Cryo-ET Reveals Molecular Details of Multi-Megadalton Bacterial Protein Complexes

Thesis by
Przemysław Dutka

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
(Defended April 28, 2023)

© 2023
Przemysław Dutka
ORCID: 0000-0003-3819-1618

ABSTRACT

Cryo-electron tomography (cryo-ET) is a powerful method for investigating the 3D structure of intact cells, organelles, and complex protein macromolecules that cannot be crystallized or are too heterogenous for single-particle cryo-electron microscopy (cryo-EM). However, obtaining high-resolution cryo-ET structures for many biologically important targets is still a challenge. To address this challenge, cryo-ET can be combined with other methods, including X-ray crystallography, single-particle cryo-EM, structure predictions, cross-linking mass spectrometry, biochemistry, and evolutionary analysis to produce integrative models. Recently, with the development of AI-based tools such as AlphaFold2, structure prediction has played an increasingly important role in integrative modeling. The combination of cryo-ET and structure prediction in particular has provided unprecedented insights into the ultrastructure of cellular components. This thesis focuses on two bacterial multi-megadalton protein complexes which are difficult to study by classical structural biology approaches: gas vesicles (GVs) and the *Legionella pneumophila* Dot/Icm type IV secretion system (T4SS). GVs are gas-filled protein nanostructures that regulate the position of certain microorganisms in water and consequently their access to sunlight and nutrients. Here, we investigate the mechanical properties of GVs and reveal the molecular structure of GVs and its implication for the assembly mechanism. The Dot/Icm T4SS is a macromolecular complex formed by approximately 27 proteins, utilized by *L. pneumophila* to hijack the host cell's biology for its replication purposes. A nearly-complete integrative model of this complex provides crucial insights into its structural organization and its evolution from conjugation to secretion, as well as the transportation of substrates into the host cell.

PUBLISHED CONTENT AND CONTRIBUTIONS

Przemysław Dutka, Dina Malounda, Lauren Ann Metskas, Songye Chen, Robert C. Hurt, George J. Lu, Grant J. Jensen[#], and Mikhail G. Shapiro[#]. 2021. "Measuring Gas Vesicle Dimensions by Electron Microscopy." *Protein Science* 30 (5): 1081–86. https://doi.org/10.1002/pro.4056

> Contributions: I worked on conceptualization; methodology; investigation; formal analysis; visualization; writing—original draft preparation; writing—review & editing.

Hossein Salahshoor*, Yuxing Yao*, **Przemysław Dutka***, Nivin N. Nyström, Zhiyang Jin, Ellen Min, Dina Malounda, Grant J. Jensen, Michael Ortiz, and Mikhail G. Shapiro[#]. 2022. "Geometric Effects in Gas Vesicle Buckling under Ultrasound." *Biophysical Journal* 121 (21): 4221–28. https://doi.org/10.1016/j.bpj.2022.09.004

> Contributions: I conceived and designed the study, conducted *in vitro* experiments and analyzed the experimental data, wrote and edited the manuscript.

Przemysław Dutka, Lauren Ann Metskas, Robert C. Hurt, Hossein Salahshoor, Ting-Yu Wang, Dina Malounda, George J. Lu, Tsui-Fen Chou, Mikhail G. Shapiro[#], and Grant J. Jensen[#]. *2023*. "Structure of Anabaena Flos-Aquae Gas Vesicles Revealed by Cryo-ET." *Structure*. https://doi.org/10.1016/j.str.2023.03.011.

> Contributions: I conceived experiments, prepared samples, acquired and analyzed data, performed data exploration, drafted the manuscript, and prepared the figures.

Przemysław Dutka*, Yuxi Liu*, Stefano Maggi*, Debnath Ghosal, Jue Wang, Stephen D. Carter, Wei Zhao, Sukhithasri Vijayrajratnam, Joseph P. Vogel, Grant J. Jensen[#]. 2023. "Structure and Function of the Dot/Icm T4SS". bioRxiv. https://doi.org/10.1101/2023.03.22.533729

> Contributions: I prepared samples, reconstructed tomograms and picked particles for a subset of data, performed subtomogram averaging, performed data exploration, located proteins forming the system, participated in building the model, drafted the manuscript, and prepared the figures.

Robert C. Hurt*, Marjorie T. Buss*, Mengtong Duan*, Katie Wong, Mei Yi You, Daniel P. Sawyer, Margaret B. Swift, **Przemysław Dutka**, Pierina Barturen-Larrea, David R. Mittelstein, Zhiyang Jin, Mohamad H. Abedi, Arash Farhadi, Ramya Deshpande, and Mikhail G. Shapiro[#]. 2023. "Genomically Mined Acoustic Reporter Genes for Real-Time in Vivo Monitoring of Tumors

and Tumor-Homing Bacteria." *Nature Biotechnology*, January. https://doi.org/10.1038/s41587-022-01581-y.

Contributions: I planned and performed experiments relative to TEM imaging and helped analyze data.

Mohammed Kaplan*, Yi-Wei Chang*, Catherine M. Oikonomou, William J. Nicolas, Andrew I. Jewett, Stefan Kreida, **Przemysław Dutka**, Lee A. Rettberg, Stefano Maggi, and Grant J. Jensen#. 2022. "Dynamic Structural Adaptations Enable the Endobiotic Predation of *Bdellovibrio bacteriovorus*." bioRxiv. https://doi.org/10.1101/2022.06.13.496000.

Contributions: I performed the ribosome subtomogarm averaging, helped in the data visualization and analysis relative to the ribosome averages, and participated in the correction of the manuscript.

Mohammed Kaplan*, Doulin C. Shepherd*, Naveen Vankadari, Ki Woo Kim, Charles L. Larson, **Przemysław Dutka**, Paul A. Beare, Edward Krzymowski, Robert A. Heinzen, Grant J. Jensen#, and Debnath Ghosal#. 2022. "Structural Remodeling of *Coxiella burnetii* during Its Biphasic Developmental Cycle Revealed by Cryo-Electron Tomography." bioRxiv. https://doi.org/10.1101/2022.08.23.505044.

Contributions: I performed segmentations, helped in the data visualization, and participated in the correction of the manuscript.

William J. Nicolas, Florian Fäßler, **Przemysław Dutka**, Florian K. M. Schur, Grant Jensen, and Elliot Meyerowitz#. 2022. "Cryo-Electron Tomography of the Onion Cell Wall Shows Bimodally Oriented Cellulose Fibers and Reticulated Homogalacturonan Networks." *Current Biology*, CB 32 (11): 2375–89.e6. https://doi.org/10.1016/j.cub.2022.04.024

Contributions: I performed subtomogram averaging of the fibers, helped in the data analysis relative to the fiber averages, and participated in the correction of the manuscript.

*denotes co-first authors
#denotes corresponding authors

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Chapter 1

INTRODUCTION

Gaining insight into the relationship between protein structure and function is crucial for understanding the biological processes that occur in living organisms. One of the earliest and widely adopted techniques that allowed scientists to determine protein structure was crystallography (Jaskolski, Dauter, and Wlodawer 2014). This was possible by growing crystals of the molecule of interest and using X-ray diffraction to determine the positions of atoms within the crystal lattice. Despite being highly effective in unveiling the structures of numerous biological macromolecules, crystallography has its limitations, particularly in the case of large and complex structures. Furthermore, protein crystallization occurs in non-native conditions, which hinders probing of the protein's conformational heterogeneity and might occasionally result in the capture of conformations that, although present, are not necessarily the most prevalent ones within the cell.

The inherent limitations of crystallography prompted the development of single-particle cryo-electron microscopy (cryo-EM) (Kühlbrandt 2014). Cryo-EM overcomes the most significant drawback of X-ray crystallography, which is obtaining a well-diffracting crystal. Instead, in cryo-EM, the sample is rapidly frozen in a thin layer of ice and imaged using an electron microscope, producing high-resolution images of biological molecules in a frozen-hydrated state (preserving the protein in a near-native state). The images are then used to generate a three-dimensional density map of the macromolecule by averaging thousands of two-dimensional images captured from

different angles. Single-particle cryo-EM has revolutionized structural biology by enabling structure determination of complex macromolecules that are difficult or impossible to crystallize (Callaway 2020). This technique has revolutionized structural biology of membrane proteins and large and heterogeneous complexes and made significant contributions to our understanding of fundamental biological processes such as DNA replication (Xie et al. 2023; Jain et al. 2019; Lancey et al. 2021) and protein synthesis (Fromm et al. 2023), and is helping to advance drug discovery and design (Renaud et al. 2018). With many hardware and software developments in recent years, single-particle cryo-EM has become the method of choice for studying many difficult targets (Punjani et al. 2017; Zivanov et al. 2018).

Single-particle cryo-EM has undoubtedly been a game-changing technique; however, it does have its limitations. One such limitation is the requirement for sample purification, which can often result in the loss of valuable information regarding the cellular context. Furthermore, larger and more complex protein assemblies can be challenging to purify while maintaining their integrity. These limitations have contributed to a growing interest in cryo-electron tomography (cryo-ET), which offers the unique advantage of studying the 3-D structure of intact cells and organelles (Oikonomou and Jensen 2017). Cryo-ET involves acquiring a series of 2-D images of a sample as it is tilted, which can then be reconstructed into a 3-D volume called tomogram. This technique is particularly useful for investigating large and complex structures.

Cryo-ET has recently delivered some spectacular high-resolution structures allowing us to understand the ultrastructure of Rubisco inside alpha-carboxysomes (Metskas et al. 2022), visualize protein biogenesis at the ER membrane (Xue et al. 2022; Gemmer et al. 2023) or reveal

ribosome-antibiotic complexes inside intact cells (Tegunov et al. 2021). However, in all these cases achieving high resolution was possible either because of a large number of particles, the large size of the complex, and/or symmetry. Unfortunately, for many biological samples the resolution produced by cryo-ET is stuck at a single-digit nanometer scale. There are multiple factors limiting the resolution of cryo-ET maps (Hylton and Swulius 2021). The primary problem is that electron dose has to be distributed among many images to produce tilt-series, which results in images with significantly lower signal-to-noise ratio compared to the single-particle approach. Moreover frequently cryo-ET samples are thick and their thickness increases at higher-tilt angles, further reducing the signal-to-noise ratio. Among other problems are difficulties locating molecules of interest in the tomogram and a low number of particles per tomogram. Despite significant advances in data collection speed (Chreifi, Chen, and Jensen 2021; Chreifi et al. 2019) and software (Tegunov and Cramer 2019; Tegunov et al. 2021; Zivanov et al. 2022), it is still challenging to reach a near-atomic resolution for many biologically important targets.

The limited resolution of cryo-ET can be overcome by combining it with other methods to produce integrative models (Koukos and Bonvin 2020). In this approach, X-ray crystallography, single-particle cryo-EM, and structure predictions provide information about individual components of the system and their interactions, which can then be combined and docked into the molecular envelope obtained by cryo-ET with the help of cross-linking mass spectrometry, biochemistry, and evolutionary analysis.

In the last couple of years, structure prediction has played an increasingly important role in integrative modeling. This was largely made possible by the development of AlphaFold2 (Jumper

et al. 2021; Mirdita et al. 2022), an artificial intelligence (AI)-based tool that predicts accuratley protein structure. The combination of cryo-ET and AlphaFold has opened a new era in structural biology and has produced some groundbreaking structures. Some notable examples include *de novo* protein identification in mammalian sperm (Chen et al. 2022) and a nearly-complete model of the nuclear pore complex (Mosalaganti et al. 2022). Both of these are examples of large macromolecular protein complexes that would be difficult to purify and study with single-particle cryo-EM or crystallography. Many important biological processes are driven by similarly complex assemblies, and understanding the structure and function of these complexes is critical for gaining insights into biological processes, developing treatments for diseases associated with their dysfunction, and structure-based engineering for biotechnology.

