetry. Six flexible N1E-N2E linkers would connect this upper part to six "lower" N2E-CTE domains with approximate cyclic C_6 symmetry. This generates a $GspL_6•GspE_6$ assembly with approximate overall C_3 symmetry. This point has been discussed in the paper reporting the *V. parahaemolyticus* peri-GspL structure [73]. (An example of flexible linkers connecting "upper" and "lower" parts of an assembly which do not obey the same symmetry axes has been reported for the glutamate receptor tetramer [74].) Here we propose that the α-interface is the predominant GspL interface on the cytoplasmic side of the $(GspL•GspE)_6$ assembly (Figure 2-10).

- Other inner membrane proteins, like GspM, are known to interact with GspL [56, 58] and may be incorporated into a linear pre-assembly complex comprising multiple $(GspM•GspL•GspE)_6$ assemblies.
- Given the fact that the hexamer composed of N2E-CTE domains can adopt different arrangements (Chapter 3), alterations in the mutual positions of these domains during ATP hydrolysis by GspE likely conveys motions of the N2E-CTE domains via the N1E-N2E linker to the N1E•cyto-GspL heterodimer. Changes in positions of cyto-GspL domains will alter positions of periplasmic GspL domains, and of other inner membrane platform proteins, thereby promoting assembly of the pseudopilus in the periplasm.
- The arrangement of full-length-GspL dimers as sketched above might only exist transiently in the fully assembled T2SS since it has recently been proposed, on the basis of disulfide cross-linking studies, that the periplasmic domains of full-length peri-GspL are part of the time forming homodimers, and part of the time peri-GspM•peri-GspL heterodimers [75]. Whether or not the α-interface of the GspL dimer would be maintained in the cytoplasm during such proposed rearrangements of GspL domains in the periplasm remains to be determined.

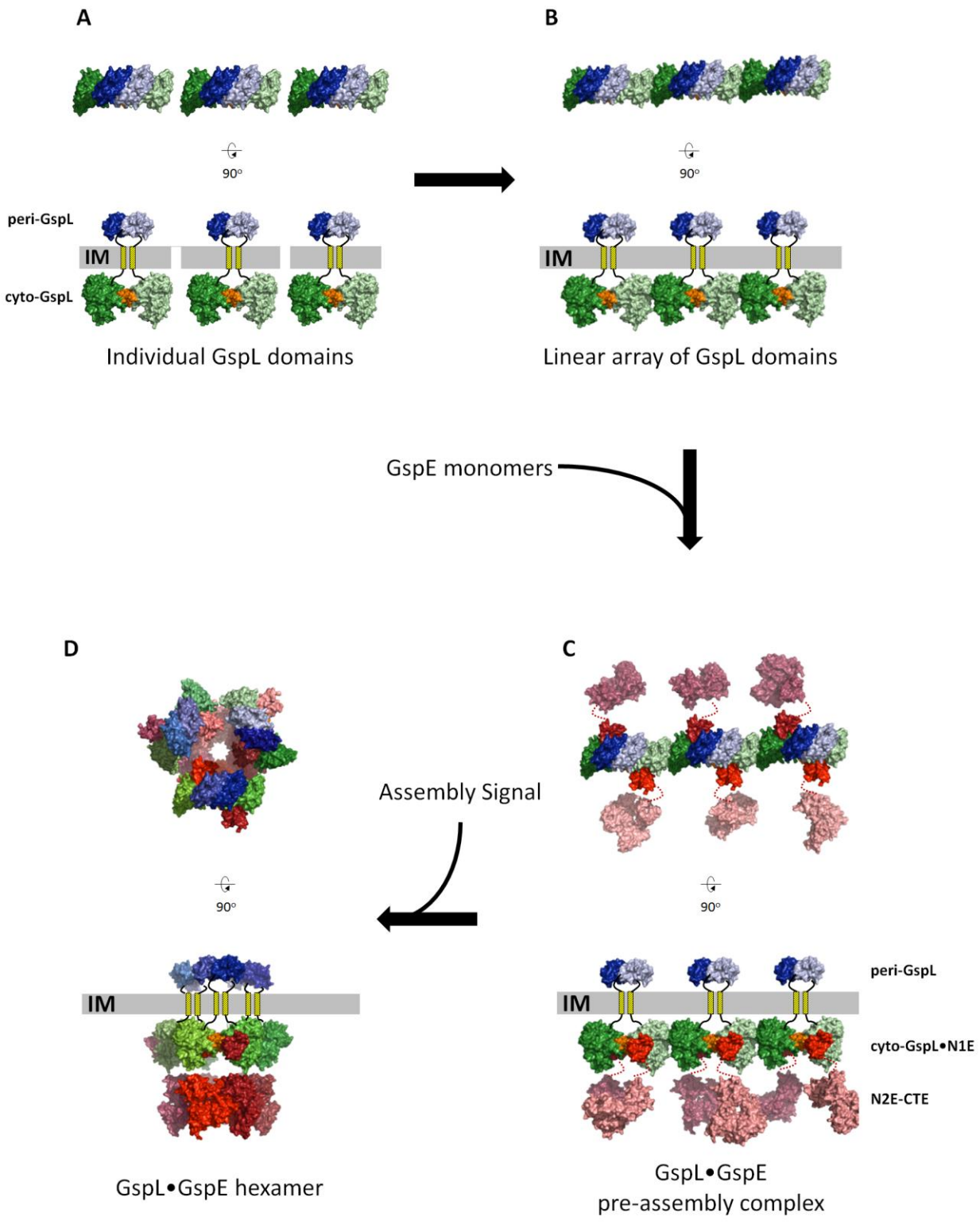

A
B
90°
90°
peri-GspL
IM
cyto-GspL
IM
Individual GspL domains
Linear array of GspL domains
GspE monomers
D
C
Assembly Signal
90°
90°
IM
IM
peri-GspL
cyto-GspL•N1E
N2E-CTE
GspL•GspE hexamer
GspL•GspE
pre-assembly complex

Figure 2-10: A possible pre-assembly complex of the T2SS Inner Membrane Platform. The upper views are perpendicular to the membrane. The lower views are parallel to the membrane. (A) Thre separate dimers of GspL with interactions across the α-interface (orange) of the cytoplasmic domains (green colors) and the p-interface in the periplasmic domains (blue colors). (B) Multiple GspL dimers form linear arrays by β-interfaces amongst cytoplasmic domains. (C) N1E domains of GspE interact with cyto-GspL domains of a linear array, yielding a pre-assembly complex. Additional proteins like GspM (not shown) might also be part of the pre-assembly complex (see text). (D) After an assembly signal, six GspL and six GspE subunits form an assembly with the lower N2E-CTE didomain as a hexamer with approximate C_6 symmetry, connected by N1E-N2E linker residues to the N1E•GspL complex with approximate C_3 symmetry. The latter C_3 axis relates three N1E•GspL dimers with each of these dimers containing an approximate C_2 axis. The aforementioned approximate C_6, C_3 and C_2 axes run parallel to each other, perpendicular to the inner membrane plane, with the C_6 and C_3 axes coinciding. The C_2 axes have a different position, approximately related by the C_3 axis.

Recent electron microscopy studies on a *Thermus thermophilus* T4aPS sub-assembly have proposed the presence of hetero-multimers formed by two copies each of the inner membrane proteins PilM, PilN and PilO, which are the homologs of the T2SS GspL and GspM Inner Membrane Platform proteins [76, 77]. This is in line with the dimers of the cytoplasmic and periplasmic domains of GspL and of GspM observed in crystal structures of these T2SS proteins [26, 59, 73, 78]. However, there are also important differences between inner membrane proteins of the T2SS and T4aPS. Particularly relevant for the current discussion is the fact that the conserved T2SS residues in the α-interface of cyto-GspL are not conserved in PilM (Figure 2-11). In view of such key differences, it seems best to refrain from extending the ideas presented here regarding the biogenesis, composition and symmetry of the T2SS Inner Membrane Complex to that of the T4aPS.

It is clear from the above that the intriguing but fascinating T2SS is revealing many of its secrets in a most reluctant manner. Further structural and biochemical studies are obviously required to confirm current hypotheses and to add new information.

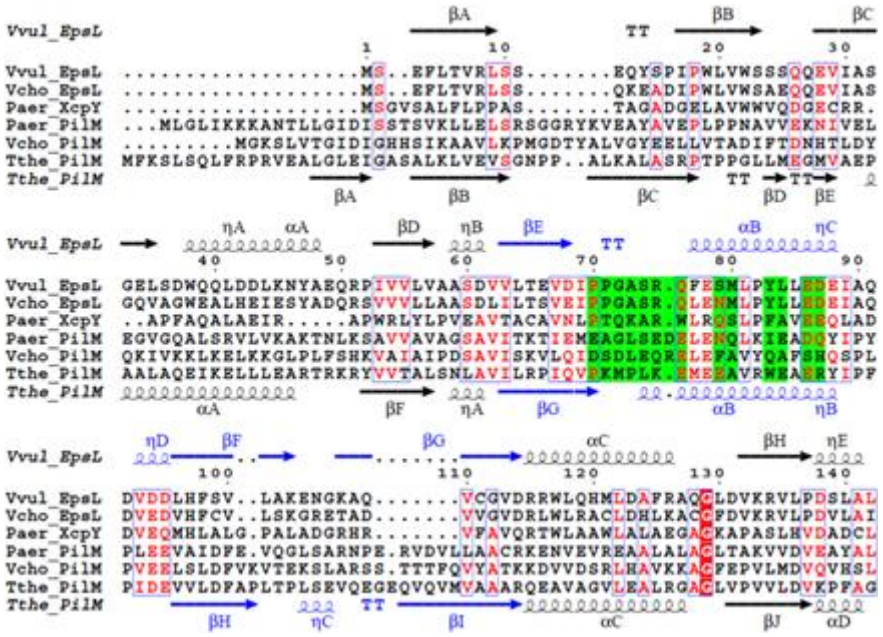


Figure 2-11: Representative aligned sequences from T2SS cyto-GspL and T4aPS PilM subdomains I and II. Shown are sequences of T2SS cyto-GspL from *Vibrio vulnificus, Vibrio cholerae* and *Pseudomonas aeruginosa,* and T4aPS PilM from *Vibrio cholerae, Pseudomonas aeruginosa* and *Thermus thermophilus.* The secondary structure elements of *Vibrio vulnificus* cyto-GspL and *Thermus thermophilus* PilM are annotated at the top and bottom, respectively, with subdomain I in black and subdomain II in blue. The α-interface residues are highlighted in green (bottom).

2.5 REFERENCE

1. Korotkov, K.V., M. Sandkvist, and W.G. Hol, *The type II secretion system: biogenesis, molecular architecture and mechanism.* Nature reviews. Microbiology, 2012. **10**(5): p. 336-51.
2. Sikora, A.E., et al., *Proteomic analysis of the Vibrio cholerae type II secretome reveals new proteins, including three related serine proteases.* The Journal of biological chemistry, 2011. **286**(19): p. 16555-66.
3. Hirst, T.R., et al., *Mechanism of toxin secretion by Vibrio cholerae investigated in strains harboring plasmids that encode heat-labile enterotoxins of Escherichia coli.* Proceedings of the National Academy of Sciences of the United States of America, 1984. **81**(24): p. 7752-6.
4. Merritt, E.A. and W.G. Hol, *AB5 toxins.* Current Opinion in Structural Biology, 1995. **5**(2): p. 165-71.
5. Qadri, F., et al., *Reduction in capsular content and enhanced bacterial susceptibility to serum killing of Vibrio cholerae O139 associated with the 2002 cholera epidemic in Bangladesh.* Infection and immunity, 2005. **73**(10): p. 6577-83.
6. Wenneras, C. and V. Erling, *Prevalence of enterotoxigenic Escherichia coli-associated diarrhoea and carrier state in the developing world.* Journal of health, population, and nutrition, 2004. **22**(4): p. 370-82.
7. Steffen, R., et al., *Vaccination against enterotoxigenic Escherichia coli, a cause of travelers' diarrhea.* Journal of travel medicine, 2005. **12**(2): p. 102-7.
8. Goldwater, P.N. and K.A. Bettelheim, *Treatment of enterohemorrhagic Escherichia coli (EHEC) infection and hemolytic uremic syndrome (HUS).* BMC medicine, 2012. **10**: p. 12.
9. Ho, T.D., et al., *Type 2 secretion promotes enterohemorrhagic Escherichia coli adherence and intestinal colonization.* Infection and immunity, 2008. **76**(5): p. 1858-65.
10. Toshima, H., et al., *Prevalence of enteric bacteria that inhibit growth of enterohaemorrhagic Escherichia coli O157 in humans.* Epidemiology and infection, 2007. **135**(1): p. 110-7.
11. Ochoa, T.J. and C.A. Contreras, *Enteropathogenic escherichia coli infection in children.* Current opinion in infectious diseases, 2011. **24**(5): p. 478-83.
12. Baldi, D.L., et al., *The type II secretion system and its ubiquitous lipoprotein substrate, SslE, are required for biofilm formation and virulence of enteropathogenic Escherichia coli.* Infection and immunity, 2012. **80**(6): p. 2042-52.
13. Jyot, J., et al., *Type II secretion system of Pseudomonas aeruginosa: in vivo evidence of a significant role in death due to lung infection.* The Journal of infectious diseases, 2011. **203**(10): p. 1369-77.
14. Ball, G., et al., *A novel type II secretion system in Pseudomonas aeruginosa.* Molecular microbiology, 2002. **43**(2): p. 475-85.
15. Rossier, O., J. Dao, and N.P. Cianciotto, *The type II secretion system of Legionella pneumophila elaborates two aminopeptidases, as well as a metalloprotease that contributes to differential infection among protozoan hosts.* Applied and environmental microbiology, 2008. **74**(3): p. 753-61.
16. DebRoy, S., et al., *Legionella pneumophila type II secretome reveals unique exoproteins and a chitinase that promotes bacterial persistence in the lung.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(50): p. 19146-51.
17. Nivaskumar, M. and O. Francetic, *Type II secretion system: A magic beanstalk or a protein escalator.* Biochim Biophys Acta, 2014.
18. Cianciotto, N.P., *Type II Secretion and Legionella Virulence.* Current topics in microbiology and immunology, 2013. **376**: p. 81-102.