This thesis focuses on two multi-megadalton bacterial protein complexes: gas vesicles (GVs) (A. E. Walsby 1994a; Pfeifer 2012) and the *Legionella pneumophila* defective in organelle trafficking/intracellular multiplication (Dot/Icm) type IV secretion system (T4SS) (Kubori and Nagai 2016; Hubber and Roy 2010). Chapters 2-4 focus on GVs, a unique class of bacterial gas-filled protein nanostructures found in certain microorganisms. They act as buoyancy devices that regulate the position of these microorganisms in water and their access to sunlight and nutrients (A. E. Walsby 1994a). GVs are encoded by large gene clusters and are primarily composed of the highly-hydrophobic major structural protein GvpA, which forms the rib-like shell structure of GVs. Some gene clusters also encode for the accessory structural protein GvpC, which provides additional strengthening to the shell (Hayes, Buchholz, and Walsby 1992). GVs' unique biophysical properties have enabled their use as genetically-encoded reporters and actuators of cellular function deep in tissue (Shapiro, Goodwill, et al. 2014; Bourdeau et al. 2018; Farhadi et

al. 2019; Wu et al. 2023; Farhadi et al. 2020; Lakshmanan et al. 2020; Hurt et al. 2023). Other applications take advantage of GVs' refractive index, gas permeability, and susceptibility to magnetic fields (Shapiro, Ramirez, et al. 2014; Lu et al. 2018, 2020).

Despite being discovered in the 19th century and their increasing use in bioengineering, little was known about the molecular structure of GVs or their assembly mechanism. In Chapter 2, the analysis of GVs derived from various bacteria is carried out to address discrepancies in dimensions reported in the literature, which arise from different methodologies used to measure them, and to establish accurate measurements that consolidate the existing literature. GV dimensions directly affect their mechanics, which is crucial for their applications as acoustic reporter agents. Thus, in Chapter 3, the effect of GV geometry on their buckling properties is investigated using a combination of computational and experimental approaches. This work shows that diameter is a key factor influencing GVs' mechanical properties while changes in length play a minor role. Finally, Chapter 4 describes the structure of the *Anabaena flos-aquae* GVs obtained through cryo-ET and integrative modeling based on the homologous structure of *Bacillus megaterium* GVs. The structure reveals a corrugated pattern of the shell arising from the polymerization of GvpA which changes polarity at the center of the GV cylinder - a site that may act as an elongation center. The GV cylinder is additionally reinforced by the accessory protein GvpC, which forms a helical cage around the GvpA shell.

Chapter 5 focuses on the Dot/Icm T4SS, another example of a bacterial macromolecular complex. During the infection process, *L. pneumophila* secretes over 300 proteins to hijack the host cell's biology for its own replication purposes (Kubori and Nagai 2016; Qiu and Luo 2017). The Dot/Icm

T4SS is one of the largest and most complex biological assemblies known and is formed by approximately 27 proteins. It spans across the outer and inner membranes. Despite its significance as a potential drug target, our current understanding of its atomic structure is limited to isolated subcomplexes. Due to the sheer size and complexity of this system, no successful purification attempts have been reported. Through improvements in the cryo-ET map quality and integrative modeling, a nearly-complete model of the T4SS has been constructed, providing important clues on how substrates might be transported into the host cell. Additionally, evolutionary and structural analysis reveals how a conjugative T4SS evolved its primary function from conjugation to secretion by incorporating a channel-forming protein, TssM (IcmF), and its interacting partner, TssL (DotU), from a bacterial type VI secretion system (T6SS).

Chapter 2

MEASURING GAS VESICLE DIMENSIONS BY ELECTRON MICROSCOPY

Adapted from:

Przemysław Dutka, Dina Malounda, Lauren Ann Metskas, Songye Chen, Robert C. Hurt, George J. Lu, Grant J. Jensen, and Mikhail G. Shapiro. 2021. "Measuring Gas Vesicle Dimensions by Electron Microscopy." *Protein Science: A Publication of the Protein Society* 30 (5): 1081–86. https://doi.org/10.1002/pro.4056.

Abstract

Gas vesicles (GVs) are cylindrical or spindle-shaped protein nanostructures filled with air and used for flotation by various cyanobacteria, heterotrophic bacteria, and Archaea. Recently, GVs have gained interest in biotechnology applications due to their ability to serve as imaging agents and actuators for ultrasound, magnetic resonance and several optical techniques. The diameter of GVs is a crucial parameter contributing to their mechanical stability, buoyancy function and evolution in host cells, as well as their properties in imaging applications. Despite its importance, reported diameters for the same types of GV differ depending on the method used for its assessment. Here, we provide an explanation for these discrepancies and utilize electron microscopy (EM) techniques to accurately estimate the diameter of the most commonly studied types of GVs. We show that during air drying on the EM grid, GVs flatten, leading to a ~1.5-fold increase in their apparent diameter. We demonstrate that GVs' diameter can be accurately determined by direct measurements from cryo-EM samples or alternatively indirectly derived from widths of flat collapsed and negatively stained GVs. Our findings help explain the inconsistency in previously reported data and provide accurate methods to measure GVs dimensions.

2.1 Introduction

Gas vesicles (GVs) are hollow, gas-filled protein nanostructures natively expressed in certain types of cyanobacteria, heterotrophic bacteria, and Archaea as a buoyancy aid (Walsby 1994a). Recently, it was discovered that the unique physical properties of GVs enable them to serve as genetically encodable contrast agents for ultrasound and other imaging methods, allowing deep tissue imaging of cellular function (Shapiro, Goodwill, et al. 2014; Shapiro, Ramirez, et al. 2014; Bourdeau et al. 2018; Lu et al. 2018; Farhadi et al. 2019, 2020; Lakshmanan et al. 2020). In addition, GVs are being applied to acoustic manipulation and therapeutic uses of engineered cells (Bar-Zion et al. 2019; Wu et al. 2019).

Fully formed GVs adopt two predominant shapes—cylinders with conical ends or spindle-like. The GVs may be 0.1–2 μm in length, or even longer when heterologously expressed in more spacious mammalian cells (Farhadi et al. 2019). The mean diameter of GVs isolated from different species widely varies, but is relatively constant for the same type of GV. There is an inverse correlation between diameter and critical collapse pressure (Hayes and Walsby 1986). This correlation has important evolutionary consequences. While wider GVs can provide buoyancy at a lower energetic cost, they collapse at lower pressure. This is perhaps best reflected by analyzing the widths and collapse pressure of GVs isolated from *Planktothrix spp*. from Nordic lakes of different depths (Beard et al. 2000, 1999). Three types of GVs isolated from *Planktothrix spp*. had widths of ~51, 58, and 67 nm with respective collapse pressures of 1.1, 0.9, and 0.7 MPa, allowing them to adapt to the hydrostatic pressure in different lakes (Beard et al. 2000; Dunton and Walsby 2005).

Despite the importance of GVs' diameter for their biophysical properties, there are significant discrepancies in values reported in the literature. For example, the width of GVs from *Anabaena*

flos-aquae (Ana) measured inside cells by thin-section electron microscopy (EM) was ~70 nm (Anthony Edward Walsby and Fogg 1971), which is considerably smaller than the value obtained by negative stain EM (ns-EM) for isolated GVs—136 nm (Lakshmanan et al. 2017). Similar discrepancies can be observed for GVs from *Halobacterium salinarum* (Halo), whose reported values range from 45 to 250 nm (Lakshmanan et al. 2017; Simon 1981; Offner et al. 1998; Pfeifer 2012). To some extent, these discrepancies could be explained by natural variability in diameter. However, analysis of width distributions for GVs from several species of cyanobacteria (Hayes and Walsby 1986) or *Bacillus megaterium* (Mega) (Farhadi et al. 2018) shows a narrow range. This inconsistency in diameter measurement was investigated almost 50 years ago by Walsby (Walsby and Fogg 1971). He observed that Ana GVs have a constant width of 70 nm when measured inside cells by thin-section EM, which was close to the value measured for the purified sample imaged using a freeze-etching technique (75 nm). In contrast, estimations by ns-EM ranged from 70 to 114 nm (Walsby and Fogg 1971). He suggested that the stain used in EM leads to swelling of GVs, which increases their diameter but has little effect on the length. As an alternative approach for assessing GV diameter, Walsby proposed indirect measurement based on the widths of flat collapsed GVs. The diameter of Ana GVs measured using this strategy was ~85 nm (Walsby and Bleything 1988). Archer and King gave another potential explanation for discrepancies in GV measurements. They proposed that the isolation process leads to deformations, increasing the width of GVs (Archer and King 1984). Regardless of these concerns, the diameter of GVs has been routinely assessed for isolated specimens by ns-EM.

As GVs have attracted more attention in biotechnology applications, accurate estimates of their diameter have become a critical input into GV engineering. For that reason, we investigated the discrepancies in reported GV diameters using modern microscopy tools. Using these updated

techniques, we provide measurements for the most commonly studied GVs: Ana, Mega, and Halo. For Halo, we analyzed two different GV types, which are products of the independent gene clusters p-vac and c-vac.

2.2 Results and Discussion

To more closely evaluate the behavior of stained and air-dried GVs on the EM grid, we collected projection images for different types of GVs at 0° and 50° tilt and analyzed their morphology (**Figure 2-1A, B**). Although we predicted some degree of distortions to the cylindrical shape of GVs, the observed differences were unexpectedly large. For Ana GVs, there was an average of ~55 nm width difference between measurements at these two angles. The pattern was similar for both Mega and Halo GVs, although to a different degree. This data indicates that all types of GVs flatten during the staining procedure, adopting an elliptic cylinder shape.

Certain limitations of the ns-EM technology, such as specimen flattening or stain thickness irreproducibility, were previously described (Frank 2006). However, the observed deformation of the GV protein shell is not like the typical flattening reported before, where sample was mainly affected in z-direction with little to no effect on the x,y-dimensions (Frank 2006). Since GVs produce strong contrast on EM even without staining, we decided to take advantage of this unique property and evaluate the effect of the stain. Analysis of unstained, air-dried Ana GVs samples at 0° and 50° tilts show on average ~20 nm difference in diameter (**Figure 2-1C**), which is significantly less than the stained sample, but not negligible.