19. Howard, S.P., *Assembly of the type II secretion system.* Research in Microbiology, 2013. **164**(6): p. 535-544.
20. Douzi, B., A. Filloux, and R. Voulhoux, *On the path to uncover the bacterial type II secretion system.* Philosophical Transactions of the Royal Society B-Biological Sciences, 2012. **367**(1592): p. 1059-1072.
21. McLaughlin, L.S., R.J.F. Haft, and K.T. Forest, *Structural insights into the Type II secretion nanomachine.* Current Opinion in Structural Biology, 2012. **22**(2): p. 208-216.
22. Johnson, T.L., et al., *Type II secretion: from structure to function.* FEMS microbiology letters, 2006. **255**(2): p. 175-86.
23. Filloux, A., *The underlying mechanisms of type II protein secretion.* Biochimica et biophysica acta, 2004. **1694**(1-3): p. 163-79.
24. Py, B., L. Loiseau, and F. Barras, *An inner membrane platform in the type II secretion machinery of Gram-negative bacteria.* EMBO Rep, 2001. **2**(3): p. 244-8.
25. Shiue, S.J., et al., *XpsE oligomerization triggered by ATP binding, not hydrolysis, leads to its association with XpsL.* The EMBO journal, 2006. **25**(7): p. 1426-35.
26. Abendroth, J., et al., *The X-ray structure of the type II secretion system complex formed by the N-terminal domain of EpsE and the cytoplasmic domain of EpsL of Vibrio cholerae.* Journal of molecular biology, 2005. **348**(4): p. 845-55.
27. Sandkvist, M., et al., *Interaction between the autokinase EpsE and EpsL in the cytoplasmic membrane is required for extracellular secretion in Vibrio cholerae.* The EMBO journal, 1995. **14**(8): p. 1664-73.
28. Arts, J., et al., *Interaction domains in the Pseudomonas aeruginosa type II secretory apparatus component XcpS (GspF).* Microbiology, 2007. **153**(Pt 5): p. 1582-92.
29. Korotkov, K.V. and W.G. Hol, *Structure of the GspK-GspI-GspJ complex from the enterotoxigenic Escherichia coli type 2 secretion system.* Nature structural & molecular biology, 2008. **15**(5): p. 462-8.
30. Douzi, B., et al., *Deciphering the Xcp Pseudomonas aeruginosa type II secretion machinery through multiple interactions with substrates.* The Journal of biological chemistry, 2011. **286**(47): p. 40792-801.
31. Yanez, M.E., et al., *Structure of the minor pseudopilin EpsH from the Type 2 secretion system of Vibrio cholerae.* Journal of molecular biology, 2008. **377**(1): p. 91-103.
32. Campos, M., O. Francetic, and M. Nilges, *Modeling pilus structures from sparse data.* Journal of Structural Biology, 2011. **173**(3): p. 436-44.
33. Korotkov, K.V., et al., *Crystal structure of the N-terminal domain of the secretin GspD from ETEC determined with the assistance of a nanobody.* Structure, 2009. **17**(2): p. 255-65.
34. Yanez, M.E., et al., *The crystal structure of a binary complex of two pseudopilins: EpsI and EpsJ from the type 2 secretion system of Vibrio vulnificus.* Journal of molecular biology, 2008. **375**(2): p. 471-86.
35. Kohler, R., et al., *Structure and assembly of the pseudopilin PulG.* Mol Microbiol, 2004. **54**(3): p. 647-64.
36. Reichow, S.L., et al., *Structure of the cholera toxin secretion channel in its closed state.* Nature structural & molecular biology, 2010. **17**(10): p. 1226-32.
37. Chami, M., et al., *Structural insights into the secretin PulD and its trypsin-resistant core.* The Journal of biological chemistry, 2005. **280**(45): p. 37732-41.
38. Chen, Y.-L. and N.-T. Hu, *Function-Related Positioning of the Type II Secretion ATPase of Xanthomonas campestris pv. campestris.* PLoS ONE, 2013. **8**(3).
39. Ayers, M., P.L. Howell, and L.L. Burrows, *Architecture of the type II secretion and type IV pilus machineries.* Future microbiology, 2010. **5**(8): p. 1203-18.
40. Craig, L. and J. Li, *Type IV pilli: paradoxes in form and function.* Current Opinion in Structural Biology, 2008. **18**(2): p. 267-277.

41. Giltner, C.L., Y. Nguyen, and L.L. Burrows, *Type IV pilin proteins: versatile molecular modules.* Microbiology and molecular biology reviews : MMBR, 2012. **76**(4): p. 740-72.
42. Pelicic, V., *Type IV pili: e pluribus unum?* Molecular microbiology, 2008. **68**(4): p. 827-37.
43. Lassak, K., A. Ghosh, and S.V. Albers, *Diversity, assembly and regulation of archaeal type IV pili-like and non-type-IV pili-like surface structures.* Research in Microbiology, 2012. **163**(9-10): p. 630-44.
44. Korotkov, K.V., T. Gonen, and W.G. Hol, *Secretins: dynamic channels for protein transport across membranes.* Trends in biochemical sciences, 2011. **36**(8): p. 433-43.
45. Patrick, M., et al., *Oligomerization of EpsE coordinates residues from multiple subunits to facilitate ATPase activity.* The Journal of biological chemistry, 2011. **286**(12): p. 10378-86.
46. Camberg, J.L., et al., *Synergistic stimulation of EpsE ATP hydrolysis by EpsL and acidic phospholipids.* The EMBO journal, 2007. **26**(1): p. 19-27.
47. Camberg, J.L. and M. Sandkvist, *Molecular analysis of the Vibrio cholerae type II secretion ATPase EpsE.* Journal of bacteriology, 2005. **187**(1): p. 249-56.
48. Lu, C., et al., *Hexamers of the type II secretion ATPase GspE from Vibrio cholerae with increased ATPase activity.* Structure, 2013. **21**(9): p. 1707-17.
49. Robien, M.A., et al., *Crystal structure of the extracellular protein secretion NTPase EpsE of Vibrio cholerae.* Journal of molecular biology, 2003. **333**(3): p. 657-74.
50. Possot, O.M. and A.P. Pugsley, *The conserved tetracysteine motif in the general secretory pathway component PulE is required for efficient pullulanase secretion.* Gene, 1997. **192**(1): p. 45-50.
51. Chen, Y., et al., *Structure and function of the XpsE N-terminal domain, an essential component of the Xanthomonas campestris type II secretion system.* The Journal of biological chemistry, 2005. **280**(51): p. 42356-63.
52. Misic, A.M., K.A. Satyshur, and K.T. Forest, *P. aeruginosa PilT structures with and without nucleotide reveal a dynamic type IV pilus retraction motor.* Journal of Molecular Biology, 2010. **400**(5): p. 1011-21.
53. Satyshur, K.A., et al., *Crystal structures of the pilus retraction motor PilT suggest large domain movements and subunit cooperation drive motility.* Structure, 2007. **15**(3): p. 363-76.
54. Reindl, S., et al., *Insights into FlaI Functions in Archaeal Motor Assembly and Motility from Structures, Conformations, and Genetics.* Molecular Cell, 2013. **49**(6): p. 1069-82.
55. Yamagata, A. and J.A. Tainer, *Hexameric structures of the archaeal secretion ATPase GspE and implications for a universal secretion mechanism.* The EMBO journal, 2007. **26**(3): p. 878-90.
56. Sandkvist, M., et al., *Direct interaction of the EpsL and EpsM proteins of the general secretion apparatus in Vibrio cholerae.* J Bacteriol, 1999. **181**(10): p. 3129-35.
57. Gray, M.D., et al., *In vivo cross-linking of EpsG to EpsL suggests a role for EpsL as an ATPase-pseudopilin coupling protein in the Type II secretion system of Vibrio cholerae.* Molecular microbiology, 2011. **79**(3): p. 786-98.
58. Sandkvist, M., et al., *Two regions of EpsL involved in species-specific protein-protein interactions with EpsE and EpsM of the general secretion pathway in Vibrio cholerae.* Journal of bacteriology, 2000. **182**(3): p. 742-8.
59. Abendroth, J., et al., *The structure of the cytoplasmic domain of EpsL, an inner membrane component of the type II secretion system of Vibrio cholerae: an unusual member of the actin-like ATPase superfamily.* Journal of molecular biology, 2004. **344**(3): p. 619-33.

60. Karuppiah, V. and J.P. Derrick, *Structure of the PilM-PilN inner membrane type IV pilus biogenesis complex from Thermus thermophilus.* J Biol Chem, 2011. **286**(27): p. 24434-42.
61. Yamagata, A., et al., *Structure of an essential type IV pilus biogenesis protein provides insights into pilus and type II secretion systems.* Journal of Molecular Biology, 2012. **419**(1-2): p. 110-24.
62. Kabsch, W., *Xds.* Acta crystallographica. Section D, Biological crystallography, 2010. **66**(Pt 2): p. 125-32.
63. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
64. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallographica Section D-Biological Crystallography, 2004. **60**(Part 12 Special Issue 1): p. 2126-2132.
65. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of Macromolecular Structures by the Maximum-Likelihood Method.* Acta Crystallogr D, 1997. **53**(Part 3): p. 240-255.
66. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions.* Acta Crystallographica Section D Biological Crystallography, 2004. **60**(Part 12 Special Issue 1): p. 2256-2268.
67. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* Journal of molecular biology, 2007. **372**(3): p. 774-97.
68. Gouet, P., X. Robert, and E. Courcelle, *ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins.* Nucleic Acids Res, 2003. **31**(13): p. 3320-3.
69. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta crystallographica. Section D, Biological crystallography, 2010. **66**(Pt 1): p. 12-21.
70. Collins, R.F., et al., *Structure and mechanism of the PilF DNA transformation ATPase from Thermus thermophilus.* The Biochemical journal, 2013. **450**(2): p. 417-25.
71. Possot, O. and A.P. Pugsley, *Molecular characterization of PulE, a protein required for pullulanase secretion.* Molecular microbiology, 1994. **12**(2): p. 287-99.
72. Rivas, S., et al., *TrwD, a protein encoded by the IncW plasmid R388, displays an ATP hydrolase activity essential for bacterial conjugation.* The Journal of biological chemistry, 1997. **272**(41): p. 25583-90.
73. Abendroth, J., A.C. Kreger, and W.G. Hol, *The dimer formed by the periplasmic domain of EpsL from the Type 2 Secretion System of Vibrio parahaemolyticus.* Journal of structural biology, 2009. **168**(2): p. 313-22.
74. Sobolevsky, A.I., M.P. Rosconi, and E. Gouaux, *X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor.* Nature, 2009. **462**(7274): p. 745-56.
75. Lallemand, M., et al., *Dynamic interplay between the periplasmic and transmembrane domains of GspL and GspM in the type II secretion system.* PLoS One, 2013. **8**(11): p. e79562.
76. Sampaleanu, L.M., et al., *Periplasmic domains of Pseudomonas aeruginosa PilN and PilO form a stable heterodimeric complex.* Journal of Molecular Biology, 2009. **394**(1): p. 143-59.
77. Karuppiah, V., et al., *Structure and assembly of an inner membrane platform for initiation of type IV pilus biogenesis.* Proc Natl Acad Sci U S A, 2013. **110**(48): p. E4638-47.
78. Abendroth, J., et al., *The crystal structure of the periplasmic domain of the type II secretion system protein EpsM from Vibrio cholerae: the simplest version of the ferredoxin fold.* Journal of molecular biology, 2004. **338**(3): p. 585-96.

Chapter 3

Hexamers of the Type II Secretion ATPase GspE from *Vibrio cholerae* with Increased ATPase Activity

A modified version of this chapter was previously published in Lu et al., 2013 [1].

SUMMARY

The type II secretion system (T2SS), a multiprotein machinery spanning two membranes in Gram-negative bacteria, is responsible for the secretion of folded proteins from the periplasm across the outer membrane. The critical multidomain T2SS assembly ATPase GspE had so far not been structurally characterized as a hexamer. Here, four hexamers of *Vibrio cholerae* GspE are obtained when fused to Hcp1 as an assistant hexamer, as shown with native mass spectrometry. The enzymatic activity of the GspE-Hcp1 fusions is ~20 times higher than that of a GspE monomer indicating that increasing the local concentration of GspE by the fusion strategy was successful. Crystal structures of GspE-Hcp1 fusions with different linker lengths reveal regular and elongated hexamers of GspE with major differences in domain orientation within subunits, and in subunit assembly. SAXS studies on GspE-Hcp1 fusions suggest that even further variability in GspE hexamer architecture is likely.

3.1 INTRODUCTION

The type II secretion system (T2SS) is a complex, two-membrane-spanning machinery occurring in a wide variety of pathogenic and non-pathogenic Gram-negative bacteria, where it plays a crucial role in the secretion of folded proteins from the periplasm across the outer membrane into the extracellular milieu [2-4]. For instance, the *Vibrio cholerae* T2SS is responsible for secreting the major virulence factor cholera toxin across

the outer membrane in such a manner that the delicate AB_5 heterohexamer architecture of the toxin [5] is maintained without unfolding [6]. Key subassemblies of the T2SS machinery are: an outer membrane channel formed by the secretin GspD; a filamentous pseudopilus with, as its major component, the pseudopilin GspG; a central multiprotein inner membrane complex that interacts with all other subassemblies; and a "secretion ATPase" GspE in the cytosol which interacts with the inner membrane protein GspL. The T2SS GspE is an ATPase that enables formation of the pseudopilus, which likely acts as a piston during exoprotein translocation across the outer membrane.

There are several other important machineries with components that are closely or distantly related to the T2SS, including the type IV pilus systems (T4PS), the filamentous phage assembly system in Gram-negative bacteria, the archaella (also called the archaeal flagella) assembly system in Archaea, and the bacterial transformation system in Gram-positive bacteria [2, 7]. The closest relative of the T2SS is the T4PS, which also spans the inner and outer membranes in Gram-negative bacteria. The T4PS and T2SS share several features, yet there are also major differences [8]. One distinct difference is that the pilus assembled by a T4PS often contains thousands of pilin subunits and extends far into the milieu surrounding the bacterium, while the pseudopilus assembled by a T2SS consists probably of 5-20 subunits, and this pseudopilus remains in the periplasm under physiologic conditions. Moreover, most T4PS machineries have two or even three ATPases including an assembly ATPase and one or more retraction ATPases [8-10]. In contrast, T2SS machineries have one ATPase, GspE. Here we focus on the architecture and activity of the T2SS ATPase and its relatives, in particular from the T4PS.

The ~56 kDa T2SS secretion ATPase GspE belongs to the family of Type II/IV secretion ATPases [11, 12]. Almost all T2SS ATPases have three major domains: two N-terminal domains, N1E and N2E, and a C-terminal domain (CTE), although, in a few species, an additional N-terminal N0 domain occurs [13]. The N1E and N2E are connected by a linker of ~20 residues, and the N2E by a ~15 residue linker to the CTE. The N1E interacts with the

cytoplasmic domain of the inner membrane protein GspL [14-16]. The CTE consists of three subdomains: the major nucleotide binding domain (C1E), a four-helical domain (C2E), and a metal-binding domain (CME) (Figure 3-1). The metal binding domain occurs in almost all T2SS ATPases from Gram-negative bacteria and contains an essential tetra-cysteine motif which coordinates zinc ions [17, 18]. While the assembly ATPase of the T4PS that is required for polymerization of pilins into pili has a similar domain organization as the T2SS ATPase, the retraction ATPase that promotes depolymerization of pili lacks the N1E and CME (Figure 3-1).

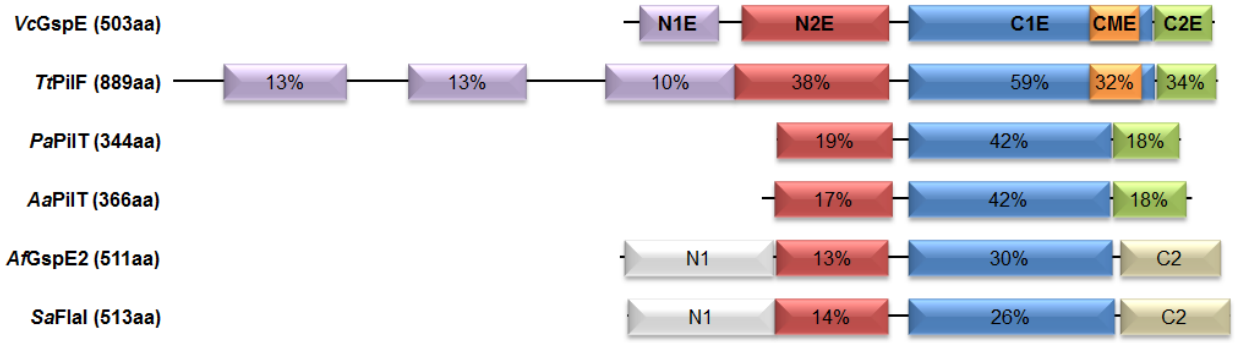


Figure 3-1: Domain bar diagram of GspE and homologous ATPases with electron microscopy reconstructions or crystal structures. *Vc*GspE: GspE*,* the assembly ATPase from the T2SS in the Gram-negative *Vibrio cholerae. Tt*PilF: PilF, an assembly ATPase from the T4PS in the Gram-negative eubacterium *Thermus thermophilus. Pa*PilT: PilT, a retraction ATPase from the T4PS in the Gram-negative eubacterium *Pseudomonas aeruginosa. Aa*PilT: PilT, a retraction ATPase from the T4PS in the Gram-negative eubacterium *Aquifex aeolicus; Af*GspE2: one of three related ATPases in the Archaeon *Archaeoglobus fulgidus*; *Sa*FlaI: FlaI, an ATPase from the archaellum assembly system in the Archaeon *Sulfolobus acidocaldarius.* For domains that are homologous in structures to GspE, the percentage sequence identity is given compared to *Vc*GspE.

Many non-T2SS GspE homologs, such as the retraction ATPase PilT and archaeal flagellar ATPases, form hexamers with various point group symmetries [19-23]. In contrast, obtaining T2SS ATPases as homogeneous hexamers in solution, or as hexamers in crystals, has not been possible so far. The crystal structure of a truncated form of *Vibrio cholerae* GspE, also called EpsE [15], has been elucidated. This variant of GspE lacked the N1E and crystallized in a helical arrangement with 6_1 symmetry [17]. Solution studies indicate hexamer formation of the GspE from the *Xanthomonas campestris* T2SS upon the addition of nucleotides [24]. Activity measurements of *V. cholerae* GspE showed that oligomers of an approximately hexameric size according to gel filtration analysis possess much higher activity than monomers, but hexamers are only present in small amounts [25]. Full-length *V. cholerae* GspE forms oligomers, possibly hexamers, in the presence of the cytoplasmic domain of GspL and acidic phospholipids, specifically cardiolipin, with a concomitant considerable increase in ATPase activity [18]. Mutations interrupting putative intersubunit interfaces of *V. cholerae* GspE result in activity loss, indicating that a precise interaction between adjoining subunits is essential for the function of GspE [21]. High-resolution structures of retraction T4PS ATPase hexamers are known [20, 22] and recently an electron microscopy reconstruction of a hexamer of the T4PS assembly ATPase PilF from *Thermus thermophilus*, an extremophile eubacterium, has been reported [26]. So far, however, no hexamers of T2SS ATPases have been characterized structurally.

To overcome the reluctance of GspE to form hexamers, we embarked on a strategy of fusing GspE to another protein known to form hexamers by itself. The rationale behind this approach is to increase the local concentration of GspE subunits, thereby increasing the probability of GspE hexamer formation. Here we selected Hcp1 [27] as a fusion partner for *V. cholerae* $^{\Delta N1}$GspE, and observed that the fusion protein forms hexamers in solution with a concomitant ~20-fold increase in ATPase activity with respect to the monomeric enzyme. The application of the assistant hexamer strategy was clearly successful.

In crystallographic studies of the GspE-Hcp1 fusion proteins, *V. cholerae* $^{\Delta N1}$GspE hexameric assemblies with two different symmetries, and distinctly different subunit conformations were observed. Small angle X-ray scattering (SAXS) studies of $^{\Delta N1}$GspE-Hcp1 fusions, when combined with the two $^{\Delta N1}$GspE-Hcp1 hexamer crystal structures, indicate that additional types of GspE hexamers are likely. These results show considerable interdomain flexibility, consistent with large dynamic motions expected to occur in the T2SS while in action. Comparison of the two structures with related ATPases indicates that a construction unit consisting of two domains from adjacent subunits is a hallmark of this large family of secretion ATPases, as also reported in a recent study of the archaellum assembly ATPase FlaI [23]. The global general principle of these ATPases appears to be that construction units are largely fixed and subunits are highly flexible.

3.2 METHODS

3.2.1 Design of $^{\Delta N1}$GspE-Hcp1 Fusions

P. aeruginosa Hcp1 (PDB ID: 1Y12; [27]), a 165-residue single domain protein that forms stable hexameric rings in solution and crystal structure, was chosen as the assistant hexamer. A hexameric ring model with C_6 symmetry of *V. cholerae* GspE without its N-terminal 99 residues ($^{\Delta N1}$GspE) was generated based on the C_6 hexameric structure of *Aquifex aeolicus* PilT bound to ATP (*Aa*PilT; PDB ID: 2EWW) [20], which shares 25% amino acid sequence identity with $^{\Delta N1}$GspE. The CTEs of $^{\Delta 90}$GspE were superimposed onto the C-domains of the AaPilT hexamer, and N2Es were placed in positions to maintain the most extensive and conserved CTE•N2E' interface observed in the *V. cholerae* $^{\Delta 90}$GspE helical structure (PDB ID: 1P9W) [17]. By aligning the six fold axes of the $^{\Delta 90}$GspE hexameric model with that of the Hcp1 hexamer, the two hexamers were placed close to each other without clashes. An approximate distance of 28 Å was obtained between the last observed amino acid at the C-terminus in the $^{\Delta 90}$GspE structure and the first observed amino acid at the N-terminus of the Hcp1 structure. The five C-terminal residues of GspE, that were not

included in the construct of the $^{\Delta 90}$GspE structure [17], where maintained in the current constructs. Several variants of $^{\Delta N1}$GspE-Hcp1 fusions with different linker sequences were constructed and verified by DNA sequencing. The constructs are labeled as $^{\Delta N1}$GspE-naa-Hcp1, which consists of *V. cholerae* GspE residues 100-503, a linker of 'n' amino acids, and the full-length Hcp1 sequence. Two "6aa" variants were made to check the effect of amino acids in the linker on ATPase activity: $^{\Delta N1}$GspE-6aa(KLASGA)-Hcp1 and $^{\Delta N1}$GspE-6aa(GSGSGS)-Hcp1. The activities were very similar (Figure 3-3). The crystal structure was determined of $^{\Delta N1}$GspE-6aa(GSGSGS)-Hcp1, which is for convenience called $^{\Delta N1}$GspE-6aa-Hcp1 throughout this chapter.

3.2.2 Purification and Characterization

Four variants of $^{\Delta N1}$GspE -linker-Hcp1, each containing a C-terminal hexa-histidine tag, were cloned into a modified pETDuet vector. Protein was expressed at 12°C in BL21(DE3), by induction with 0.1 mM IPTG at OD_{600}~0.8. Cells were harvested at approximately 18 hours and resuspended in purification buffer consisting of 20 mM Tris-HCl pH 8, 500 mM NaCl, 5% (v/v) glycerol, 1 mM TCEP/HCl, 0.1 mM ATP, 5 mM $MgCl_2$ supplemented with 50 mM imidazole and a protease inhibitor cocktail. All the purification procedures were carried out at 4 °C. The cells were disrupted by sonication. The soluble fraction was loaded onto Ni-NTA resin (Qiagen), washed with 75 mM imidazole added to the purification buffer, and eluted in the purification buffer with 250 mM imidazole. Proteins were then purified by Superose 6 gel-permeation chromatography (Amersham). The final protein solution was concentrated to approximately 5 mg/ml supplemented with 5 mM ADP and 5 mM $MgCl_2$.

3.2.3 ATPase Activity Measurements

ATPase activities were measured using BIOMOL Green reagent (Enzo Life Sciences). The reaction mixture contained 0.25 μM monomeric protein, 5 mM ATP, and 5 mM $MgCl_2$, in

100 mM HEPES, pH 8.5, 65 mM NaCl, 5% glycerol. Reactions were incubated at 37 °C for 0, 5, 10, 15, and 20 minutes and assayed for the release of inorganic phosphate. The amount of phosphate was determined by comparing the absorbance at 650 nm with a phosphate standard curve. Data reported were from three separate samples of the same purified proteins assayed in duplicate.

3.2.4 Native Mass Spectrometry

All native mass spectra were acquired using a hybrid electrospray/quadrupole/ion-mobility/time-of-flight mass spectrometer (Waters Synapt G2 HDMS). For each protein, a ~20 µL sample solution was prepared by exchanging 5 µL of 2.5-6.4 mg/mL of the given protein in a storage buffer into a buffer containing 500 mM ammonium acetate buffer (pH 8.0), 50 µM $MgCl_2$, 50 µM ADP using a Corning® Spin-X® UF centrifugal concentrator with a 10K MWCO. Ions were formed by nanospray using borosilicate capillaries that have inner diameters of 0.78 mm and a tip at one end with an inner diameter ~1–3 µm. The capillary was loaded with 2-4 µL of the sample solution and a potential of ~1 kV was applied using a platinum wire electrode inserted into the back of the capillary that makes direct contact with the solution. Mass spectra were calibrated externally using electrospray generated ions from a 50 mg/mL solution of CsI.

3.2.5 Crystallization and Data Collection

$^{\Delta N1}$GspE-5aa-Hcp1 crystals were grown in sitting drop vapor diffusion experiments at 4 ^{o}C using 1 µL protein solution in the presence of 5 mM AMPPNP and 5 mM $MgCl_2$, and 1 µL reservoir solution of 12.5% PEG 20,000, 0.1 M Bicine pH 9, and 2% v/v 1,4-dioxane. Crystals were cryo-protected with the above precipitant and nucleotides supplemented with 25% glycerol. Data were collected at the SSRL beamline 12-1. $^{\Delta N1}$GspE-5aa-Hcp1 data were processed to 7.3 Å in space group $P4_22_12$, with cell dimensions a = b = 205.1 Å and c = 234.7 Å, using HKL2000 [28].

$^{\Delta N1}$GspE-6aa-Hcp1 crystals were grown in sitting drops at 4 ^{o}C using 1 µL protein solution in the presence of 5 mM ADP, 5 mM AMPPNP and 5 mM $MgCl_2$, and 1 µL reservoir solution of 7% PEG6000 and 0.1 M Bicine pH 9. Crystals were cryo-protected with the above precipitant and nucleotides supplemented with 20% glycerol. Data were collected in-house using a Saturn 94 CCD detector on a Rigaku Micromax HF-7 rotating anode. The Zn K-edge data were collected at SSRL beamline 12-1. Native $^{\Delta N1}$GspE-6aa-Hcp1 data were processed to 7.6 Å in space group $P4_22_12$, with cell dimensions a=b= 205.1 Å and c= 235.0 Å, using HKL2000. The Zn-edge $^{\Delta N1}$GspE-6aa-Hcp1 data were isomorphous and was processed using XDS [29]. The solvent content of these tetragonal crystals is ~61.9 %.

$^{\Delta N1}$GspE-7aa-Hcp1 crystals were grown in sitting drops at 4 ^{o}C using 1 µL protein solution and 1 µL reservoir solution of 7% PEG3350, 0.12 M ammonium citrate pH 7.0, 5 mM ADP, 5 mM $MgCl_2$, and cryoprotected by the addition of 25% glycerol. Data were collected in-house using a Saturn 94 CCD detector on a Rigaku Micromax HF-7 generator. The $^{\Delta N1}$GspE-7aa-Hcp1 data were processed to 6.95 Å in space group $P22_12_1$ with cell dimension of a= 106.7, b=132.4, c=149.7 Å, using HKL2000.

$^{\Delta N1}$GspE-8aa-Hcp1 crystals were grown in sitting drops at 4 ^{o}C using 1 µL protein solution and 1 µL reservoir solution of 16% PEG300, 0.2 M ammonium sulfate, 0.1 M BisTris pH 6.1, 5 mM ADP, 5 mM $MgCl_2$, 5 mM $AlCl_3$, and 15 mM NaF. Crystals were cryo-protected with the above precipitant with 30% PEG300. Data was collected at the SSRL beamline 12-1. The $^{\Delta N1}$GspE-8aa-Hcp1 data were processed in space group $P22_12_1$ with cell dimension of a = 112.5, b = 132.9, c = 142.8 Å, using XDS. The $^{\Delta N1}$GspE-8aa-Hcp1 data were highly anisotropic, extending to ~4.1 Å in the a and b directions and to only ~5 Å in the c direction. The solvent content of these orthorhombic crystals is ~55.6 %.

3.2.6 Structure Determinations

The structure of $^{\Delta N1}$GspE-5aa-Hcp1 was solved by molecular replacement with Phaser [30] using the CTE•N2E' construction unit from a full length structure of GspE (chapter 2), i.e. employing the same procedure as used in the structure determination of the unliganded PaPilT structure [22]. Six CTE•N2E' construction units and one Hcp1 hexamer were placed in the asymmetric unit. TLS and rigid body refinements in Phaser were performed by defining each N2E and CTE as an independent group, and the Hcp1 hexamer as the thirteenth group.

The structure of $^{\Delta N1}$GspE-6aa-Hcp1 was solved by molecular replacement with Phaser using the procedure as used for the $^{\Delta N1}$GspE-5aa-Hcp1 structure. Phaser placed six CTE•N2E' construction units in the asymmetric unit of the $^{\Delta N1}$GspE-6aa-Hcp1 $P4_22_12$ crystal. The resultant $^{\Delta N1}$GspE-5aa-Hcp1 and $^{\Delta N1}$GspE-6aa-Hcp1 structures are essentially the same. The structure of $^{\Delta N1}$GspE-8aa-Hcp1 was solved using Phaser and the same search model as above. Three CTE•N2E' construction units were placed in the asymmetric unit of the $^{\Delta N1}$GspE-8aa-Hcp1 $P22_12_1$ crystal, with a crystallographic two fold generating a hexameric ring of $^{\Delta N1}$GspE subunits. Subsequently, Phaser was able to position the Hcp1 hexamer (PDB: 1Y12). The $^{\Delta N1}$GspE-8aa-Hcp1 structure was then used as a model to obtain the structure of $^{\Delta N1}$GspE-7aa-Hcp1 in the same $P22_12_1$ space group yielding very similar arrangements of the $^{\Delta N1}$GspE and Hcp1 hexamers, but at lower resolution.

For the $^{\Delta N1}$GspE-6aa-Hcp1 and $^{\Delta N1}$GspE-8aa-Hcp1 structures further adjustments to the models were performed with REFMAC [31], initially with each hexameric ring defined as a rigid body. Final rigid body and TLS refinements were carried out for the N2E and CTE of each independent $^{\Delta N1}$GspE subunit and the Hcp1 hexamer. Since the $^{\Delta N1}$GspE-5aa-Hcp1 structure is very similar to $^{\Delta N1}$GspE-6aa-Hcp1 structure and the $^{\Delta N1}$GspE-7aa-Hcp1 variant is very similar to the $^{\Delta N1}$GspE-8aa-Hcp1 structure but with lower resolution, we focus here on the $^{\Delta N1}$GspE-6aa-Hcp1 and $^{\Delta N1}$GspE-8aa-Hcp1 structure in the structure analysis and description.

3.2.7 SAXS

SAXS data for $^{\Delta N1}$GspE-6aa-Hcp1 and $^{\Delta N1}$GspE-8aa-Hcp1 were collected at the beamline 4-2 at SSRL using a Rayonix MX225-HE detector. $^{\Delta N1}$GspE-6aa-Hcp1 and $^{\Delta N1}$GspE-8aa-Hcp1 were purified in the presence of 5 mM nucleotide analogs and 5 mM $MgCl_2$ in 20 mM Tris-HCl pH 8, 500 mM NaCl, 5% (v/v) glycerol, 1 mM TCEP/HCl, at 4 °C, and then centrifuged to remove the aggregated material. 35 µL of purified samples from the size exclusion chromatographic separation, with corresponding matching buffers, were automatically loaded into the capillary flow cell with a Hamilton syringe robot. Measurements were taken at room temperature with a sample-to-detector distance of 1700 mm and an X-ray energy of 11 keV. The SAXS data were measured at three protein concentrations for each sample (data not shown) and merged together using PRIMUS [32]. Scattering curves were calculated from the $^{\Delta N1}$GspE-8aa-Hcp1 and $^{\Delta N1}$GspE-6aa-Hcp1 crystal structures, and compared with the experimental scattering curves using FOXS [33]. The chi-square value was used to compare calculated and experimental SAXS curves [34].

3.3 RESULTS

3.3.1 Characterization of GspE-Hcp1 Fusions

Eight fusion proteins were successfully expressed, purified and characterized, and four were crystallized: $^{\Delta N1}$GspE-KLASG-Hcp1, $^{\Delta N1}$GspE-GSGSGS-Hcp1, $^{\Delta N1}$GspE-KLASGAG-Hcp1 and $^{\Delta N1}$GspE-KLASGAGH-Hcp1, called hereafter $^{\Delta N1}$GspE-5aa-Hcp1, $^{\Delta N1}$GspE-6aa-Hcp1, $^{\Delta N1}$GspE-7aa-Hcp1 and $^{\Delta N1}$GspE-8aa-Hcp1, respectively. Homogeneous hexamer formation of these four *V. cholerae* $^{\Delta N1}$GspE-linker-Hcp1 fusions was determined by native mass spectrometry (Figure 3-2). The expected and observed molecular masses differed by less than ~1 kDa, which is in the expected range of mass differences for such ~385 kDa complexes. In the case of the five amino acid linker, the mass spectrometry data also provided evidence for the presence of pentamers. The hexamerization greatly enhanced the ATPase activity of these four $^{\Delta N1}$GspE-linker-Hcp1 variants compared to monomeric $^{\Delta N1}$GspE; in all cases, an increased activity by a factor of approximately 20 was observed (Figure 3-

3). This increase in activity might be an underestimate because there is some tendency for the control $^{\Delta N1}$GspE protein to dimerize according to dynamic light scattering and size exclusion experiments, but not according to native mass spectrometry (data not shown). In addition, the activity of the oligomeric forms was 2-4 fold greater than the cardiolipin-stimulated activity of full- length GspE when in complex with the cytoplasmic domain of GspL [18, 21].