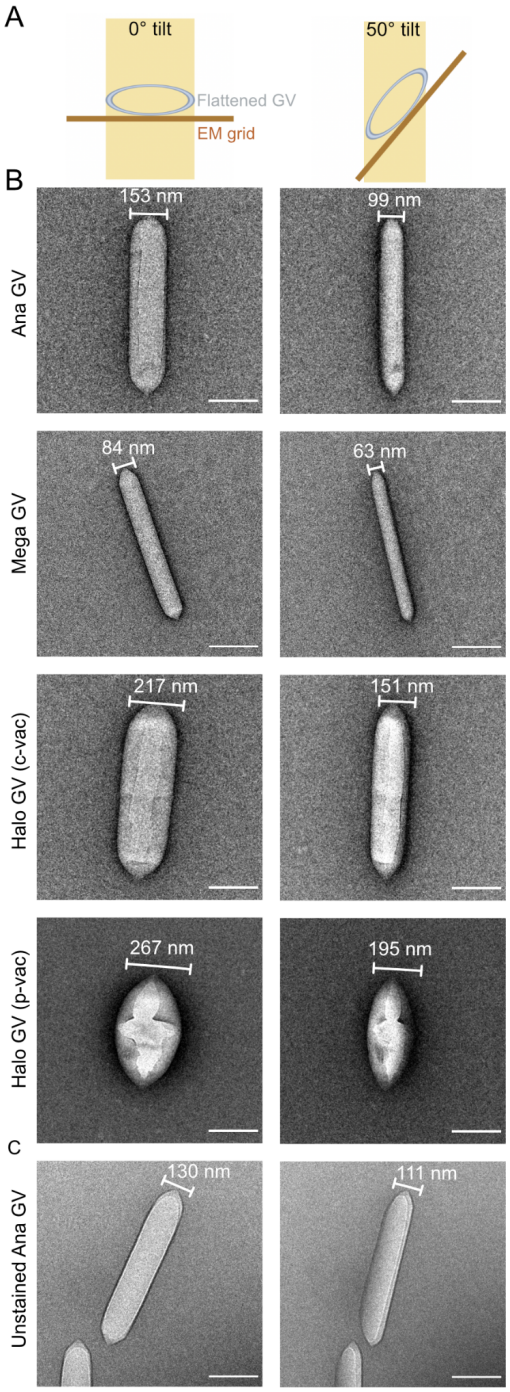


Figure 2-1. Gas vesicles (GVs) flattening on the electron microscopy (EM) grid. (A) Schematic showing cross-section of the flattened GV at 0° and 50° tilt. (B, C) Representative projection images at 0° and 50° tilt for (B) negatively stained and air-dried Ana, Mega, and Halo GV; and (C) unstained, air-dried Ana GV. Scale bar, 200 nm.

Distortions to the GV shape are the effect of the unique mechanical properties of GVs' protein shell. In ns-EM, the sample lies on a carbon support; thus, we suspect that GVs are compressed by the surface tension of evaporating water. Notably, the degree of deformation appears to be correlated with critical collapse pressure. Halo GVs, which experience the most flattening, are also the least robust among investigated GVs, with collapse pressure of 0.1 Mpa (Lakshmanan et al. 2017). In contrast, Mega GVs, which have a much higher collapse pressure of ~0.7 Mpa (Lakshmanan et al. 2017), flatten the least.

To obtain more accurate measurements of GV diameter, we used two complementary methods. First, we imaged the GVs with cryo-EM, which preserves GVs' cylindrical shape. Unfortunately, cryo-EM is a more demanding technique, requiring time-consuming sample optimization, larger sample quantities, and access to a more sophisticated instrument. Alternatively, we inferred GV diameter from the widths of flat collapsed GVs with negative staining, as measured by Walsby and Bleything (Walsby and Bleything 1988). This method, which equates the collapsed GV width with half of the intact cylindrical circumference, should allow for a faster and more accessible estimation of GV dimensions. We decided to analyze diameter distribution for Mega, Ana, and Halo GVs using both strategies.

Cryo-EM of intact GVs and collapsed ns-GV imaging resulted in similar values for each analyzed GV type (**Figure 2-2, Table 2-T1**), with differences within statistical error. Mega and Ana GVs appear to have a uniform diameter, varying within a narrow range (**Figure 2-2C, Table 2-T1**). In contrast, Halo GV diameters varied. Halo is capable of producing two types of GVs. Spindle-shaped GVs are encoded by the p-vac gene cluster located on an endogenous plasmid, while the c-vac cluster located on a mini-chromosome generates cylindrical GVs (Pfeifer 2012). According to our measurements, the diameter of both types of Halo GVs varies (**Figure 2-2C**). However,

some of this variability may be due to imperfect classification. All GV types begin their assembly as bicones, which look like smaller spindle-shape p-vac Halo GVs (Pfeifer 2012). Thus, some c-vac GVs, in their bicone phase, could have been classified as p-vac GVs. This misclassification could have made a minor contribution to the overall diameter distribution. Overall, the range of diameter values for different GV types suggest that Ana and Mega GVs have tighter regulation over diameter compared to Halo GVs. However, it is not yet known what the physiological consequences of this regulation are or how exactly the diameter is adjusted in growing GVs.

Table 2-T1. GV diameters (mean ± SD) obtained by three EM-based methods.

GV type	Intact GVs (ns-EM)[a] (nm)	Intact GVs (cryo-EM) (nm)	Collapsed GVs (ns-EM) (nm)
Mega	73 ± 14	52 ± 6	54 ± 5
Ana	136 ± 21	85 ± 4	89 ± 6
Halo (c-vac)	251 ± 51	111 ± 32	116 ± 21
Halo (p-vac)		182 ± 22	171 ± 19

Abbreviations: EM, electron microscopy; GVs, gas vesicles.
[a]Previously reported by (Lakshmanan et al. 2017)

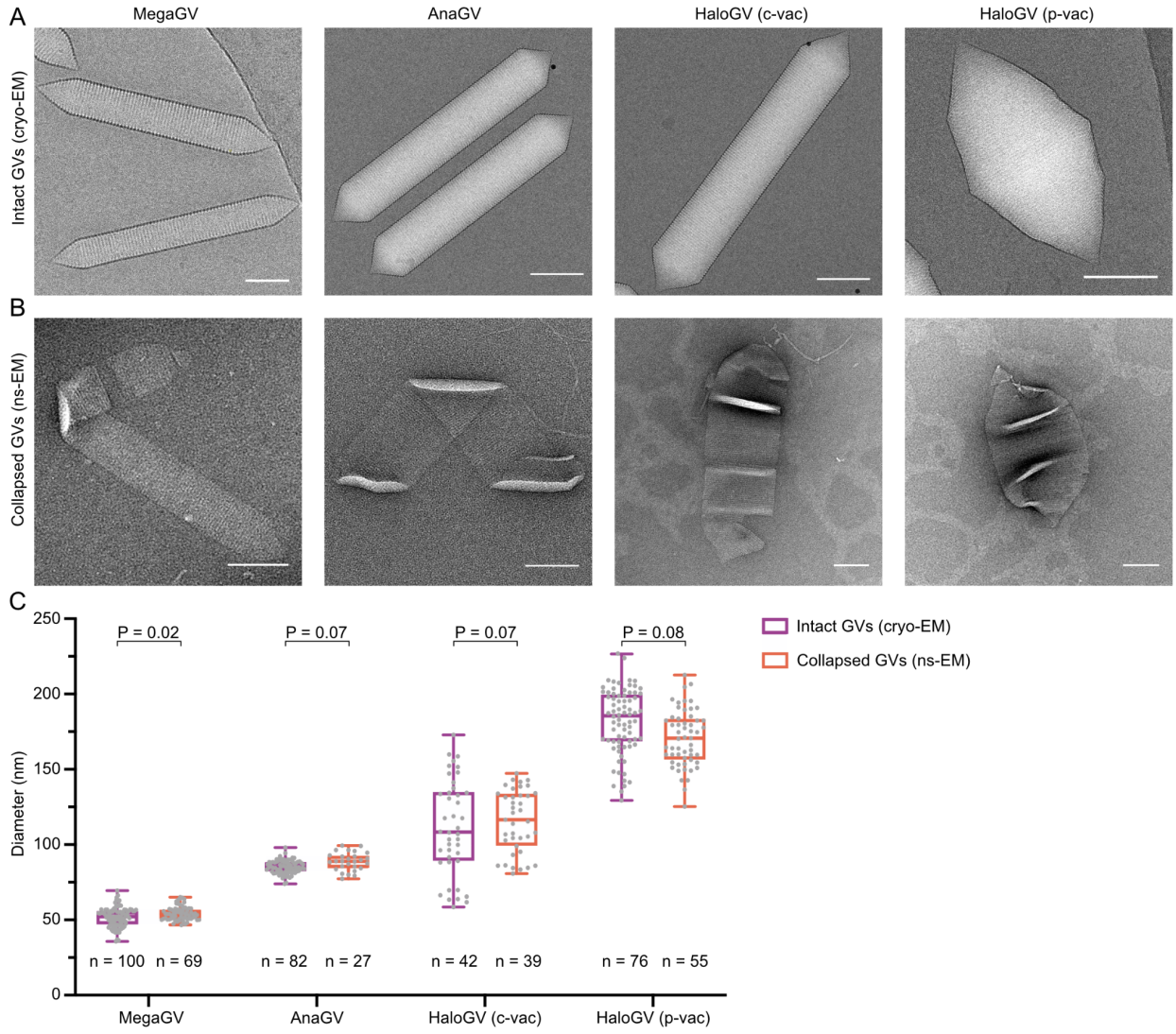


Figure 2-2. Diameter determination for Mega, Ana, and Halo gas vesicles (GVs). (**A**) Representative cryo-electron microscopy (EM) of intact GVs used for direct diameter measurement. (**B**) Representative ns-EM images of collapsed GVs used for indirect diameter assessment based on widths of flat collapsed regions. Scale bar, 100 nm. (**C**) Diameter distribution for Mega, Ana, and Halo GVs measured by cryo-EM and collapsed ns-EM. Center line indicates median, the box limits denote the interquartile range and the whiskers absolute range. Each dot represents an individual measurement. Paired t test was performed between directly measured (cryo-EM) and calculated (ns-EM) diameters for each GV type.

Taken together, our findings provide an explanation for discrepancies in previous GV diameter measurements reported in the literature. Although ns-EM is routinely used to evaluate the morphology and dimensions of intact GVs (Lakshmanan et al. 2017; Farhadi et al. 2018; N. Li and

Cannon 1998; Ramsay et al. 2011; Xu et al. 2014), our data show that this method causes GV flattening and inaccurate apparent diameter. Instead, cryo-EM of intact GVs and ns-EM of flat collapsed GVs provide correct dimensions that are mutually consistent between the two methods, as shown here for three commonly studied GVs variants.

2.3 Material and Methods

GV preparation

GVs were either isolated from native sources (Ana and Halo) or expressed heterologously in Rosetta 2(DE3) pLysS *Escherichia coli* cells (Mega) as previously described (Lakshmanan et al. 2017). In the final two or three rounds of buoyancy purification, sample buffer was exchanged to 10 mM HEPES, pH 7.5. Concentrations were measured by optical density (OD) at 500 nm using a spectrophotometer (NanoDrop ND-1000, Thermo Scientific). To prepare collapsed GV samples, diluted samples were pressurized in a sealed syringe until the solution turned transparent.

Negative stain EM

For imaging of intact GVs, the purified sample was diluted to OD500 ~ 0.5 for Ana and Halo, and OD500 ~ 0.2 for Mega. Three microliters of the target sample was applied to a freshly glow-discharged (Pelco EasiGlow, 15 mA, 1 min) Formvar/carbon-coated, 200 mesh copper grid (Ted Pella) for 1 min before blotting. Afterward, the sample was incubated for 1 min with a 0.75% uranyl formate solution before blotting and air-dried. Image acquisition was performed using a Tecnai T12 (FEI, now Thermo Fisher Scientific) EM at 120 kV, equipped with a Gatan Ultrascan 2 k × 2 k CCD.