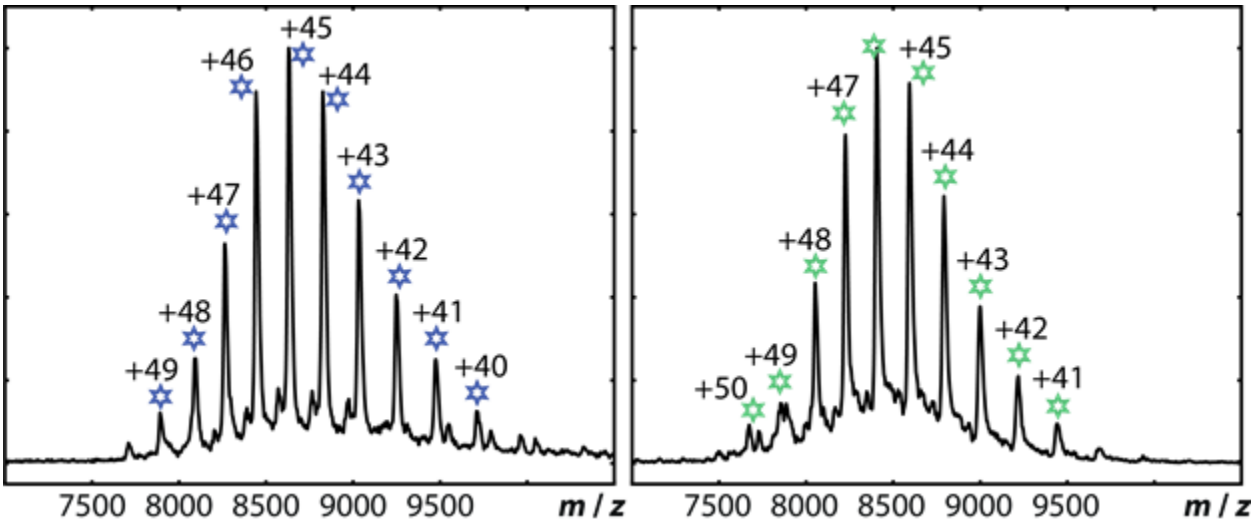


Figure 3-2: Native mass spectra of $^{\Delta N1}$GspE-Hcp1 fusion proteins. Left: $^{\Delta N1}$GspE-8aa-Hcp1. Right: $^{\Delta N1}$GspE-6aa-Hcp1. The data show that these proteins each assemble as hexamers in solution. The mass measured for each complex is only slightly greater than that expected assuming that each protein subunit in the hexamer contains one Zn and one nucleotide.

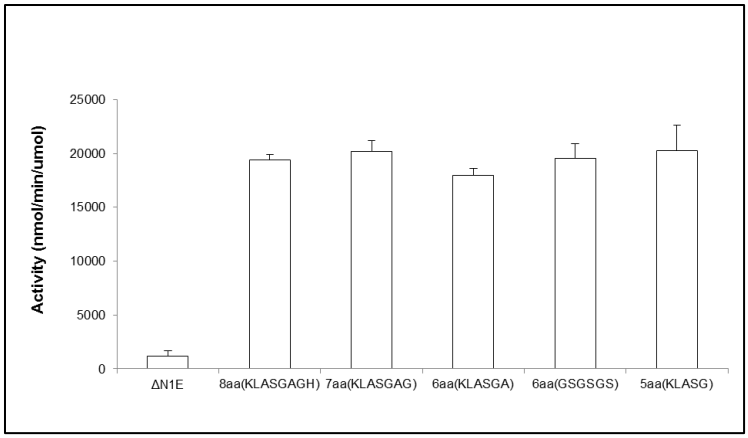


Figure 3-3: ATPase activities of *V. cholerae* $^{\Delta N1}$GspE-linker-Hcp1 variants. From left to right, the activities for $^{\Delta N1}$GspE (control monomer), $^{\Delta N1}$GspE-8aa-Hcp1, $^{\Delta N1}$GspE-7aa-Hcp1, $^{\Delta N1}$GspE-6aa(KLASGA)-Hcp1, $^{\Delta N1}$GspE-6aa(GSGSGS)-Hcp1 and $^{\Delta N1}$GspE-5aa-Hcp1. Two variant linkers were tested in the case of $^{\Delta N1}$GspE-6aa-Hcp1 fusions to evaluate the effect of linker sequence. As shown, both types of linkers gave the same increase in activity. The crystal structure was determined of $^{\Delta N1}$GspE-6aa(GSGSGS)-Hcp1, which is called $^{\Delta N1}$GspE-6aa-Hcp1 throughout this chapter. Error bars represent 1 SD.

Table 1. Crystallographic data collection and refinement[1]

	$^{\Delta N1}$GspE-6aa-Hcp1	$^{\Delta N1}$GspE-8aa-Hcp1
Data Collection		
Space Group	P 4_2 2_1 2	P 2 2_1 2_1
Unit cell dimensions a, b, c (Å)	205.1, 205.1, 235.0	112.5, 132.9, 142.8
Mosaicity (°)	0.53	0.23
Resolution range (Å)	60.0-7.6 (7.87-7.60)	38.7-4.09 (4.31-4.09)
Total number of reflections	45,135	111,225
Unique reflections	6,587	17,346
Average redundancy	6.9 (7.0)	6.4 (6.1)
Completeness (%)	99.6 (99.9)	99.2 (97.5)
R_{merge}	0.25 (>1)	0.083 (0.959)
$\langle I/\sigma(I) \rangle$	10.9 (2.0)	7.1 (0.8)
Refinement		
Resolution (Å)	50.0-7.6 (7.9-7.6)	38.7-4.2 (4.38-4.20)
Number of reflections	6265 (432)	15196 (1014)
R_{work}/R_{free}	0.349/0.360	0.384/0.376
Number of protein chains per asymmetric unit	Six $^{\Delta N1}$GspE-6aa-Hcp1 chains (383 kDa)	Three $^{\Delta N1}$GspE-8aa-Hcp1 chains (193 kDa)

Numbers in parentheses refer to outer resolution shell

3.3.2 Crystal Structure of Two GspE Hexamers

Crystals of all four *V. cholerae* $^{\Delta N1}$GspE-linker-Hcp1 variants were obtained and the structures could be solved by molecular replacement (see Methods and Table 1). The tetragonal crystals of $^{\Delta N1}$GspE-5aa-Hcp1 and $^{\Delta N1}$GspE-6aa-Hcp1 exhibited space group $P4_22_12$ with essentially the same cell dimensions and diffracted to a resolution of ~7.6 Å. The asymmetric unit contains one Hcp1 hexamer and a quite regular arrangement of six $^{\Delta N1}$GspE subunits with quasi C_6 point group symmetry (Figure 3-4), hereafter called the qC_6 hexamer. The 7aa- and 8aa-linked fusions crystallized in space group $P22_12_1$ with similar cell dimensions. The crystals of $^{\Delta N1}$GspE-7aa-Hcp1 protein reached a resolution of ~7 Å, while those of $^{\Delta N1}$GspE-8aa-Hcp1 diffracted highly anisotropically with a resolution of ~5 Å along the a axis and ~4.1 Å in the other two directions. The asymmetric unit contains three $^{\Delta N1}$GspE-8aa-Hcp1 subunits, which are each different as described below. A crystallographic two-fold axis creates an elongated hexamer with C_2 point group symmetry from the three crystallographically independent $^{\Delta N1}$GspE subunits (Figure 3-4), hereafter called the C_2 hexamer. Below, we describe only the $^{\Delta N1}$GspE-6aa-Hcp1 and $^{\Delta N1}$GspE-8aa-Hcp1 structures, because these are essentially the same as those of, respectively, the 5aa- and 7aa-linked proteins.

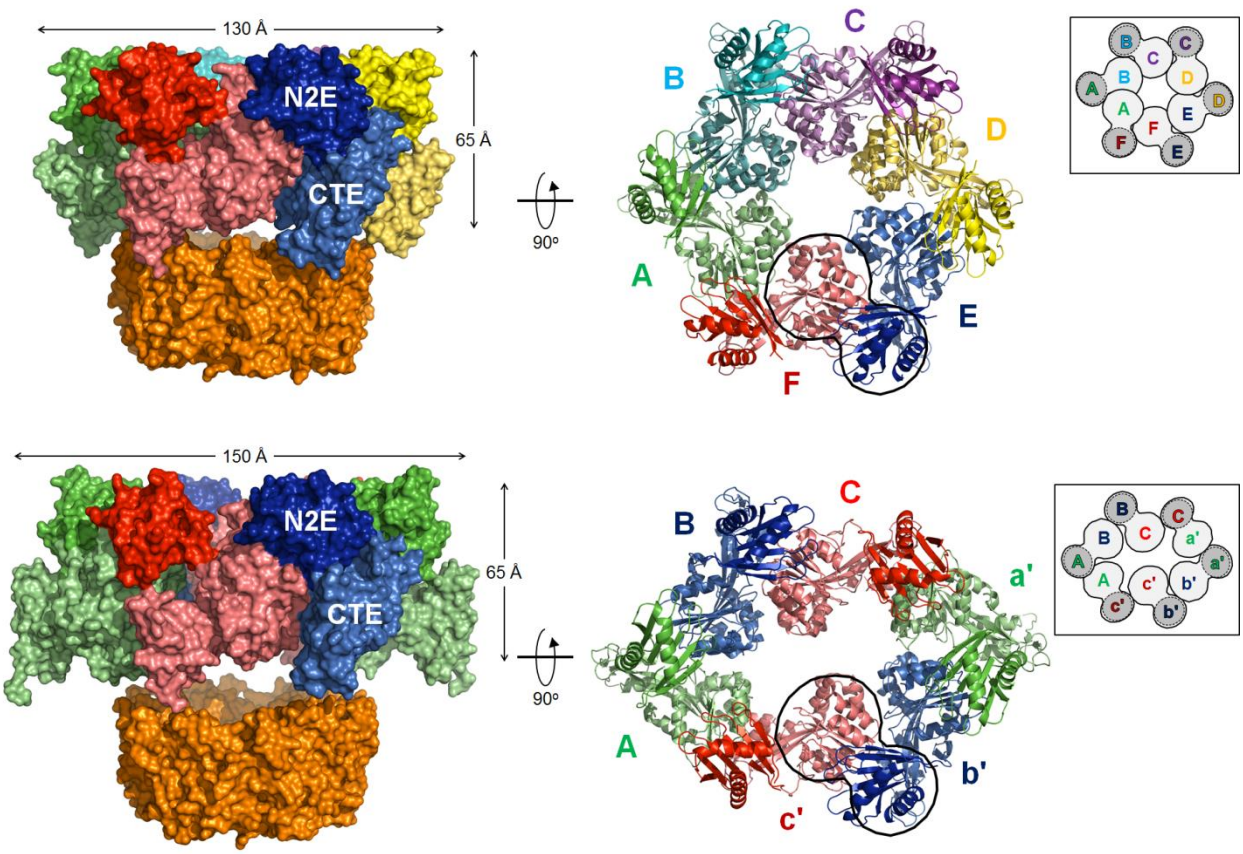


Figure 3-4: Crystal structures of hexameric $^{\Delta N1}$GspE from *Vibrio cholerae* fused to Hcp1. The CTEs are shown in a lighter shade of the same color as the N2Es of the same subunit. The Hcp1 assistant hexamer is shown in orange. One CTE•N2E' construction unit is outlined. Insets: schematic view of the hexamer outlining all six CTE•N2E' construction units. CTE in light grey and N2E' in dark grey. Upper: the structure of *V. cholerae* $^{\Delta N1}$GspE-6aa-Hcp1. This fusion forms a $^{\Delta N1}$GspE hexamer with quasi C_6 point group symmetry. Shown are subunits A (green), B (cyan), C (purple), D (yellow), E (blue) and F (red). The shape of the hexamer in this view is very regular. Upper left: view perpendicular to the quasi six-fold depicting the Hcp1 hexamer. Upper right: view along the quasi six-fold axis of the $^{\Delta N1}$GspE hexamer, with the Hcp1 hexamer omitted. Lower: the structure of *V. cholerae* $^{\Delta N1}$GspE-8aa-Hcp1. This fusion forms a $^{\Delta N1}$GspE hexamer with C_2 point group symmetry. Shown are subunits A (green), B (blue), C (red) – each occurring twice in the hexamer. The shape of the hexamer in this view is an approximate ellipsoid of 105 Å by 150 Å. Lower left: view

perpendicular to the two-fold depicting also the Hcp1 hexamer. Lower right: view along the two-fold axis of the $^{\Delta N1}$GspE hexamer, with the Hcp1 hexamer omitted.

Table 2. Pairwise comparisons of the N2E-*vs*-CTE orientation in *V. cholerae* GspE, *Aa*PilT and *Pa*PilT subunits.

			*Vc*GspE				*Aa*PilT				*Pa*PilT		
			qC_6	C_2			C_6	qC_2			C_2		
				A	B	C		D	E	F	A	B	C
Vc GspE	qC_6			34.0	16.0	40.9	40.2	43.0	43.0	42.4	35.3	24.6	23.1
	C_2	A	34.0		31.5	48.4	26.0	38.8	29.4	35.6	20.7	21.5	18.8
		B	16.0	31.5		46.7	37.9	47.2	35.2	44.4	37.7	26.8	19.2
		C	40.9	48.4	46.7		27.6	14.7	73.4	17.2	30.0	30.2	38.0
Aa PilT	C_6		40.2	26.0	37.9	27.6		16.8	51.8	13.6	11.8	15.3	19.9
	qC_2	D	43.0	38.8	47.2	14.7	16.8		66.4	3.9	20.3	23.7	31.9
		E	43.0	29.4	35.2	73.4	51.8	66.4		63.2	48.3	44.0	36.1
		F	42.4	35.6	44.4	17.2	13.6	3.9	63.2		16.5	20.9	29.1
Pa PilT	C_2	A	35.3	20.7	37.7	30.0	11.8	20.3	48.3	16.5		11.0	17.7
		B	24.6	21.5	26.8	30.2	15.3	23.7	44.0	20.9	11.0		9.2
		C	23.1	18.8	19.2	38.0	19.9	31.9	36.1	29.1	17.7	9.2	

Each pairwise comparison of two subunits is based on a superposition of two CTEs. The resultant superposition operation is applied to the entire subunit. Subsequently the two N2Es are superimposed. The rotation angle of this second superposition is the difference in N2E-*vs*-CTE orientation given in the Table. *Aa*PilT with C_6 hexamer symmetry (PDB: 2EWV [20]); *Aa*PilT with quasi C_2 (qC_2) hexamer symmetry (PDB: 2GSZ [20]); *Pa*PilT with C_2 hexamer symmetry (PDB: 3JVV [22]).

Comparing the two $^{\Delta N1}$GspE hexamers reveals striking differences such as very different shapes of the outside envelopes and of the central cavities (Figure 3-4). The regular qC_6 hexamer has a cylindrical shape with an exterior diameter of ~130 Å and an inner diameter of ~25 Å. The C_2 hexamer, in contrast, has an elongated appearance of ~105 by 150 Å. The inner diameter of the C_2 hexamer is quite irregular, measuring ~14 Å at the narrowest point and ~50 Å at the widest point. The six C1Es in the C_2 hexamer are less tightly packed together than in the qC_6 hexamer. In both hexamers, the Hcp1 rings are positioned at the expected position, near the C-termini of the $^{\Delta N1}$GspE rings. While the six-fold axis of the Hcp1 hexamer in the $^{\Delta N1}$GspE-8aa-Hcp1 structure coincides with the crystallographic two-fold axis that is shared with the two-fold axis of the C_2 hexamer, in the $^{\Delta N1}$GspE-6aa-Hcp1 structure the six-fold axis of the Hcp1 hexamer does not align perfectly with the quasi six-fold axis of the qC_6 hexamer (Figure 3-4).

Including the $^{\Delta 90}$GspE structure [17], we can now compare $^{\Delta N1}$GspE subunits in three different crystal forms. In the $^{\Delta N1}$GspE-6aa-Hcp1 structure with $^{\Delta N1}$GspE hexamers with qC_6 symmetry, the N2E-*vs*-CTE domain orientation within one subunit is essentially the same for all six subunits, differing by only 2.0 to 5.2 degrees. In contrast, in the $^{\Delta N1}$GspE-8aa-Hcp1 structure with $^{\Delta N1}$GspE hexamers with C_2 symmetry, the N2E-*vs*-CTE orientation is highly variable, differing by 32 to 48 degrees when compared pairwise (Table 2). The N2E-*vs*-CTE orientation observed in the three independent subunits in the C_2 hexamer deviates by 16 to 41 degrees from the average orientation in the qC_6 hexamer (Figure 3-5). Interestingly, the N2E-*vs*-CTE orientations in subunit C of the C_2 hexamer and the subunit in the helical structure [17] differ by only 2 degrees and hence are essentially the same (Figure 3-5). Therefore, the three available $^{\Delta N1}$GspE crystal structures, with ten (i.e. 6+3+1) independent views of $^{\Delta N1}$GspE subunits, reveal essentially four different N2E-*vs*-CTE domain orientations.

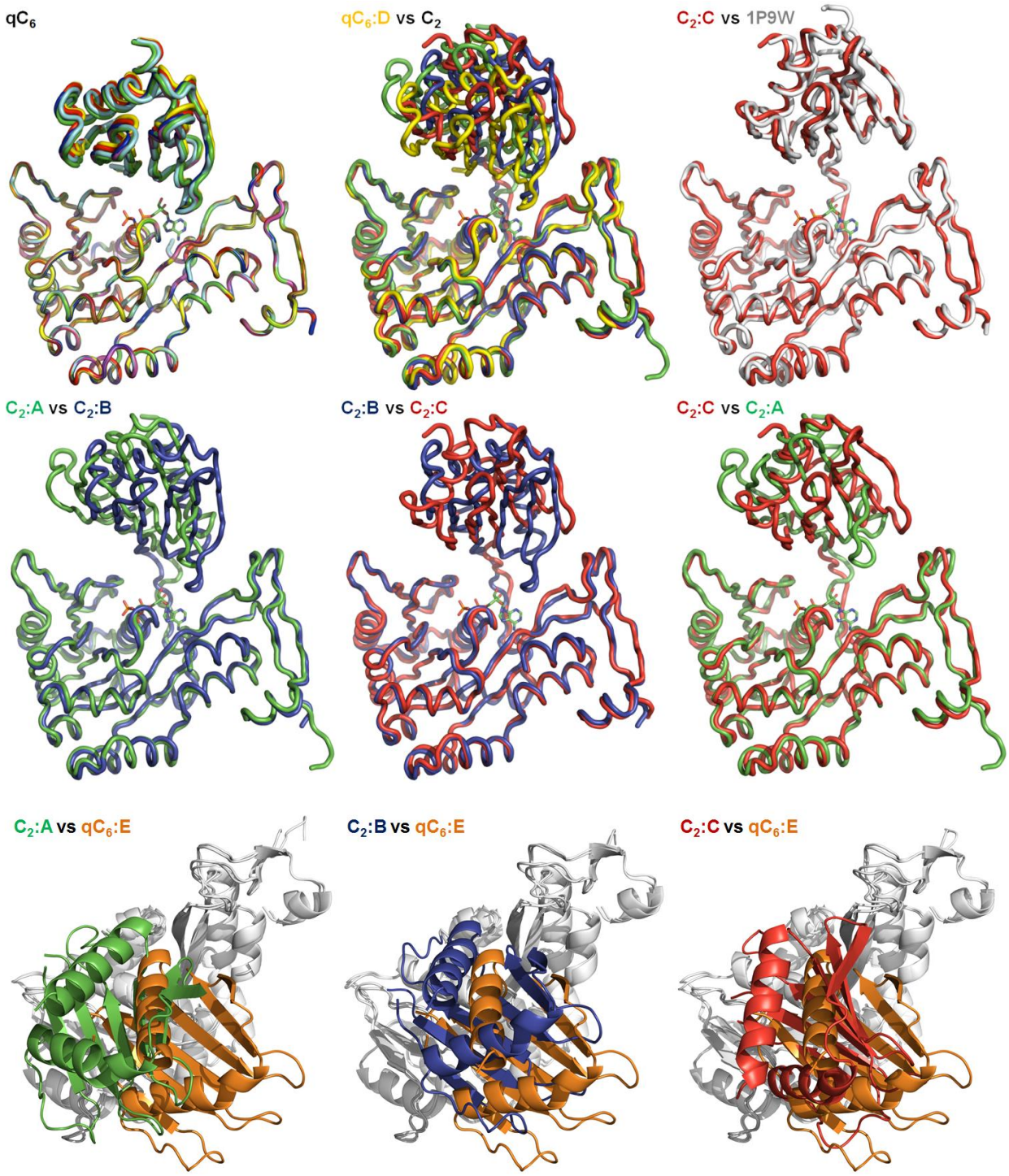

qC6
qC6:D vs C2
C2:C vs 1P9W
C2:A vs C2:B
C2:B vs C2:C
C2:C vs C2:A
C2:A vs qC6:E
C2:B vs qC6:E
C2:C vs qC6:E

Figure 3-5: The variability of the N2E-*vs*-CTE orientations in *V. cholerae* GspE. Top and middle: Superimposed subunits in this "canonical view" are shown with the CTEs superimposed below and the N2Es on top (colored as Figure 3-4). The nucleotide shown for reference is AMPPNP from the helical $^{\Delta 90}$GspE structure (PDB: 1p9w [17]). For N2E-*vs*-CTE orientations see also Table 2. Top left: superposition of the six subunits of the $^{\Delta N1}$GspE qC_6 hexamer from the $^{\Delta N1}$GspE-6aa-Hcp1 structure, revealing only small differences, by 1 to 5 degrees, in N2E-*vs*-CTE orientations. Top middle: superposition of subunit D (yellow) from the $^{\Delta N1}$GspE qC_6 hexamer and the three subunits of the $^{\Delta N1}$GspE C_2 hexamer from the $^{\Delta N1}$GspE-8aa-Hcp1 structure. N2E-*vs*-CTE orientations vary by 16 to 41 degrees. Top right: superposition of subunit C (red) from the $^{\Delta N1}$GspE C_2 hexamer and $^{\Delta N1}$GspE (grey) from the helical $^{\Delta 90}$GspE structure (PDB: 1p9w [17]). The difference in N2E-*vs*-CTE orientation is only 2 degrees. Middle row from left to right is superposition of subunits A (green) and B (blue), subunits B (blue) and C (red), and subunits C (red) and A (green) of the $^{\Delta N1}$GspE C_2 hexamer. The differences in N2E-*vs*-CTE orientation are 32, 48, 47 degrees, respectively. Bottom: "Orthogonal views" of the subunits in *V. cholerae* GspE. Pairwise comparison of $^{\Delta N1}$GspE subunits after superposition of the CTEs (grey, as background) viewed in a direction approximately perpendicular to the "canonical view" in top and middle. The N2E of subunit E from the $^{\Delta N1}$GspE qC_6 hexamer (orange) is used as reference for each case.

Despite the large freedom of domain motion within one subunit revealed by this comparison, the qC_6 and C_2 hexamers of $^{\Delta N1}$GspE share the manner in which the CTE of one subunit interacts with the N2E of an adjacent subunit (N2E') (Figure 3-6). The same CTE•N2E' arrangement is also observed in the helical structure of $^{\Delta 90}$GspE [17]. Apparently, the $^{\Delta N1}$GspE hexamers reported here, and the $^{\Delta 90}$GspE helix with 6_1 symmetry, are all built from essentially the same CTE•N2E' construction unit. The different $^{\Delta N1}$GspE hexamers are obtained by different conformations of the N2E-to-CTE linkers connecting six such construction units. This results in strikingly different orientations and positions of all domains in the two hexamers. For instance, the distances between the Cα atoms of residue Ile458, located near the end of the first helix α10 in the C2E, is 35.5 ± 2.5 Å in non-adjacent subunits in the qC_6 hexamer but ranges from 39 to 70 Å in the C_2 hexamer. Such variations suggest that large conformational changes of the subunits during functioning of this assembly ATPase in the T2SS may occur. A similar architectural principle, with essentially the same CTE•N2E' building block, has recently been reported for the distantly related FlaI ATPase of the archaellum assembly system from the crenarchaeon *Sulfolobus acidocaldarius* (Reindl et al., 2013).

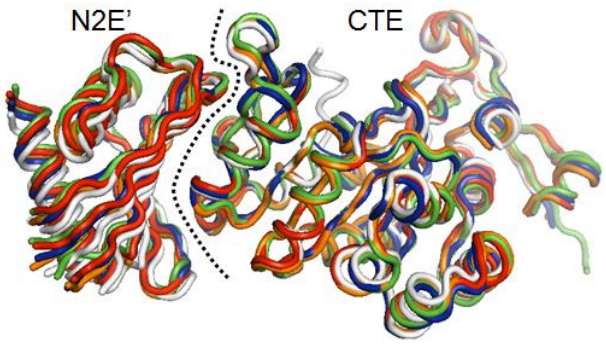


Figure 3-6: The CTE•N2E' construction unit in GspE hexamers and the helical GspE structure. Superposition of the CTEs shows that the CTE•N2E' construction units are essentially the same for all *V. cholerae* GspE structures. Depicted are: one CTE•N2E' unit from the qC_6 hexamer (orange), the three independent units from the C_2 hexamer (colored blue, red and green as in Figure 3-4), and the CTE•N2E' unit from the helical $^{\Delta 90}$GspE structure (grey; PDB: 1P9W [17]). The dashed line indicates the separation between the N2E' and CTE.

Crystals of the four *V. cholerae* $^{\Delta N1}$GspE-linker-Hcp1 variants were obtained in the presence of 5 mM nucleotide(s) and 5 mM $MgCl_2$. Electron density at the expected position of the nucleotide, close to the CTE of $^{\Delta N1}$GspE subunits, could be observed in the three crystallographically independent subunits of the C_2 hexamer, and also in the six subunits of the qC_6 hexamer (Figure 3-7). The densities for the phosphoryl groups of the nucleotides occur at approximately the same position as Mg•AMPPNP in the helical $^{\Delta 90}$GspE structure [17]. The $^{\Delta N1}$GspE-8aa-Hcp1 structure was solved with Mg•ADP plus $AlCl_3$ and NaF but no clear evidence for density beyond the β-phosphoryl group was present. The densities vary in the nine crystallographically independent subunits which might indicate that the nucleotide is present with different occupancies, but given the uncertainties due to the low resolution of the structures it seems best not to discuss this point here in detail. Higher resolution structures will be required to reveal the mode of nucleotide binding more precisely.

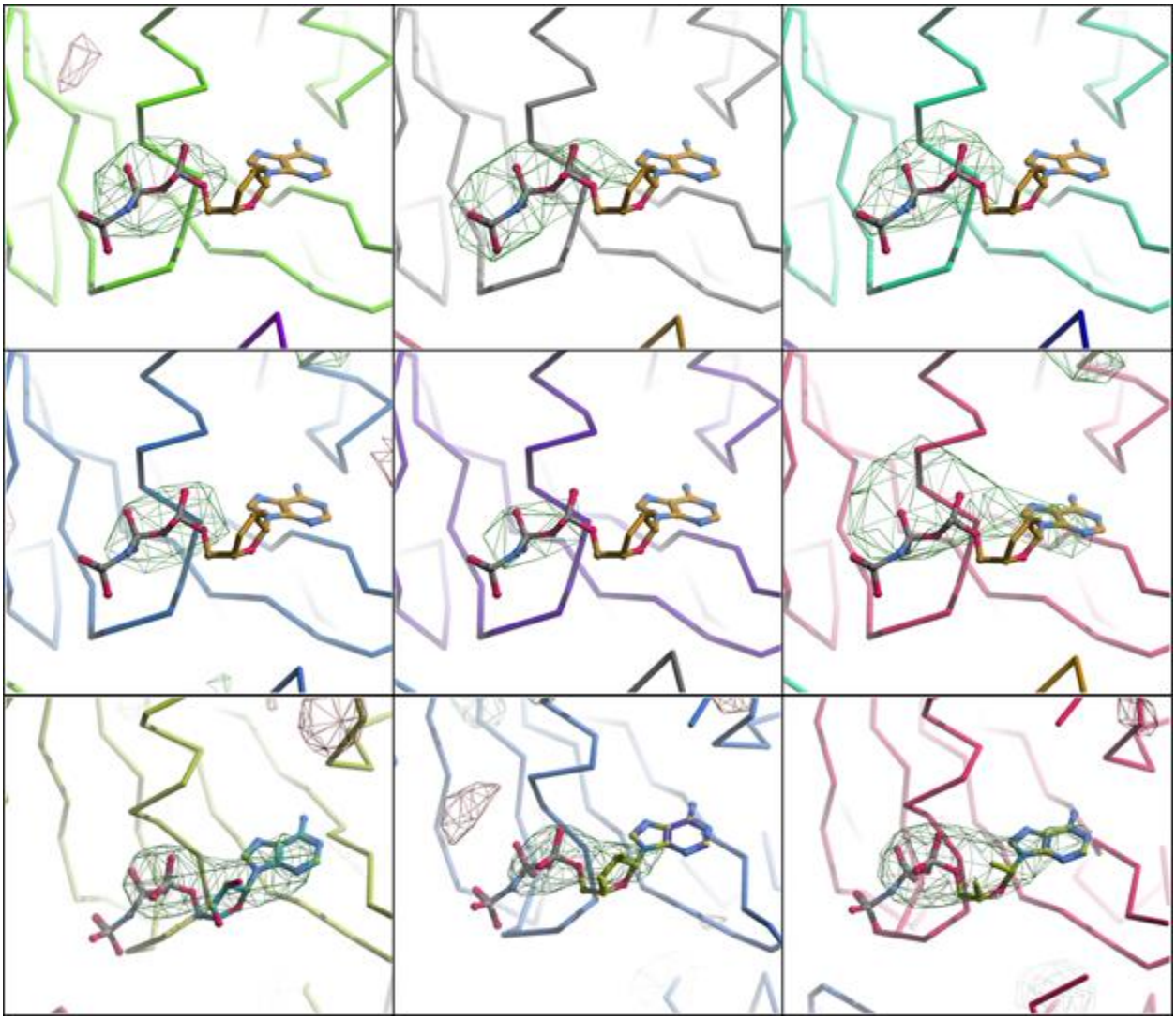


Figure 3-7: Electron densities of nucleotides in *V. cholerae* $^{\Delta N1}$GspE-Hcp1 fusion structures. The Cα traces of each crystallographically independent subunit are shown in different colors. The AMPPNP coordinates shown are superposed from the CTE of the $^{\Delta 90}$GspE structure (PDB: 1P9W [17]) onto the CTE of each crystallographically independent subunit. (Fobs-Fcalc) difference electron densities at the 3 sigma level are shown. The phases were obtained without including nucleotide coordinates. Top and middle: electron densities at the nucleotide position in the six subunits of $^{\Delta N1}$GspE-6aa-Hcp1. The nucleotides present in the protein solution were ADP and AMPPNP. Bottom: electron densities at the nucleotide position in the $^{\Delta N1}$GspE-8aa-Hcp1 structure. The nucleotide present in the protein solution was ADP, $AlCl_3$, and NaF.

The metal binding domains are located on the periphery of both $^{\Delta N1}$GspE hexamers. Despite the ~7 Å resolution of the data, the presence of zinc ions in the CMEs was unambiguously confirmed by six peaks in the anomalous difference density at the expected positions in the $^{\Delta N1}$GspE-6aa-Hcp1 crystal structure (Figure 3-8). The positions of zinc ions of non-adjacent subunits are 114 ± 8 Å apart in the qC_6 hexamer. These inter-zinc distances range from 91 to 140 Å in the C_2 hexamer, again showing that the two hexamers are dramatically different.