Cryo-EM

For cryo-EM, Quantifoil R2/2 200 Mesh, extra thick carbon, copper grids (EMS) were glow discharged (Pelco EasiGlow, 10 mA, 1 min). Freshly purified Mega (OD500 ~ 1), Ana (OD500 ~ 15), and Halo (OD500 ~ 8) GVs sample was frozen using a Mark IV Vitrobot (FEI, now Thermo Fisher Scientific) (4°C, 100% humidity, blot force 3, blot time 4 s). Micrographs were collected on a 300 kV Titan Krios microscope (FEI, now Thermo Fisher Scientific) with an energy filter (Gatan) and equipped with a K3 6k × 4 k direct electron detector (Gatan). Data were collected using SerialEM software with a pixel size of either 1.4 Å (×64,000 magnification) or 2.15 Å (×42,000 magnification) and −2.5 μm defocus (Mastronarde 2005).

Diameter determination

All measurements were made using IMOD software (Kremer, Mastronarde, and McIntosh 1996). The cylinder/spindle diameter direct measurements from cryo-EM micrographs were performed only once for each GVs at its widest region. Indirectly diameter was calculated as $2w/\pi$, where w is the width of the flat collapsed GV measured from the ns-EM micrograph. Sample from at least two independent preparations were used for each measurement.

Statistical analysis was performed in GraphPad PRISM. To ensure normal distribution of the data a Shapiro–Wilk normality test, Kolmogorov–Smirnov test, and D'Agostino & Pearson test was performed. For all data sets, at least two calculated tests suggested normal distribution, thus a paired *t* test was employed.

2.4 Acknowledgments

This work was supported by the National Institutes of Health (grant R35-GM122588 to Grant J. Jensen and R01-EB018975 to Mikhail G. Shapiro) and the Caltech Center for Environmental Microbial Interactions (CEMI). Electron microscopy was performed in the Beckman Institute Resource Center for Transmission Electron Microscopy at Caltech. Related research in the Shapiro Laboratory is also supported by the Heritage Medical Research Institute, the Pew Scholarship in the Biomedical Sciences, and the Packard Fellowship for Science and Engineering.

Chapter 3

GEOMETRIC EFFECTS IN GAS VESICLE BUCKLING UNDER ULTRASOUND

Adapted from:

Hossein Salahshoor*, Yuxing Yao*, **Przemysław Dutk**a*, Nivin N. Nyström, Zhiyang Jin, Ellen Min, Dina Malounda, Grant J. Jensen, Michael Ortiz, and Mikhail G. Shapiro#. 2022. "Geometric Effects in Gas Vesicle Buckling under Ultrasound." *Biophysical Journal* **121** (21): 4221–28. https://doi.org/10.1016/j.bpj.2022.09.004

Abstract

Acoustic reporter genes based on gas vesicles (GVs) have enabled the use of ultrasound to noninvasively visualize cellular function in vivo. The specific detection of GV signals relative to background acoustic scattering in tissues is facilitated by nonlinear ultrasound imaging techniques taking advantage of the sonomechanical buckling of GVs. However, the effect of geometry on the buckling behavior of GVs under exposure to ultrasound has not been studied. To understand such geometric effects, we developed computational models of GVs of various lengths and diameters and used finite element simulations to predict their threshold buckling pressures and postbuckling deformations. We demonstrated that the GV diameter has an inverse cubic relation to the threshold buckling pressure, whereas length has no substantial effect. To complement these simulations, we experimentally probed the effect of geometry on the mechanical properties of GVs and the corresponding nonlinear ultrasound signals. The results of these experiments corroborate our computational predictions. This study provides fundamental insights into how geometry affects the sonomechanical properties of GVs, which, in turn, can inform further engineering of these nanostructures for high-contrast, nonlinear ultrasound imaging.

3.1 Introduction

Ultrasound imaging has demonstrated tremendous potential for monitoring biological processes due to its deep tissue penetration and noninvasive operation. Recently, the gas vesicle (GV)—a unique genetically encoded, gas-filled, protein-shelled nanostructure—was developed as a new type of contrast agent (Shapiro, Goodwill, et al. 2014; Lakshmanan et al. 2016), reporter gene (Bourdeau et al. 2018; Farhadi et al. 2019; Hurt et al., n.d.), and biosensor (Lakshmanan et al. 2020) to connect ultrasound images to dynamic biological activities such as gene expression and enzyme activity. To enable the sensitive detection of GVs in intact animals, imaging techniques must overcome the background linear scattering of tissues. This task is accomplished by ultrasound pulse sequences, such as amplitude modulation, which exploit the ability of GVs to produce nonlinear ultrasound scattering (Maresca et al. 2017; Maresca, Sawyer, et al. 2018; Rabut et al. 2021). This ability hinges on the mechanical buckling of GVs—an abrupt transition in mechanical response due to an external load. Specifically, above a threshold acoustic pressure known as the buckling pressure, the protein shell of a GV abruptly undergoes mechanical instability by exhibiting large, reversible deformations, which, in turn, lead to nonlinear scattering of ultrasound waves (Maresca et al. 2017; Cherin et al. 2017; Zhang et al. 2020). Previous work has shown that the protein composition of the GV shell can affect GV mechanical properties and acoustic buckling behavior (Lakshmanan et al. 2016, 2020; Cherin et al. 2017). However, the effect of GV geometry on buckling mechanics and ultrasound responsiveness remains uncharacterized. Distinct classes of GVs exhibit different characteristic dimensions with respect to length and diameter (Dutka et al. 2021), and the distribution of these parameters can depend on the cell type expressing the GVs (Farhadi et al. 2019).

In this work, combining computational modeling and experimental studies, we systematically investigate how the geometry of cylindrical GVs can affect their buckling behavior upon application of ultrasound pressure. Based on the dimensions of wild-type GVs obtained from cryogenic-electron microscopy (cryo-EM), we developed a series of finite element models of GVs, each with a distinct length or diameter. Our computational simulations predict that the diameter, rather than the length, can significantly alter the threshold buckling pressure of GVs under ultrasound. We then aimed to corroborate these predictions through experiments. To this end, we sorted GVs expressed by cyanobacteria into different populations based on diameter and recorded their respective nonlinear ultrasound scattering. We show that GVs with larger diameters exhibit stronger scattering of nonlinear ultrasound signals for a given acoustic pressure. This work reveals a fundamental relationship between GV geometry and buckling behavior, which provides guidance for the engineering of GVs with different sonomechanical characteristics for enhanced ultrasound imaging and potential multiplexed detection (Lakshmanan et al. 2016).

3.2 Results and discussion

Computational analysis

We first investigated the effect of geometric features of GVs on their buckling response under ultrasound. We chose wild-type GVs expressed by cyanobacterium AnaS as a model system (Ana GVs) due to their common use in ultrasound studies. The shell wall of Ana GV is made of GvpA, a primary GV structural protein, and GvpC, a secondary GV structural protein (A. E. Walsby 1994a). Previous experiments showed that stripped Ana GVs (AnaS), in which GvpC units have been selectively removed or digested, buckle and scatter nonlinearly above a certain acoustic pressure (Lakshmanan et al. 2016, 2020; Maresca et al. 2017; Maresca, Sawyer, et al. 2018). As

described in the material and methods section, for finite element analysis, we modeled the buckling of a stripped GV subjected to ultrasound overpressure. We first conducted simulations using a GV with an average length and diameter of 500 and 85 nm, respectively, which correspond to the average dimensions of wild-type Ana GVs (Dutka et al. 2021). We conducted an LBA, in which an eigenvalue problem is formulated upon the construction of the pertinent stiffness and mass matrix. We solved this problem using the Lanczos algorithm and obtained the first 10 modes of buckling. **Figure 3-S3** depicts these buckling modes, with the first threshold buckling pressure predicted to occur at 332 kPa. Next, we solved the deformed postbuckling configurations and validated the results of the LBA. The compliant nature of the GV protein shell leads to large deformations upon buckling, which requires nonlinear analysis to resolve. The combination of a compliant protein shell and subsequent nonlinear geometric effects under ultrasound results in an output of ill-conditioned tangent matrices. To compute threshold buckling pressures under these conditions, we utilized a dynamic relaxation approach through explicit analysis. To compute the threshold buckling pressures for each buckling mode obtained, we conducted a series of simulations, independent of the LBA analysis, for an individual stripped GV, where the overpressure varies over a period of 1 μs, starting at 100 kPa and increasing in steps of 20 kPa until a pressure that causes the structure to buckle. Each simulation was designed with a total simulation time of 1 μs at 11.4 MHz frequency.

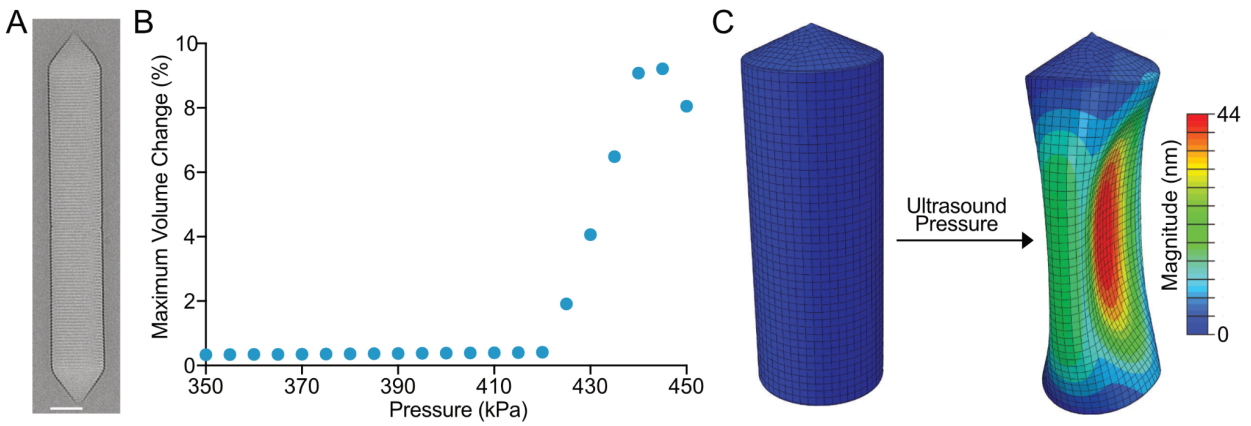


Figure 3-1. Geometric characterizations and computational modeling of GVs. (**A**) Representative cryo-EM image of a stripped GV (AnaS) isolated from cyanobacterium *Anabaena flos-aquae*. Scale bar, 50 nm. (**B**) Maximum percentage volume change in a GV as a function of applied ultrasound pressure. The sudden departure from a linear response indicates the onset of buckling in a GV, which is reminiscent of pitch-fork instability in bifurcation theory (Thompson 2015). (**B**) Depictions from a finite element model of a GV with length and diameter dimensions of 500 and 85 nm, respectively. Both the initial configuration (left) and the buckled configuration at 331 kPa (right) are depicted.

We quantified GV deformations by measuring the change in volume, which, prior to the onset of buckling, increases negligibly with externally applied cycles of ultrasound pressure. At the threshold pressure for the onset of buckling, an abrupt transition occurs in the GV deformation mechanics. Notably, this transition may not occur in response to all the cycles within an ultrasound pulse, due to the tapered nature of pulse amplitudes and to the nonlinear geometric effects of a GV exposed to ultrasound, which may induce the onset of buckling only after the GV experiences a few cycles of ultrasound pressure. We then identified the exact threshold buckling pressure that causes this nonlinear response, within a narrow range of 1 kPa, via a bisection method in which we ascertain the interval that contains the threshold buckling pressure by repeatedly bisecting each pressure interval and selecting the subinterval in which buckling commences. This bisection algorithm determined the threshold buckling pressure of AnaS GVs with dimensions of 500 nm in length and 85 nm in diameter to be 331 kPa (**Figure 3-1**).