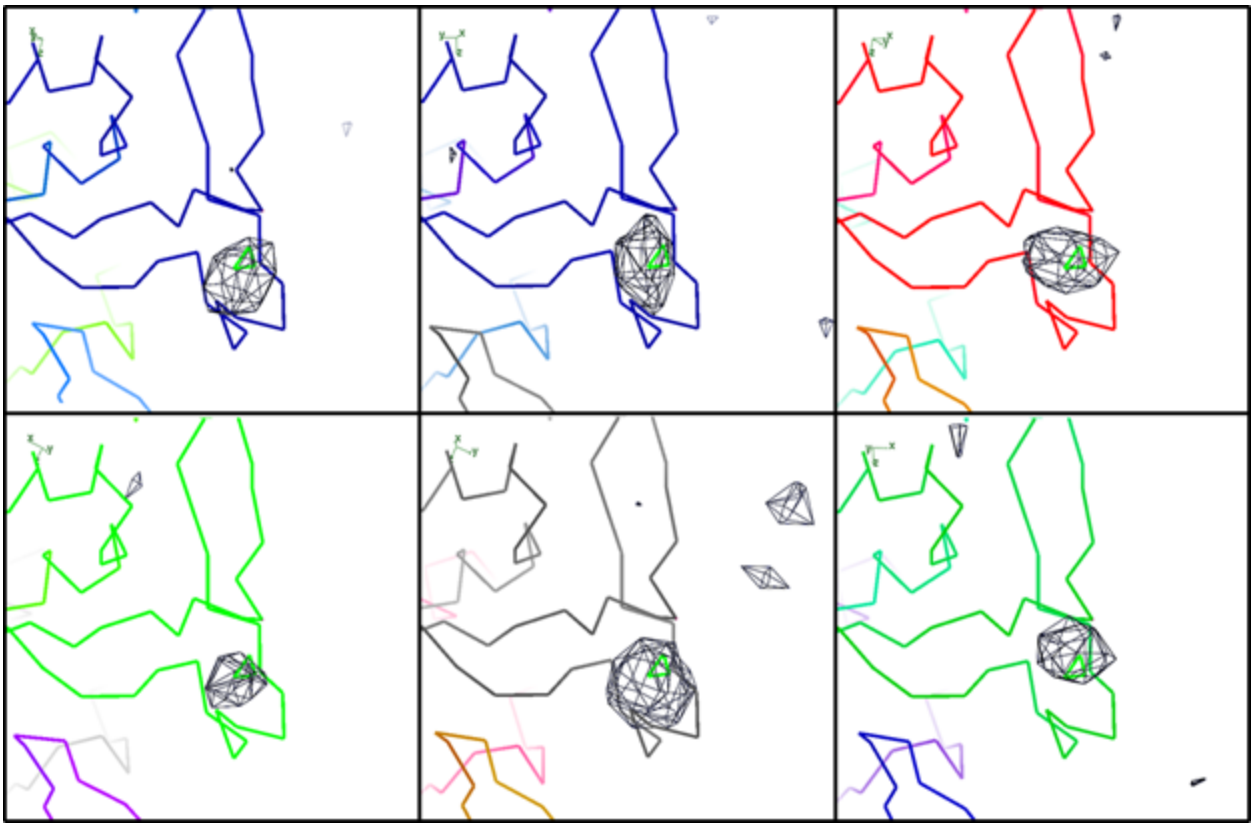


Figure 3-8: Anomalous electron difference densities of zinc sites in V. cholerae $^{\Delta N1}$GspE-6aa-Hcp1. The peak heights at the zinc positions (as located in the $^{\Delta 90}$GspE structure (PDB: 1P9W; [17]) in the six independent crystallographic subunits of the $^{\Delta N1}$GspE-6aa-Hcp1 hexamer are: 4.4, 4.5, 4.5, 4.8, 5.3 and 6.7 sigma.

3.3.3 SAXS Studies

SAXS studies were carried out to investigate whether the crystal structures correspond with the $^{\Delta N1}$GspE hexamers in solution and if the presence of different nucleotides affects the structure of the $^{\Delta N1}$GspE hexamer of the fusion proteins. The observed scattering curve of $^{\Delta N1}$GspE-8aa-Hcp1 in the presence of 5 mM ADP, 5 mM $MgCl_2$, 5 mM $AlCl_3$ and 15 mM NaF closely follows the calculated scattering curve based on the crystal structure of $^{\Delta N1}$GspE-8aa-Hcp1 elucidated in the presence of the same compounds and containing the elongated C_2 hexamer (Figure 3-9, left). In contrast, the calculated scattering curve using the crystal structure of $^{\Delta N1}$GspE-6aa-Hcp1 with the $^{\Delta N1}$GspE qC_6 hexamer matches the observed scattering curve of $^{\Delta N1}$GspE-8aa-Hcp1 poorly. Hence it appears that in solution the structure of $^{\Delta N1}$GspE-8aa-Hcp1 is close to that observed in the crystal structure.

Surprisingly, the observed scattering curve of $^{\Delta N1}$GspE-6aa-Hcp1 deviates from the calculated scattering curve based on the crystal structure of $^{\Delta N1}$GspE-6aa-Hcp1. The calculated scattering curve using the crystal structure of $^{\Delta N1}$GspE-8aa-Hcp1 also resulted in a poor fit with the observed curve of $^{\Delta N1}$GspE-6aa-Hcp1 (Figure 3-9, right). It is therefore likely that the architecture of the $^{\Delta N1}$GspE hexamer of the $^{\Delta N1}$GspE-6aa-Hcp1 complex in solution is neither that of the qC_6 nor of the C_2 hexamer observed in the crystals.

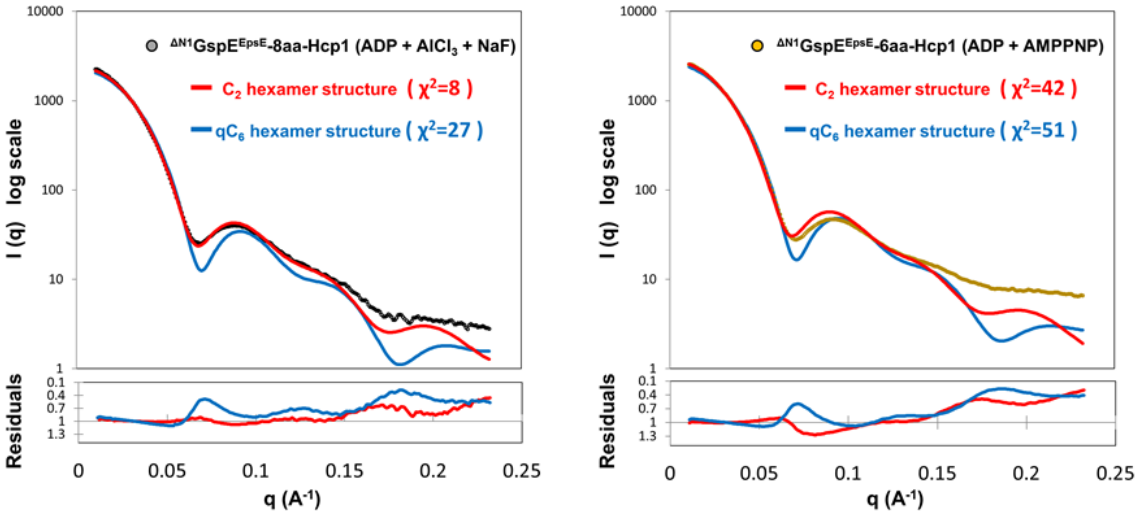


Figure 3-9: SAXS studies on $^{\Delta N1}$GspE-8aa-Hcp1 and $^{\Delta N1}$GspE-6aa-Hcp1. Experimental (black and yellow) and calculated (red and blue) SAXS scattering curves for $^{\Delta N1}$GspE-8aa-Hcp1 and $^{\Delta N1}$GspE-6aa-Hcp1 with the same nucleotides as in the mother liquor of the crystal structures (see the right upper corner, and for specific details see methods). The calculated SAXS curve based on the $^{\Delta N1}$GspE-6aa-Hcp1 structure with the qC_6 hexamer of $^{\Delta N1}$GspE shown in blue, and for the $^{\Delta N1}$GspE-8aa-Hcp1 structure with the C_2 hexamer of $^{\Delta N1}$GspE shown in red. Residual plots are shown in the lower section.

3.3.4 Comparison of T2SS and T4PS ATPase Hexamers

Crystal structures of hexamers with various symmetries have been reported for the T4PS ATPases PilT [20, 22] from the Gram-negative bacteria *Aquifex aeolicus* and *Pseudomonas aeruginosa* (hereafter called *Aa*PilT and *Pa*PilT, respectively). These retraction ATPases differ substantially from the T2SS ATPases because they lack two domains: the N1E and CME. Regarding the common domains, the amino acid sequence identities of *Aa*PilT and *Pa*PilT compared to *V. cholerae* GspE are ~18 % for the N2Es, ~42 % for the C1Es, and ~18 % for the C2Es, respectively (Figure 3-1). While the ATPase hexamers of *V. cholerae* $^{\Delta N1}$GspE-6aa-Hcp1 and of *Aa*PilT, with qC_6 and C_6 symmetry respectively, have approximately the same overall dimensions (Figure 3-10), the N2E-*vs*-CTE orientations within subunits are surprisingly different in these two cases; varying by ~40 degrees in pairwise comparisons (Table 2; Figure 3-11). As a consequence, the N2Es of the $^{\Delta N1}$GspE qC_6 hexamer move further away from the central C1Es than the N2Es of the *Aa*PilT C_6 hexamer and the *Pa*PilT C_2 hexamer (Figure 3-12). The elongated $^{\Delta N1}$GspE C_2 hexamer is in turn very dissimilar from the irregular qC_2 hexamer of *Aa*PilT [20]: in this case the N2E arrangement of the $^{\Delta N1}$GspE C_2 hexamer is distinctly different from that in the *Aa*PilT qC_2 hexamer (Figure 3-12). Yet, despite these major differences in hexamer shape, the two-domain CTE•N2E' construction unit observed in the two new $^{\Delta N1}$GspE hexamers is essentially the same as in *Aa*PilT and in *Pa*PilT (Figure 3-13).

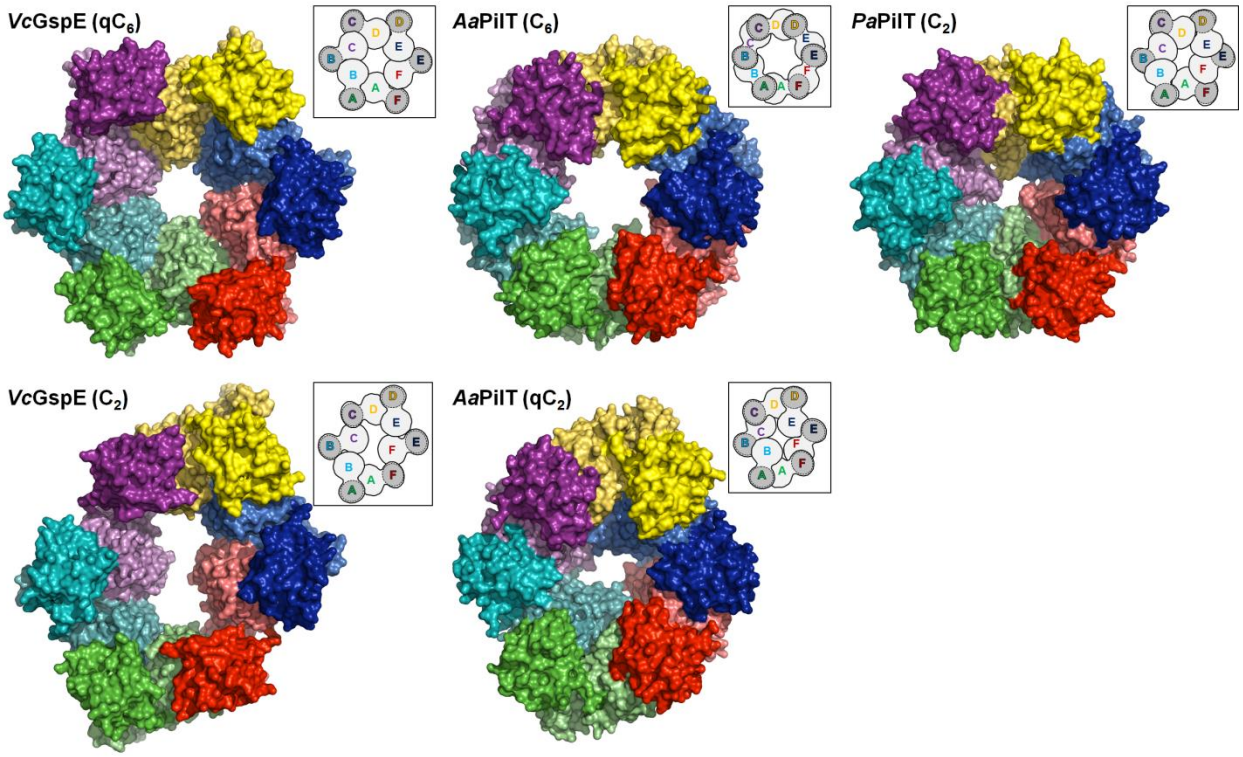


Figure 3-10: Comparison of the regular and irregular T2SS GspE and T4PS PilT hexamers. The N2E and CTE of the same subunit are represented with darker and lighter shades of the same color. Top: comparison of regular hexamers. Top left: *V. cholerae* $^{\Delta N1}$GspE qC_6 hexamer. Top middle: *A. aeolicus* PilT C_6 hexamer (PDB: 2EWV). Top right: *P. aeruginosa* PilT C_2 hexamer (PDB: 3JVV). Bottom: comparison of irregular hexamers. Bottom left: *V. cholerae* $^{\Delta N1}$GspE C_2 hexamer. Bottom middle: *A. aeolicus* PilT qC_2 hexamer (PDB: 2GSZ). Insets: hexamers with the CTE•N2E' construction units outlined.

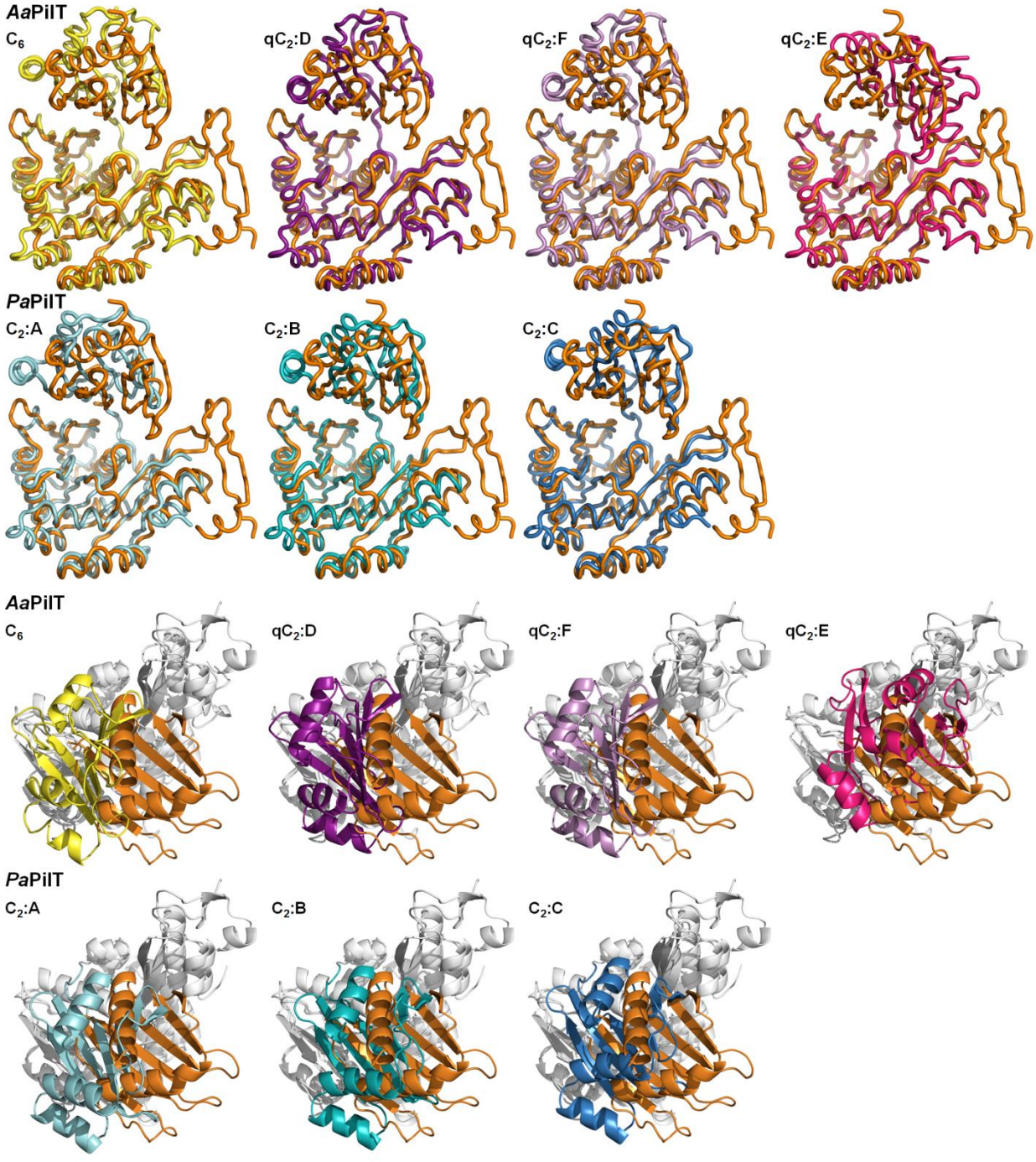

*Aa*PilT
C_6
qC_2:D
qC_2:F
qC_2:E
*Pa*PilT
C_2:A
C_2:B
C_2:C
*Aa*PilT
C_6
qC_2:D
qC_2:F
qC_2:E
*Pa*PilT
C_2:A
C_2:B
C_2:C

Figure 3-11: The variability of the N2E-*vs*-CTE orientations in T2SS GspE and T4PS PilT hexamers. Superimposed subunits are shown based on a superposition of two CTEs. Subunit E of the $^{\Delta N1}$GspE qC_6 hexamer (orange) functions as reference in all cases. None of the other subunits have the "orange" N2E-*vs*-CTE orientation. For N2E-*vs*-CTE orientations, see Table 2. Top two rows: "canonical view," with the CTEs superimposed below and the N2Es on top. Bottom two rows: "orthogonal view," viewed with the N2Es on top and the CTEs below. The direction of the "orthogonal view" is approximately orthogonal to the "canonical views." The superimposed CTEs of each pair of subunits are colored grey. First and third rows: superposition of the CTEs from the subunit of the *Aa*PilT C_6 hexamer (yellow; PDB: 2EWV) and the three independent subunits of the *Aa*PilT qC_2 hexamer (different shades of purple; PDB: 2GSZ) onto subunit E of the $^{\Delta N1}$GspE qC_6 hexamer (orange). Second and fourth rows: superposition of the CTEs of *Pa*PilT C_2 hexamer (different shades of blue; PDB: 3JVV) onto subunit E of the $^{\Delta N1}$GspE qC_6 hexamer (orange).

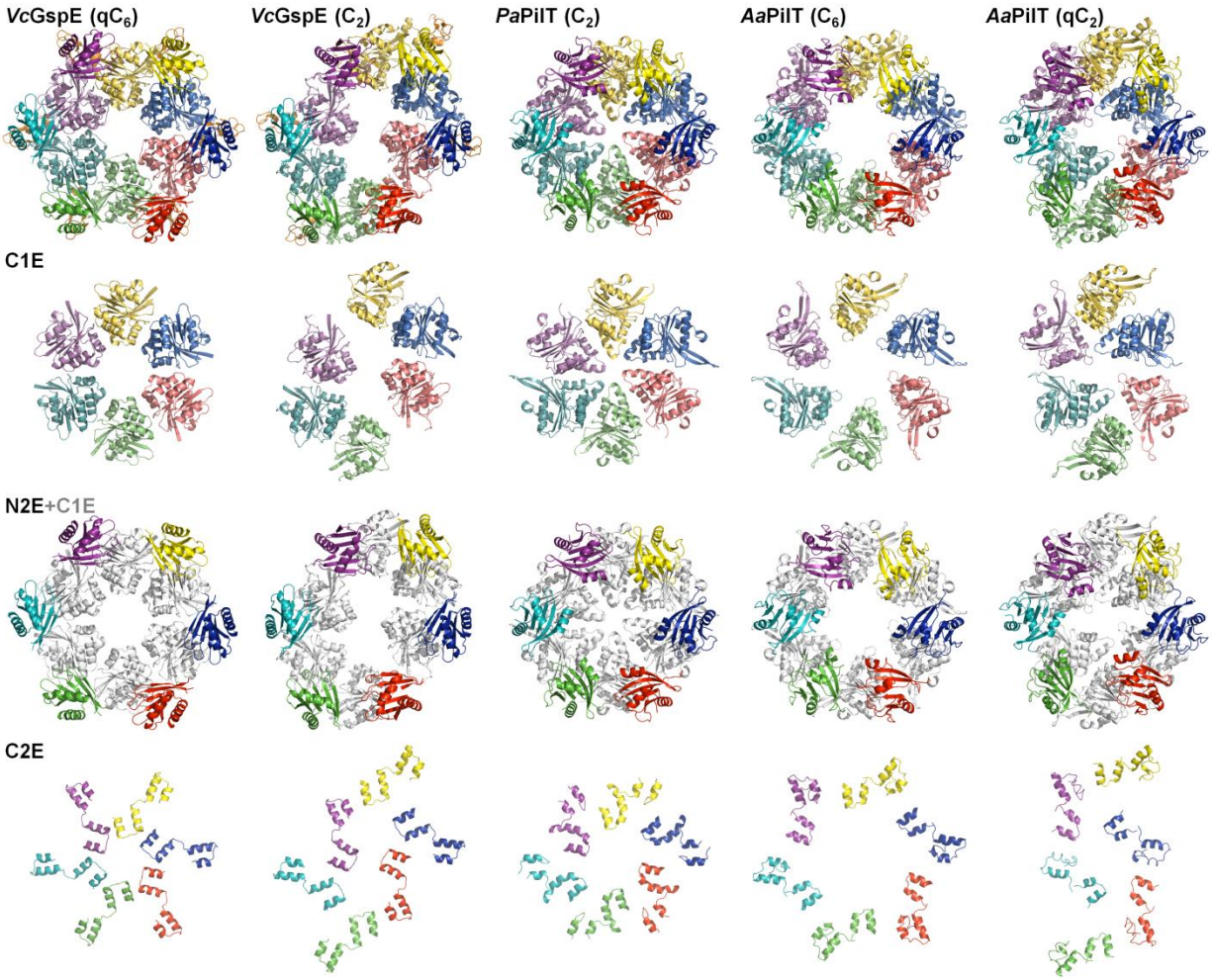


Figure 3-12: Domain rearrangements of T2SS GspE and T4PS secretion ATPases. Rows from top to bottom: hexamers, the C1Es only, the N2Es in color with C1Es in grey as background, the C2Es only. Note how different the domain positions and orientations are in the various hexamers.

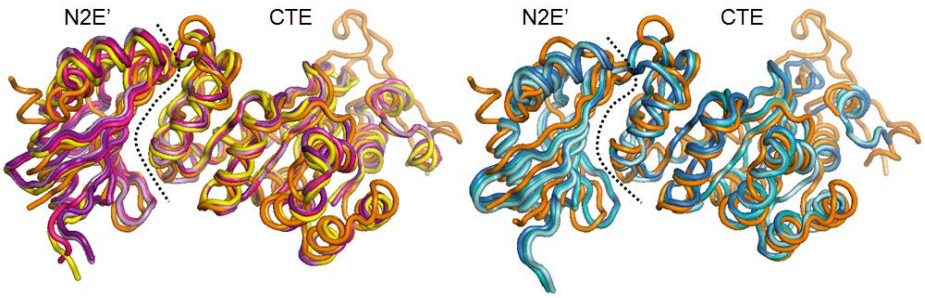


Figure 3-13: The CTE•N2E' construction units in T2SS GspE and T4PS PilT hexamers. The CTE•N2E' construction units of the T2SS GspE ATPase and of two T4PS PilT ATPases superimpose remarkably well. The dotted line indicates the separation of the N2E' and the CTE. Left: superposition of the CTEs of the *Aa*PilT C_6 hexamer (yellow; [20]) and *Aa*PilT quasi C_2 hexamer (different shades of purple) onto subunit E of the $^{\Delta N1}$GspE C_6 hexamer (orange). Right: superposition of the CTEs of *Pa*PilT (different shades of blue; [22]) onto subunit E of the $^{\Delta N1}$GspE C_6 hexamer (orange).

The five available crystal structures of PilT hexamers provide seven independent views of *Aa*PilT and *Pa*PilT subunits. Two of these, subunits D and F of *Aa*PilT qC_2 (PDB: 2GSZ) [20], are very similar to each other (Table 2). Hence, six different N2E-*vs*-CTE PilT orientations are available for comparison with the four orientations in GspE subunits in the two hexamers. For 24 mutual comparisons of four T2SS GspE *versus* six T4PS PilT ATPase subunits, the smallest difference in N2E-*vs*-CTE orientation is 14.7 degrees and the largest as much as 73.4 degrees. Clearly, none of these six PilT orientations are similar to the four GspE orientation observed in this study (Figure 3-11 and Table 2). This is particularly clear when viewed in a direction (Figure 3-11, bottom two rows) that is approximately orthogonal with respect to the "classical" direction of view (Figure 3-11, top two rows). Hence, despite considerable sequence similarity of the N2Es and C1Es of GspE and PilT, there is no similarity in the N2E-*vs*-CTE orientation among the subunits in the known structures of these two ATPases.

3.4 DISCUSSION

We have succeeded in obtaining $^{\Delta N1}$GspE hexamers in solution by fusion to an "assistant hexamer" as revealed by native mass spectrometry (Figure 3-2). In crystals, depending on linker length, two types of $^{\Delta N1}$GspE hexamers are observed (Figure 3-4). Importantly, the four $^{\Delta N1}$GspE-Hcp1 hexamers of ~385 kDa obtained show an approximately 20-fold increase in activity with respect to $^{\Delta N1}$GspE not fused to an assistant hexamer (Figure 3-3). The strategy to bring GspE subunits into proximity by linking to an assistant hexamer has resulted for the first time in structures of T2SS ATPase hexamers and in a substantial increase in ATPase activity as a result of the hexamerization of $^{\Delta N1}$GspE.

Linking domains, peptides and proteins has been applied mainly to cases where two proteins, or a peptide and a protein, with weak affinity for each other were linked together (see [35]), although sometimes also up to four domains from different proteins have been connected [36]. Another interesting example of protein linking in structural studies is the hexameric CA building block of the HIV capsid. While CA hexamers had been observed at ~9 Å resolution by cryo-electron microscopy [37], fusing CA to the assistant hexamer CcmK4 resulted in a higher- resolution crystal structure of CA hexamers [38]. Intriguingly, in spite of a linker of only two residues in this case, the CcmK4 assistant hexamer could not be localized in the crystals of the CA-2aa-CcmK4 fusion protein. Both crystal forms we obtained of $^{\Delta N1}$GspE-Hcp1 fusions did reveal $^{\Delta N1}$GspE and Hcp1 hexamers, though the crystal packing differs. In particular in the case of the C_2 hexamers, alternating layers of $^{\Delta N1}$GspE and Hcp1 hexamers occur (Figure 3-14).

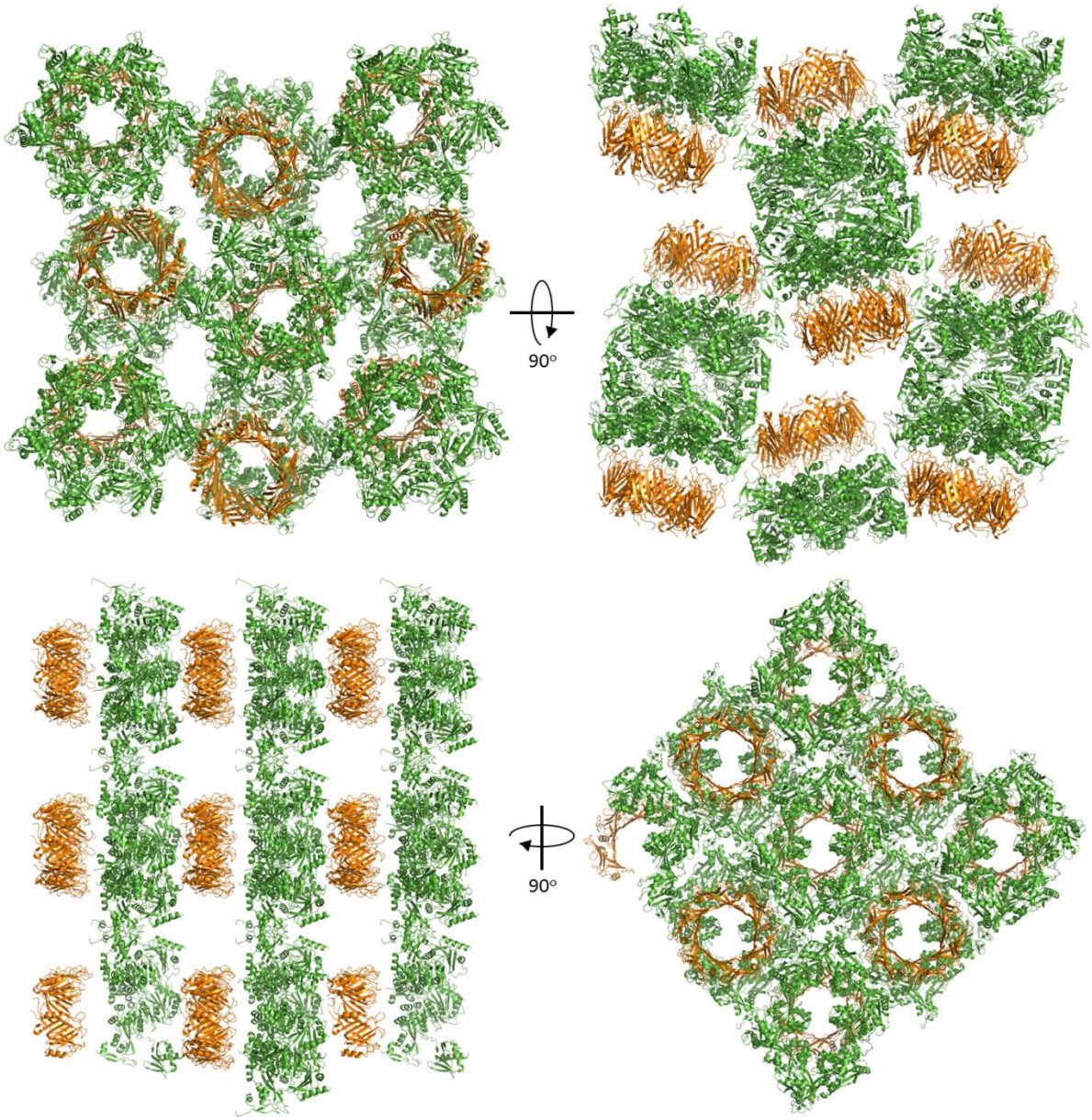


Figure 3-14: Crystal packing of $^{\Delta N1}$GspE-Hcp1 hexamers. Crystal packing of $^{\Delta N1}$GspE-6aa-Hcp1 (top) and $^{\Delta N1}$GspE-8aa-Hcp1 (bottom) with $^{\Delta N1}$GspE in green and Hcp1 in orange. In the case of $^{\Delta N1}$GspE-8aa-Hcp1 alternating layers of Hcp1 and $^{\Delta N1}$GspE are clearly present.