Notably, it is possible that accounting for nonlinear deformations using explicit dynamic analysis based on volumetric changes may lead to a threshold buckling pressure lower than that obtained via LBA. These nonlinear deformations can accommodate buckling at pressures below the values obtained from LBA. We also conducted several numerical tests covering a range of mesh sizes and verified that the results of our calculations were not affected by the discretization resolution (**Figure 3-S4**).

Geometry-dependent GV buckling

After validating a computational model that captures the ultrasound-induced buckling of a GV with fixed geometry, we aimed to model the effect of different GV lengths and diameters on the threshold buckling pressure. To this end, we conducted a thorough sensitivity analysis using an exhaustive search approach, in which we created several distinct computational GV models, each of which having identical material properties, boundary conditions, and loading conditions, including an identical ultrasound pressure waveform. In these computational models, we also fixed the GV length (or diameter) and varied the GV diameter (or length) across a physiologically relevant range of values (Dutka et al. 2021; Lakshmanan et al. 2017). In each of the corresponding finite element models, the element type and the mesh size remained invariant, leading to a different number of elements and nodes across models. Additionally, for each model, with the details delineated in the previous subsection, both LBA and explicit dynamic analysis were conducted.

We first investigated the dependence of threshold buckling pressures on GV diameter. We created two sets of models with fixed GV lengths of 300 or 500 nm, and in each set of models, we simulated a physiologically relevant range of GV diameters (Dutka et al. 2021; Lakshmanan et al.

2017). The results of the simulations are shown in **Figure 3-2** with representative snapshots of the buckled GV configuration, which demonstrates that varying the GV diameter substantially impacts the threshold buckling pressure value. We quantified this dependency using a curve fit that is defined as $P = AD^{\alpha} + B$, with P and D being the buckling pressure and the GV diameter, respectively, and A and B being fitting parameters. We consequently obtained a value of $\alpha \cong -3$.

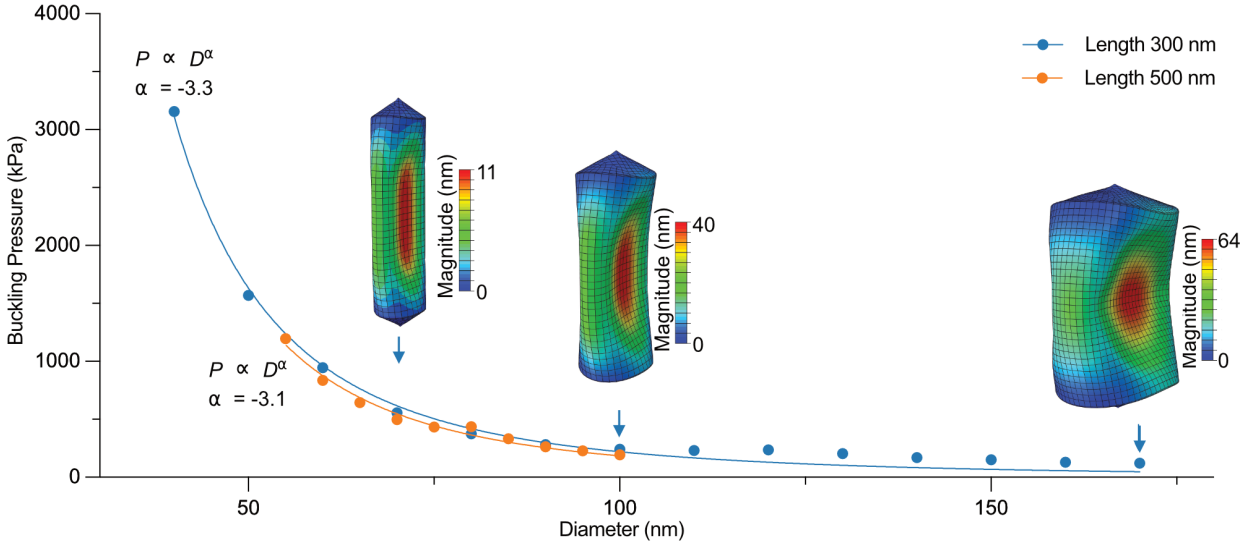


Figure 3-2. Diameter sensitivity analysis of GV buckling. The effect of GV diameter on the threshold buckling pressure at two fixed GV lengths: 300 (blue) and 500 nm (orange). Diagrams from simulations illustrate the buckled configuration of GVs with a fixed length of 300 nm and different diameters of 70 (left), 100 (middle), and 170 nm (right).

Next, we investigated the dependence of threshold buckling pressures on GV length. **Figure 3-3** shows the results of simulations conducted for two distinct GV diameters, 60 and 83 nm, with illustrative depictions of buckled configurations for three representative GVs. In dramatic contrast to our results with varying diameters, the length sensitivity analysis shows that the threshold buckling pressure is virtually unaffected by differences in GV length over the typical range exhibited by AnaS GVs. This result is apparent with the exponent α being close to zero, obtained by fitting a function of the form $P = AL^{\alpha} + B$ to the data, where P and L are the threshold

buckling pressure and the GV length, respectively. We note that the ranges of lengths that we have incorporated correspond to Anabeaba GVs, and in GV morphologies, such as spindles and bicones, we anticipate the existence of a critical length, below which it would be the driving factor in GV buckling.

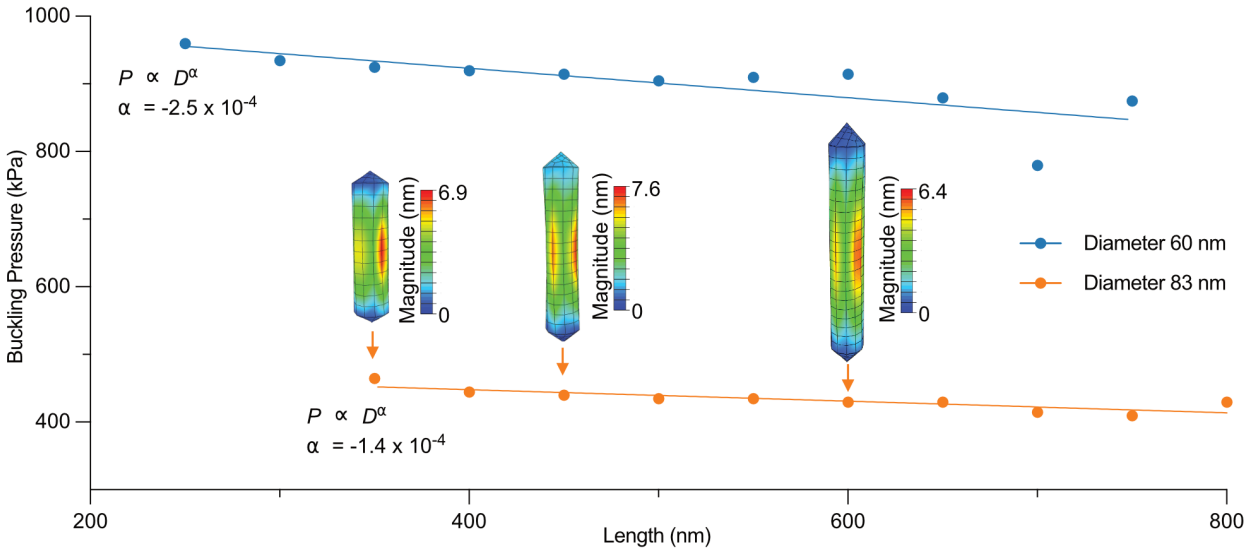


Figure 3-3. Length sensitivity analysis of GV buckling. The effect of GV length on the threshold buckling pressure at two fixed GV diameters: 60 (blue) and 83 nm (orange). Diagrams from simulations illustrate the buckled configurations of GVs with a fixed diameter of 83 nm and different lengths of 350 (left), 450 (middle), and 600 nm (right).

Considering the unsubstantial effect of GV length on the onset of buckling, we remark that the theory of cylindrical shells can help interpret our results for GV buckling. By examining the buckling theory of a shell subjected to external pressure, as well as the solutions of the corresponding eighth-order governing differential equation (also known as the Donnell stability equation), we determined the results for the limiting case of $L >> D$. For an isotropic shell, it can be shown that the buckling pressure P satisfies $P = \frac{EI}{D^3 (1 - \nu^2)}$, with E, ν, and I representing the Young's modulus, the Poisson's ratio, and the moment of inertia of the cross section, respectively (Timoshenko, Woinowsky-Krieger, and Others 1959). Although our computational models of

GVs account for an anisotropic finite-length shell with conical ends, by going to high aspect ratios, the agreement of the integer parts of the exponents obtained from our simulations with those obtained from the idealized shell theory further posits that the diameter is the dominant dimensional feature influencing GV buckling.

Experimental validation

To experimentally validate the geometry-buckling relationship revealed by our simulation results, we first fractionated AnaS into different size distributions. Given that these GVs are expressed in a single species of cyanobacteria harboring the same gene cluster, we assume the material properties of the major structural protein of the shell (GvpA) to be the same and not dependent on GV geometry. To obtain a different size distribution of AnaS, we slowly increased the hydrostatic pressure around AnaS, leading to the irreversible collapse of some GVs, and characterized the geometry of the remaining GVs. Since previous studies showed a correlation between the threshold acoustic buckling pressure of a GV and its hydrostatic collapse pressure (Lakshmanan et al. 2016, 2020), our simulation results led us to hypothesize that GVs surviving higher pressures without collapse would have smaller diameters and generate less buckling-induced nonlinear ultrasound contrast. Because collapsed GVs do not scatter light as intact GVs do, we quantified the number of intact GVs remaining by measuring the OD500 after exposure to different hydrostatic pressures. The OD500 of AnaS remained unchanged when exposed to low hydrostatic pressure, and it significantly decreased above a certain pressure until all GVs collapsed. By setting the applied hydrostatic pressure at 200 kPa, approximately 30% of GVs remain intact, whereas at a pressure of 220 kPa, only ~10% GVs remain intact (**Figure 3-4A**). Notably, GVs that remain intact at 200 kPa have a higher hydrostatic collapse pressure than the original GV population, indicating that our hydrostatic pressure treatment at 200 kPa successfully selected GVs that are mechanically

more robust and resist higher hydrostatic pressures. The absolute length and diameter distributions of pressure-treated GV samples were characterized by cryo-EM (**Figure 3-4B**). We found that the length distribution of GVs does not change significantly and is independent of pressure treatment (**Figure 3-4C**). However, it is clear that increasing applied hydrostatic pressure led to smaller average diameters in remaining intact GVs (**Figure 3-4D**). Specifically, a 200 kPa precollapse step destroys any GVs with a diameter larger than 90 nm. This observation supports the prediction that the mechanical properties of GVs depend on the diameter—but not length—resulting in significantly different susceptibility to hydrostatic pressure.

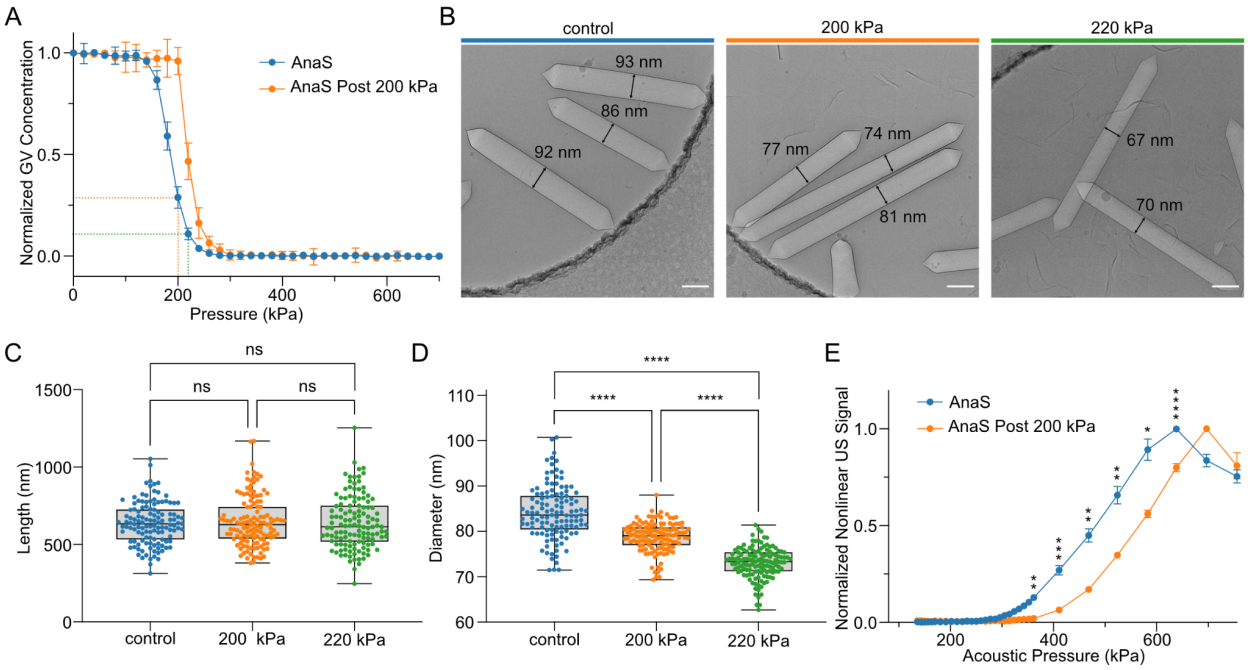


Figure 3-4. Experimental validation of the GV geometry-buckling relationship. (a) Hydrostatic collapse pressure curves for stripped GVs isolated from cyanobacterium Anabaena flos-aquae (AnaS), with (orange) and without (blue) precollapse hydrostatic pressure treatment at 200 kPa. Dashed lines indicate the pressure and corresponding OD_{500} for samples that were collected for cryo-EM and nonlinear ultrasound imaging analysis. (b) Representative cryo-EM images of AnaS used to measure lengths and diameters of GVs from the control sample (left) and after incubation at a hydrostatic pressure of 200 (middle) and 220 kPa (right). Scale bars, 100 nm. (c and d) Length (c) and diameter (d) distributions of the intact GV fraction after exposure to the indicated hydrostatic pressure. Center line indicates median, the box limits denote the interquartile range and the whiskers absolute range. Each dot represents an individual measurement. Asterisks

indicate statistical significance by one-way ANOVA tests (****$p < 0.0001$); ns, no significance. (e) Nonlinear ultrasound signals from AnaS (n = 4) as a function of acoustic pressure from samples with (orange) and without (blue) precollapse hydrostatic pressure treatment at 200 kPa. Asterisks represent statistical significance by unpaired t-tests (****$p < 0.0001$, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$). Error bars represent mean ± SEM, where not seen, are hidden by symbols.

The sonomechanical buckling behavior of GVs with different size distributions was studied using nonlinear ultrasound imaging, which detects nonlinear scattering signals generated by GV buckling (**Figure 3-S5**). As hydrostatic pressure treatment of GV samples at 220 kPa reduces the number of GVs below the level needed for reliable ultrasound imaging, we proceeded with imaging only AnaS without pretreatment or after pretreatment with 200 kPa. We found that, at the same concentration of intact GVs, pressure-treated GVs require a higher threshold pressure to generate detectable nonlinear signal compared with GVs that did not undergo precollapse treatment (**Figure 3-4** and **3-S6**). This set of results agrees well with our modeling prediction that GVs with larger diameters buckle at lower threshold pressures and would thus be expected to generate nonlinear signals at lower pressures compared with GVs with smaller diameters. Moreover, when the acoustic pressure is increasing, a larger portion of GVs will be able to buckle, therefore contributing to a higher nonlinear signal (**Figure 3-S7**). The apparent experimental buckling thresholds were 300 and 350 kPa for untreated AnaS and precollapsed AnaS, respectively. These experimental values are not far from the threshold buckling pressures predicted by our model (263 and 331 kPa, respectively) based on the largest diameter observed in a sample population of GVs, supporting the general validity of our simulations. The fact that our experimental values for threshold buckling pressure are slightly larger than the computationally predicted values can be explained by the fact that only a small fraction of GVs possess the largest diameter observed in a given sample population, and the sample may therefore not generate a

detectable amount of ultrasound signal until GVs with smaller diameters start to buckle at higher pressures. Notably, the pressure pretreated GV sample exhibited a peak nonlinear ultrasound signal at a higher pressure (above which the signal declines due to acoustic collapse of the GVs) than the GV sample not subjected to precollapse treatment, again suggesting that the pressure required to collapse GVs becomes higher after precollapse treatment (**Figure 3-S8**). Experimental validation further supports the correlation between the hydrostatic collapse pressure and threshold acoustic buckling pressure: GVs with lower hydrostatic collapse pressures tend to buckle at lower acoustic pressures and generate higher x-AM signals than GVs with higher hydrostatic collapse pressures under the same ultrasound conditions, a result that has also been observed in other studies (Lakshmanan et al. 2016, 2020).

3.3 Conclusion

The sonomechanical buckling properties of GVs were systematically investigated through finite element simulations and experiments. Computational results predicted that the GV diameter, but not the length, strongly influences the buckling behaviors of GV. We have determined that there is an inverse cubic relation between the threshold buckling pressure and the GV diameter. Above the threshold buckling pressure, ultrasound is predicted to induce large deformations of the GV shell, which agrees with the experimentally observed nonlinear acoustic backscattering response of GVs. Our computational models and analysis were corroborated by the results of experiments using nonlinear ultrasound imaging of GVs having the same genotype but different size distributions. Our results elucidate the effect of geometry on the sonomechanical buckling of GVs, which has the potential to guide future engineering of GVs as highly sensitive and specific ultrasound contrast agents, reporter genes, and bio- sensors, resulting in the advancement of high-

precision, nonlinear imaging. In addition, mechanical insights into GV interactions with ultrasound waves may benefit other GV-enabled technologies such as acoustic manipulation of engineered cells and cell-based therapeutics (Wu et al. 2019; Bar-Zion et al. 2021).

3.4 Materials and Methods

Computational modeling of GV buckling

We developed a finite element model of a single stripped GV (Lakshmanan et al. 2017) isolated from *Anabaena flos-aquae* (AnaS), in which we adopt the GV shape and geometry from a cryo-EM image (**Figure 3-1**). The adopted geometry consists of a cylindrical shell with conical ends. In view of experimental observations (**Figure 3-S1A**), we assume a uniform GV diameter within the cylindrical segment of the protein shell. We model the protein wall as a continuum shell with a thickness of 2.4 nm and a shell density of 1350 kg/m3 (Lakshmanan et al. 2017; Pfeifer 2012; Dutka et al. 2022). In order to account for the rib-like structure of the GV wall, we incorporate an elastic anisotropic material model, with elastic moduli across and along the principal axis of the GV of 0.98 and 3.92 GPa, respectively (Maresca et al. 2017). We also assign a Poisson's ratio of 0.499, which produces the desired incompressibility. While the material parameters are not obtained from direct experimental measurements, and since we aim to study the geometric effects, having values that lie within a range of parameters consistent with those of protein-based biological materials (Gosline et al. 2002) is sufficient for our purposes. The model is then discretized using shell elements. We subject the exterior and the interior surfaces of the GV to an initial pressure of 101 kPa, modeling both the inner gas pressure and the pressure of the surrounding environment. Given that the size of GVs and the wavelengths of ultrasound pulses considered in this study are below 1 μm and above 100 μm, we assume that the pressure

experienced by GVs are isotropic and uniform. Also, we assume that the ultrasound-induced shear waves in the liquid-like surrounding environment of GVs are negligible. Moreover, we note that the acoustic radiation force and the effects thereof are not considered in this study, as the acoustic radiation force on a single GV at the frequency and pressures utilized in this study are exceedingly smaller than the normal force arising from the primary acoustic pressure (Wu et al. 2019). To prevent rigid body modes in our simulations, in which the entire GV structure would undergo translations and rotations without any elastic deformation, we subject the vertices at both the top and bottom conical ends of the GV to the zero displacement Dirichlet boundary condition. We have confirmed that these boundary conditions will not affect the buckling modes and pressure. We aim to characterize the buckling pressure of GVs, where buckling refers to a mechanical instability at which a sudden abrupt deformation occurs. We note that our framework could be applied to other GV-like particles, where characterizing the reversible buckling is of interest. We first conduct linear buckling analysis (LBA) and solve the corresponding eigenvalue problem to obtain the threshold buckling pressures. LBA's computational cost is several orders of magnitude less than other computational methods for buckling analysis and is useful when only characterizing the buckling pressure is of interest. Upon the onset of buckling, the soft protein shell undergoes large deformations, which cannot be resolved using linear analysis. We therefore solve for postbuckling configurations using explicit dynamic analysis, which is a particularly powerful technique when a computational problem includes measures of discontinuity, such as buckling, in the solution (Bisagni 2000; Lanzi 2004). In this analysis, the governing equilibrium equations are solved by an explicit integration operator, using an explicit central finite difference scheme. For the explicit dynamic analysis, two reasonable assumptions are made to account for the inner gas pressure dynamics. First, we assume an isothermal buckling mechanism. Second, we neglect

diffusion of gas across the GV shell, and treat the encapsulated gas as trapped within a GV, since the gas efflux time is substantially longer than an ultrasound cycle at the frequency used in this study (Walsby et al. 1992; Kunth et al. 2018). Using these assumptions, we simulate the GV response to acoustic excitation by applying an additional oscillatory overpressure in the form of a tapered sine-burst pulse amplitude signal applied for 1 μs at a frequency of 11.4 MHz, which is a typical ultrasound setting used experimentally for imaging and is well below the resonance frequencies of GVs (Zhang et al. 2020). We have also confirmed that by conducting a separate computational frequency analysis, where we obtained the lowest resonance frequency to be 328 MHz. **Figure 3-S2** shows the first four resonant modes and the corresponding frequencies. Moreover, in the explicit dynamic analysis, we introduce numerical bulk viscosity damping to eliminate numerical artifacts and to smear nonphysical oscillations in the solutions obtained by utilizing linear and quadratic damping coefficients equal to 0.06 and 1.2, respectively (Shim et al. 2012). The selected element size in the discretization of the model is at least one tenth of the dilatational wavelength, and time steps are automatically incorporated to ensure satisfaction of the Courant-Friedrichs-Lewy stability criterion (Achenbach 2012). Calculations for both LBA and explicit dynamic analysis are carried out using Abaqus (Dassault Systèmes Simulia, France).