The fact that subunit C of the $^{\Delta N1}$GspE-8aa-Hcp1 structure is highly similar to the $^{\Delta 90}$GspE helical structure suggests that subunit conformation is important. An interesting remaining question is what the total number of major conformations is which T2SS GspE subunits adopt under physiologic conditions. Are the four N2E-*vs*-C2E orientations observed in this study the sole conformations available to GspE subunits? Are the two hexamers observed in this study the only hexameric arrangements possible for GspE? The SAXS data reveals that the elongated C_2 hexamer of $^{\Delta N1}$GspE occurs in solution as part of the $^{\Delta N1}$GspE-8aa-Hcp1 hexamer. In contrast, in the conditions tested, neither the regular qC_6 nor the elongated C_2 hexamers appeared to correspond with the SAXS data of the $^{\Delta N1}$GspE-6aa-Hcp1 hexamer (Figure 3-9). This suggests that GspE in the assembled T2SS might form more hexamer variants than the two observed in our structures. Perhaps GspE in the $^{\Delta N1}$GspE-6aa-Hcp1 fusion exists in multiple conformations but only one of these, which is not predominant in solution, is seen in the crystals. Interestingly, the $^{\Delta N1}$GspE-6aa-Hcp1 and $^{\Delta N1}$GspE-8aa-Hcp1 assemblies have essentially the same ATPase activity (Figure 3-3). This suggests that in both hexamers the subunits of $^{\Delta N1}$GspE can adopt with equal frequency the conformations capable of ATP hydrolysis.

The six different PilT N2E-*vs*-CTE orientations (Figure 3-11) possibly provide an additional set of subunit conformations available to GspE. The N2E and CTE could then differ by up to 73 degrees in mutual orientation (Table 2). An even larger set of possible hexamer arrangements is obtained when considering distant assembly systems occurring in Archaea. These prokaryotes lack an outer membrane but contain machineries that assemble flagellins, relatives of T4PS pilins and T2SS pseudopilins, into archaella. When compared with GspE, the archaellum assembly system ATPases have an amino acid sequence identity of ~28% for the C1Es and of ~13% for the N2Es, while the N1Es and C2Es have different folds, and the CME is absent (Figure 3-1). In the subunits of archaellum hexameric ATPases [19, 23], approximately three additional N2E-*vs*-CTE orientations occur that differ from the N2E-*vs*-CTE orientations observed so far in GspE and in PilT subunits (Figure 3-15). If some or all of

the known N2E-*vs*-C2E orientations in these ATPases would also be available to the T2SS GspE ATPases, then the ensemble of possible hexameric T2SS GspE conformations would become truly large. Alternatively, the T2SS, T4PS and archaellum assembly systems might each have a specific, more limited set of N2E-*vs*-CTE orientations at their disposal to perform their respective functions. Interestingly, the CTE•N2E' construction unit of the FlaI ATPase from the archaellum assembly system in *S. acidocaldarius* (Reindl et al., 2013) differs from the CTE•N2E' construction unit of *V. cholerae* GspE with an r.m.s.d. of only ~2.5 Å. Clearly, this large family of secretion ATPases is using quite inflexible construction units and very flexible subunits.

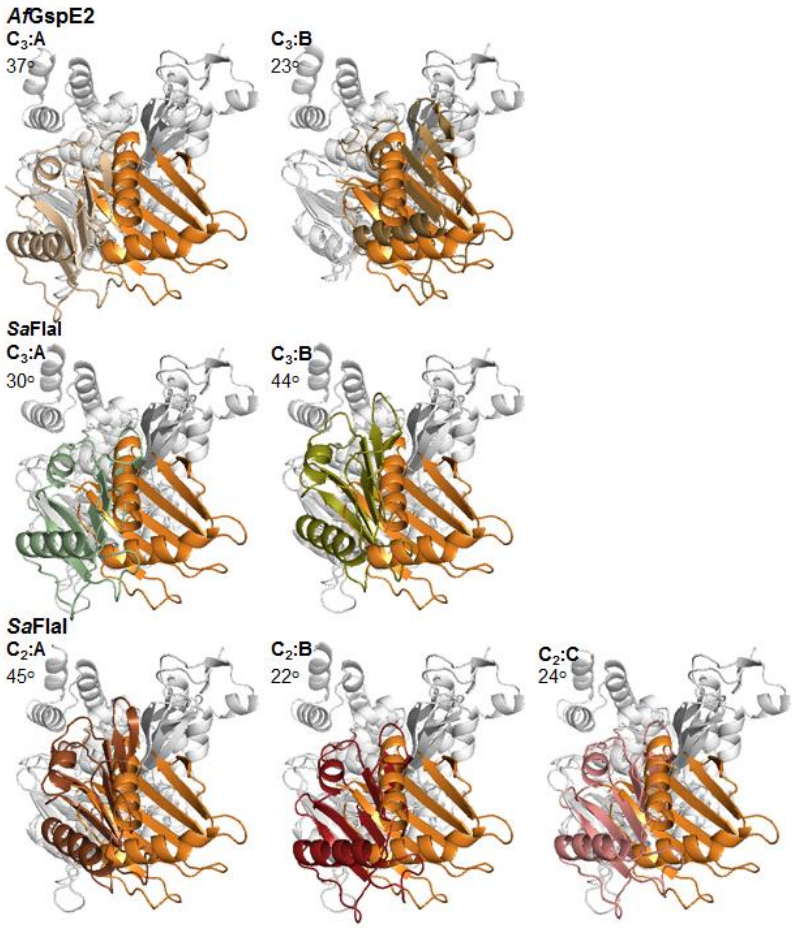


Figure 3-15: The variability of the N2E-*vs*-CTE orientations in *V. cholerae* GspE and archaeal ATPases fom *Archaeoglobus fulgidus* (*Af*GspE2) and *Sulfolobus acidocaldarius* (*Sa*FlaI). Differences in orientation of the N2E-*vs*-CTE orientations in T2SS and archaeal non-T2SS ATPases, viewed with the N2E on top and the CTE below. The view is approximately orthogonal to the "canonical view" in Figure 3-5 and 3-11. The superimposed CTEs of each pair of subunits are colored grey. Subunit E of the $^{\Delta N1}$GspE qC_6 hexamer (N2E in orange) functions as reference in all pairs. The difference in N2E orientation is shown in degrees in the left upper corner of each pair. The percentage of sequence identity per domain is given in Figure 3-1. From top to bottom, superposition of the CTEs of *Af*GspE2 C_3 hexamer (different shades of brown; PDB: 2OAP [19]), *Sa*FlaI C_3 hexamer (different shades of green; PDB: 4II7 [23]) and *Sa*FlaI C_2 hexamer (different shades of red; PDB: 4II7) onto subunit E of the $^{\Delta N1}$GspE qC_6 hexamer (orange).

Recently, a cryo-electron microscopy study of the assembly ATPase PilF of the T4PS involved in DNA transformation by the Gram-negative *Thermus thermophilus* has been reported [26]. This T4PS assembly ATPase contains the zinc-binding CME, like the T2SS assembly ATPase GspE, but has three, not one, N1 domains that are homologous to each other (Figure 3-1). In these reconstructions of *Tt*PilF, with imposed six-fold symmetry, the ATPase rings have distinctly different shapes in the absence and presence of AMPPNP. Whether *Tt*PilF can form hexamers with lower symmetry will require further investigations. Possibly the N-terminal domains of *Tt*PilF assist in the hexamerization of the assembly ATPase *Tt*PilF domains, analogous to the way the Hcp1 hexamer induces hexamer formation of GspE in our fusion proteins (Figure 3-4).

While the assistant-hexamer approach has been successful in this *in vitro* study, a crucial question for understanding the mechanism of T2SS secretion is which factors are responsible for hexamerization of GspE *in vivo*. It is obviously important to avoid futile ATP hydrolysis by the T2SS ATPase in the absence of exoproteins in the periplasm. Hence, full activation, almost certainly by hexamerization, of the T2SS GspE is likely to occur only under specific conditions. It has been proposed that the inner membrane platform depends on the assembly of the secretin GspD in the outer membrane [39]. Hence, T2SS GspE hexamerization might only occur after the GspD dodecamer is formed in the outer membrane, and the inner membrane platform is assembled with GspL likely being a key factor in the hexamerization of GspE. Maybe an exoprotein catalyzes the GspD-assisted assembly process of the inner membrane platform, as has been suggested for the T2SS in *Xanthomonas campestris*. [40]. However, there are studies that show in the related T4PS, the pilus can be formed when the T4PS secretin PilQ is not expressed [41]. Given that the T2SS and the T4PS share several features, this might suggest that in the T2SS GspD is not absolutely required for pseudopilin assembly. Evidently, despite major progress made in recent years, profound mysteries regarding the mechanism of action of the T2SS remain.

3.5 REFERENCES

1. Lu, C., et al., *Hexamers of the type II secretion ATPase GspE from Vibrio cholerae with increased ATPase activity.* Structure, 2013. **21**(9): p. 1707-17.
2. Korotkov, K.V., M. Sandkvist, and W.G. Hol, *The type II secretion system: biogenesis, molecular architecture and mechanism.* Nature reviews. Microbiology, 2012. **10**(5): p. 336-51.
3. McLaughlin, L.S., R.J.F. Haft, and K.T. Forest, *Structural insights into the Type II secretion nanomachine.* Current Opinion in Structural Biology, 2012. **22**(2): p. 208-216.
4. Douzi, B., A. Filloux, and R. Voulhoux, *On the path to uncover the bacterial type II secretion system.* Philosophical Transactions of the Royal Society B-Biological Sciences, 2012. **367**(1592): p. 1059-1072.
5. O'Neal, C.J., et al., *Crystal structures of an intrinsically active cholera toxin mutant yield insight into the toxin activation mechanism.* Biochemistry, 2004. **43**(13): p. 3772-82.
6. Hirst, T.R., et al., *Mechanism of toxin secretion by Vibrio cholerae investigated in strains harboring plasmids that encode heat-labile enterotoxins of Escherichia coli.* Proceedings of the National Academy of Sciences of the United States of America, 1984. **81**(24): p. 7752-6.
7. Korotkov, K.V., T. Gonen, and W.G. Hol, *Secretins: dynamic channels for protein transport across membranes.* Trends in biochemical sciences, 2011. **36**(8): p. 433-43.
8. Ayers, M., P.L. Howell, and L.L. Burrows, *Architecture of the type II secretion and type IV pilus machineries.* Future microbiology, 2010. **5**(8): p. 1203-18.
9. Craig, L. and J. Li, *Type IV pilli: paradoxes in form and function.* Current Opinion in Structural Biology, 2008. **18**(2): p. 267-277.
10. Giltner, C.L., Y. Nguyen, and L.L. Burrows, *Type IV pilin proteins: versatile molecular modules.* Microbiology and molecular biology reviews : MMBR, 2012. **76**(4): p. 740-72.
11. Planet, P.J., et al., *Phylogeny of genes for secretion NTPases: identification of the widespread tadA subfamily and development of a diagnostic key for gene classification.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(5): p. 2503-8.
12. Peabody, C.R., et al., *Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella.* Microbiology, 2003. **149**(Pt 11): p. 3051-72.
13. Chen, Y., et al., *Structure and function of the XpsE N-terminal domain, an essential component of the Xanthomonas campestris type II secretion system.* The Journal of biological chemistry, 2005. **280**(51): p. 42356-63.
14. Abendroth, J., et al., *The X-ray structure of the type II secretion system complex formed by the N-terminal domain of EpsE and the cytoplasmic domain of EpsL of Vibrio cholerae.* Journal of molecular biology, 2005. **348**(4): p. 845-55.
15. Sandkvist, M., et al., *Interaction between the autokinase EpsE and EpsL in the cytoplasmic membrane is required for extracellular secretion in Vibrio cholerae.* The EMBO journal, 1995. **14**(8): p. 1664-73.
16. Shiue, S.J., et al., *XpsE oligomerization triggered by ATP binding, not hydrolysis, leads to its association with XpsL.* The EMBO journal, 2006. **25**(7): p. 1426-35.
17. Robien, M.A., et al., *Crystal structure of the extracellular protein secretion NTPase EpsE of Vibrio cholerae.* Journal of molecular biology, 2003. **333**(3): p. 657-74.
18. Camberg, J.L., et al., *Synergistic stimulation of EpsE ATP hydrolysis by EpsL and acidic phospholipids.* The EMBO journal, 2007. **26**(1): p. 19-27.

19. Yamagata, A. and J.A. Tainer, *Hexameric structures of the archaeal secretion ATPase GspE and implications for a universal secretion mechanism.* The EMBO journal, 2007. **26**(3): p. 878-90.
20. Satyshur, K.A., et al., *Crystal structures of the pilus retraction motor PilT suggest large domain movements and subunit cooperation drive motility.* Structure, 2007. **15**(3): p. 363-76.
21. Patrick, M., et al., *Oligomerization of EpsE coordinates residues from multiple subunits to facilitate ATPase activity.* The Journal of biological chemistry, 2011. **286**(12): p. 10378-86.
22. Misic, A.M., K.A. Satyshur, and K.T. Forest, *P. aeruginosa PilT structures with and without nucleotide reveal a dynamic type IV pilus retraction motor.* Journal of Molecular Biology, 2010. **400**(5): p. 1011-21.
23. Reindl, S., et al., *Insights into FlaI Functions in Archaeal Motor Assembly and Motility from Structures, Conformations, and Genetics.* Molecular Cell, 2013. **49**(6): p. 1069-82.
24. Shiue, S.J., et al., *Mutation of a key residue in the type II secretion system ATPase uncouples ATP hydrolysis from protein translocation.* Molecular microbiology, 2007. **65**(2): p. 401-12.
25. Camberg, J.L. and M. Sandkvist, *Molecular analysis of the Vibrio cholerae type II secretion ATPase EpsE.* Journal of bacteriology, 2005. **187**(1): p. 249-56.
26. Collins, R.F., et al., *Structure and mechanism of the PilF DNA transformation ATPase from Thermus thermophilus.* The Biochemical journal, 2013. **450**(2): p. 417-25.
27. Mougous, J.D., et al., *A virulence locus of Pseudomonas aeruginosa encodes a protein secretion apparatus.* Science, 2006. **312**(5779): p. 1526-30.
28. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.*, in *Methods in Enzymology*, C. Carter and R. Sweet, Editors. 1997, Academic Press: New York. p. 307-326.
29. Kabsch, W., *Xds.* Acta crystallographica. Section D, Biological crystallography, 2010. **66**(Pt 2): p. 125-32.
30. McCoy, A.J., *Solving structures of protein complexes by molecular replacement with Phaser.* Acta Crystallographica Section D-Biological Crystallography, 2007. **63**(Part 1): p. 32-41.
31. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of Macromolecular Structures by the Maximum-Likelihood Method.* Acta Crystallogr D, 1997. **53**(Part 3): p. 240-255.
32. Konarev, P.V., et al., *PRIMUS: a Windows PC-based system for small-angle scattering data analysis.* Journal of Applied Crystallography, 2003. **36**: p. 1277-1282.
33. Schneidman-Duhovny, D., M. Hammel, and A. Sali, *FoXS: a web server for rapid computation and fitting of SAXS profiles.* Nucleic Acids Research, 2010. **38**(Web Server issue): p. W540-4.
34. Svergun, D., C. Barberato, and M.H.J. Koch, *CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates.* Journal of Applied Crystallography, 1995. **28**: p. 768-773.
35. Reddy Chichili, V.P., V. Kumar, and J. Sivaraman, *Linkers in the structural biology of protein–protein interactions.* Protein Science, 2013. **22**(2): p. 153-167.
36. Park, Y.J. and W.G. Hol, *Explorations of linked editosome domains leading to the discovery of motifs defining conserved pockets in editosome OB-folds.* Journal of Structural Biology, 2012. **180**(2): p. 362-73.
37. Ganser-Pornillos, B.K., et al., *Assembly properties of the human immunodeficiency virus type 1 CA protein.* Journal of virology, 2004. **78**(5): p. 2545-52.
38. Pornillos, O., et al., *X-Ray Structures of the Hexameric Building Block of the HIV Capsid.* Cell, 2009. **137**(7): p. 1282-1292.

39. Howard, S.P., *Assembly of the type II secretion system.* Research in Microbiology, 2013. **164**(6): p. 535-544.
40. Chen, Y.-L. and N.-T. Hu, *Function-Related Positioning of the Type II Secretion ATPase of Xanthomonas campestris pv. campestris.* PLoS ONE, 2013. **8**(3).
41. Wolfgang, M., et al., *Components and dynamics of fiber formation define a ubiquitous biogenesis pathway for bacterial pili.* EMBO J, 2000. **19**(23).