GV preparation and quantification

GVs were purified from AnaS as previously described (Lakshmanan et al. 2016, 2017). 6 M urea solution was added to purified native GVs, and two subsequent rounds of centrifugal flotation and removal of subnatant were preformed to prepare stripped GVs (AnaS). Two rounds of dialysis in PBS were performed to exchange the media. We determined the concentration of GVs by measuring the optical density at 500 nm (OD500) with a spectrophotometer (NanoDrop ND-1000,

Thermo Fisher Scientific, Waltham, MA, USA). For AnaS, OD500 = 1 corresponds to a concentration of 184 pM or a volume fraction of 0.04% of GVs in an aqueous suspension.

Cryo-EM characterization and image analysis

The geometry of AnaS samples subjected to precollapse pressures was characterized using cryo-EM as described before (Dutka et al. 2021). A 3-μL volume of a sample with OD500 = ~5 was applied to C-Flat 2/2-3C grids (Protochips, Cary, NC, USA) that were freshly glow discharged (Pelco EasiGlow, 10 mA, 1 min, Pelco, Fresno, CA, USA). GV samples were frozen using a Mark IV Vitrobot (FEI, now Thermo Fisher Scientific) (4°C, 100% humidity, blot force 3, blot time 4 s). Micrographs were collected on a 200 kV Talos Arctica microscope (FEI, now Thermo Fisher Scientific) equipped with a K3 6k × 4k direct electron detector (Gatan, Pleasanton, CA, USA). Multiframe images were collected using SerialEM 3.39 software (Mastronarde 2005) with a pixel size of 1.17 Å (36,000× magnification) and a defocus of –2.5 μm. Super-resolution movies were corrected for gain reference, binned by a factor of 2, and motion corrected using MotionCor2 (Zheng et al. 2017). GV dimensions were measured using IMOD 4.12 (Kremer, Mastronarde, and McIntosh 1996). Statistical analysis was performed in GraphPadPRISM.

GV diameter consistency analysis

To quantify the stability of the diameter of individual GVs from multiple cryo-EM images, we selected start and end coordinates for individual GVs and subsequently cropped the cylindrical GV tube into segments with ~10 nm distance using RELION (Zivanov et al. 2018). To obtain accurate estimates of GV diameter, we analyzed density profiles for each segment located in the central section of the GV tube using Fiji (Schindelin et al. 2012) (**Figure 3-S1A**). To evaluate diameter

consistency, we calculated the standard deviation of each GV diameter as a percentage of the mean diameter (**Figure 3-S1B**).

Collapse of GVs with defined pressure

A sample of purified AnaS with OD500 = ~20 was loaded in a sealed flow-through quartz cuvette (Hellma Analytics, Plainview, NY, USA) connected to a pressure controller (Alicat Scientific, Tucson, AZ, USA) with N2 gas supplied to apply a headspace overpressure. Nitrogen gas was chosen because it is chemically inert and easily accessible. The pressure was slowly increased by 20 kPa at each step, and the OD500 was measured with a spectrophotometer (EcoVis, OceanOptics, Winter Park, FL, USA).

Ultrasound imaging of GVs and image analysis

10 μL GVs were dispersed in 10 μL 1% (mass/volume) agarose in PBS and loaded into a homemade gel phantom made of 1% agarose, with a final OD500 = 2 measured with a spectrophotometer (NanoDrop ND-1000, Thermo Fisher Scientific). A Verasonics Vantage programmable ultrasound scanning system with an L22-14v 128-element linear array transducer (Verasonics, Kirkland, WA, USA) transmitting at 15.6 MHz was used to perform ultrasound imaging. The gel phantom and transducer tip were both immersed in a volume of PBS to conduct imaging. A customized nonlinear ultrasound imaging protocol, namely cross-amplitude modulation (x-AM) (Maresca, Sawyer, et al. 2018), was used to specifically characterize the nonlinear contrast of GVs at a distance of 5 mm from the transducer and eliminates any nonlinear ultrasound propagation within the medium. Specifically, an automated voltage ramp script implemented in MATLAB was used to acquire x-AM signals at each specified voltage step ranging

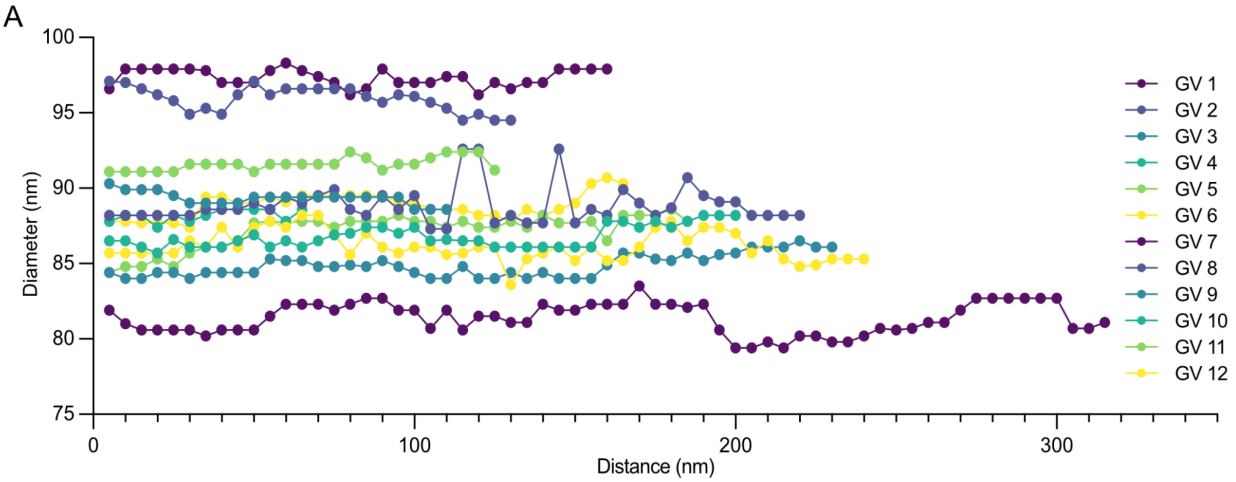
from 1.6 (corresponding to a peak positive pressure of 150 kPa) to 10 V (corresponding to a peak positive pressure of 734 kPa) with 0.5-V increments. The transmitted pressure level was calibrated using a fiber-optic hydrophone (Precision Acoustics, Dorchester, UK), and the peak positive pressure was termed "acoustic pressure," as shown in **Figure 3-4**. Since the wavelength is more than 100 times larger than the vesicle size, we assumed that the GV is experiencing an isotropic uniform pressure and that pressure waves from GV buckling would also be approximately isotropic.

3.5 Acknowledgments

The authors are grateful to Ngozi A. Eze for the helpful editorial comments. The authors are also grateful to Dr. Di Wu for insightful discussions and input. This research was supported by the National Institutes of Health grant R01-EB018975. Related research in the Shapiro lab is supported by the Packard Foundation, The Pew Charitable Trusts, and the Chan Zuckerberg Initiative. Cryo-EM was performed at the Beckman Institute Resource Center for Transmission Electron Microscopy at Caltech. M.G.S. is an investigator of the Howard Hughes Medical Institute (HHMI).
This article is subject to HHMI's Open Access to Publications policy. HHMI Investigators have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

3.6 Supplementary Data

A



B

ID	Average diameter (nm)	Standard deviation (nm)	SD as diameter fraction (%)
GV 1	81.4	1.0	1.3
GV 2	95.9	0.8	0.9
GV 3	89.3	0.4	0.5
GV 4	86.8	0.7	0.8
GV 5	91.6	0.4	0.5
GV 6	86.2	1.0	1.1
GV 7	97.3	0.6	0.6
GV 8	88.9	1.2	1.4
GV 9	84.9	0.7	0.9
GV 10	88.2	0.4	0.5
GV 11	87.2	1.1	1.3
GV 12	88.9	0.8	0.9
		Average (%)	0.8

Figure 3-S1. Diameter consistency analysis. (A) Diameter measurements of individual GVs. For each individual GV, the diameter was measured at increments of 10 nm along the main axis. (B) Table of summarized data showing diameter consistency across individual GVs.

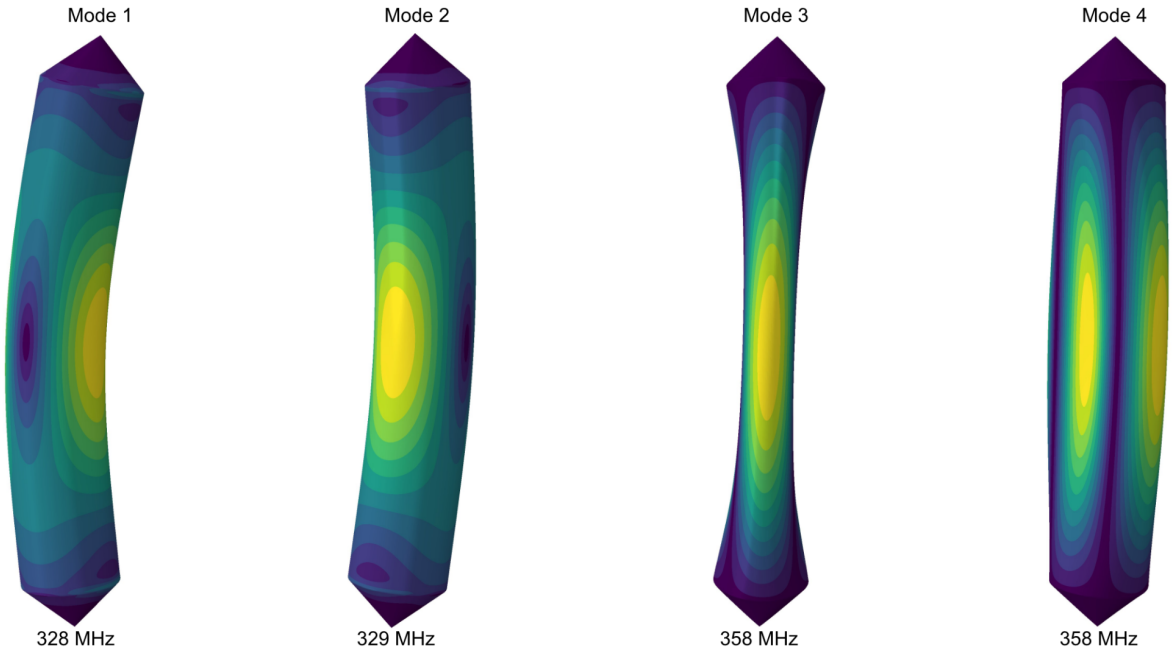


Figure 3-S2. Resonant frequencies of gas vesicle with length and diameter of 500 nm and 85 nm, respectively.

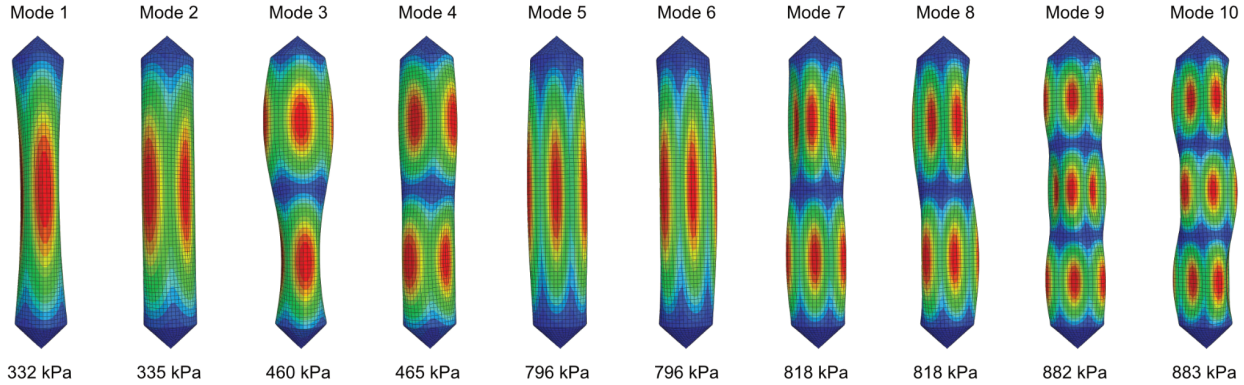


Figure 3-S3. The first ten modes of buckling (i.e., eigenvectors) and the corresponding threshold buckling pressures (i.e., eigenvalues) obtained through linear buckling analysis (LBA).

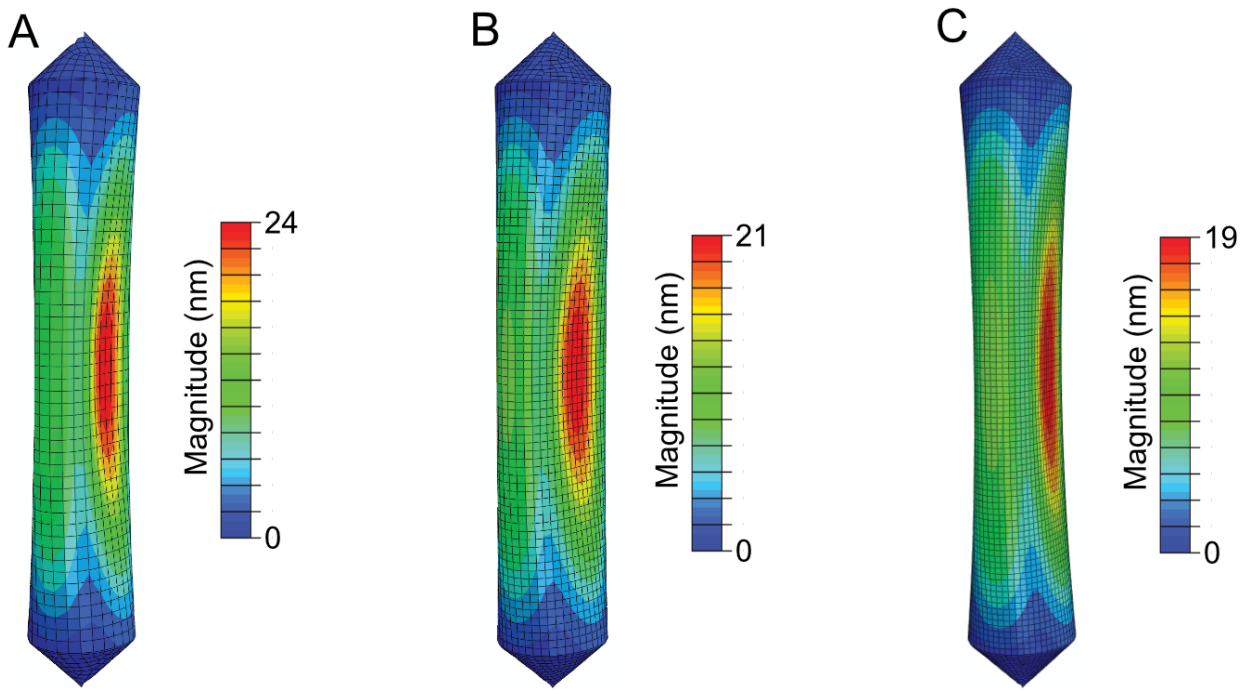


Figure 3-S4. Mesh sensitivity analysis. GV buckling simulations (with GV length and diameter set to 500 nm and 85 nm, respectively) are carried out using three different mesh sizes. A buckled configuration is depicted for each of the three models with discretization lengths as follows: (a) 9 nm, (b) 6 nm, and (c) 4.5 nm. Our results show that all three models predict an identical threshold buckling pressure of 331 kPa.

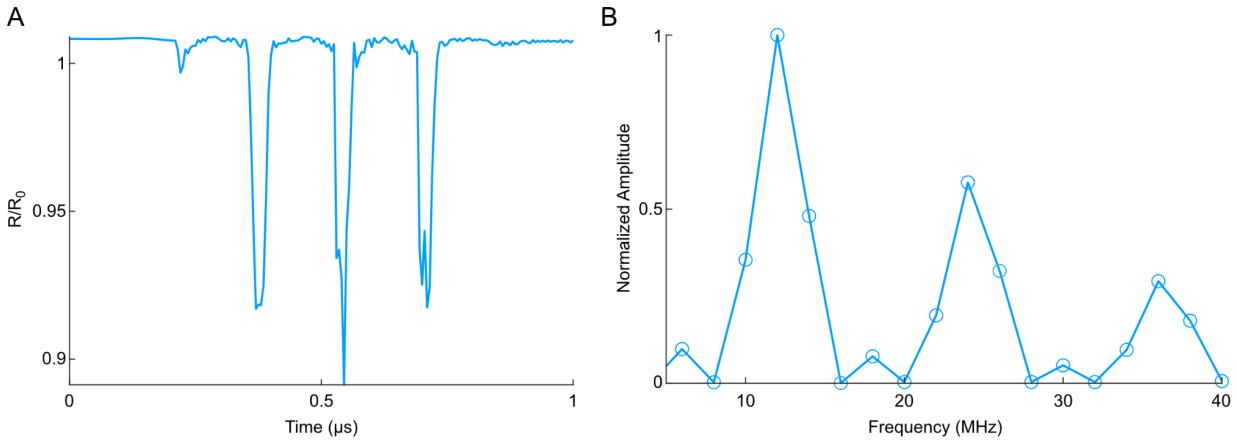


Figure 3-S5. GV volume change. (A) Simulated radius change of AnaS normalized to the original radius (R0) when exposed to 11.4 MHz and 331 kPa ultrasound. (B) The Fourier transform of GV's radial excursion.

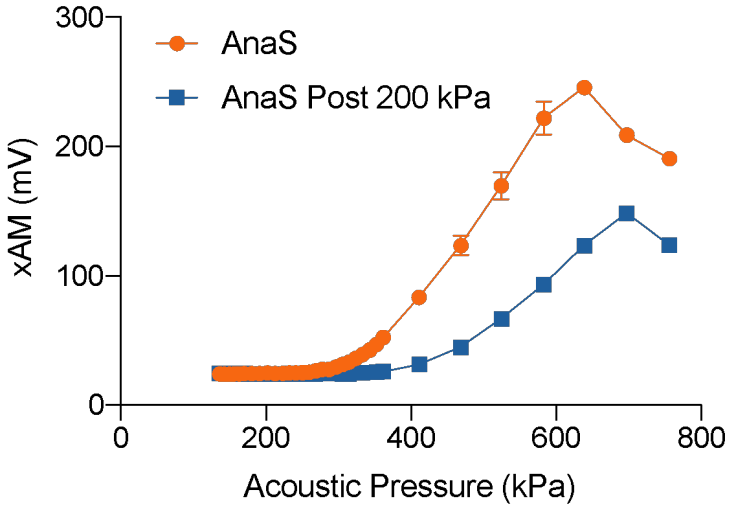


Figure 3-S6. Nonlinear ultrasound signals from AnaS (n = 4) as a function of acoustic pressure from samples with (orange) and without (blue) pre-collapse hydrostatic pressure treatment at 200 kPa. Error bars represent ± SEM, where not seen, are hidden by symbols.

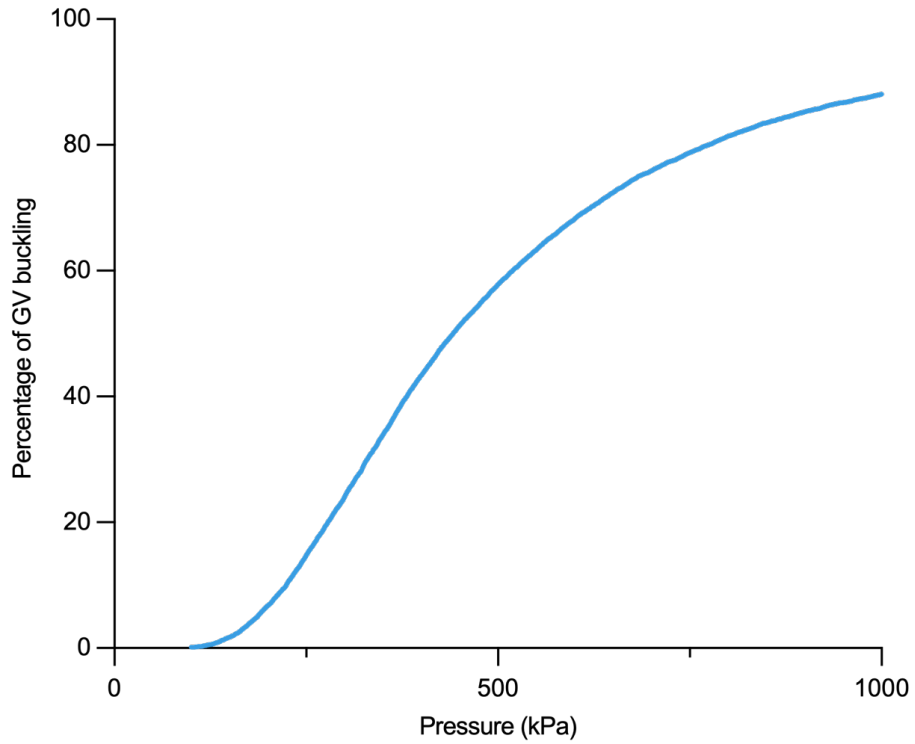


Figure 3-S7. Buckling of an ensemble of 10000 non-interacting GVs with a Gaussian distribution of lengths and diameters, with lengths varying between 500 to 1000 nm and diameters varying between 70 to 100 nm.

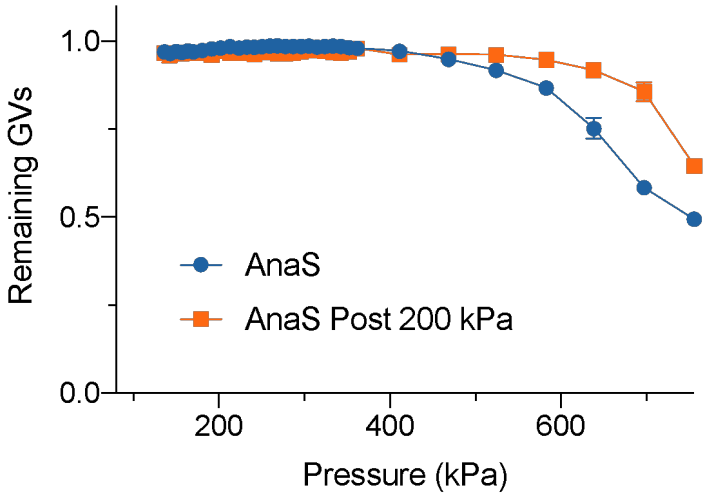


Figure 3-S8. Quantification of remaining GVs by B-mode ultrasound imaging after exposure to different acoustic pressures under xAM imaging. Error bars represent ± SEM, where not seen, are hidden by symbols.

Chapter 4

STRUCTURE OF *ANABAENA FLOS-AQUAE* GAS VESICLES REVEALED BY CRYO-ET

Adapted from:

Przemysław Dutka, Lauren Ann Metskas, Robert C. Hurt, Hossein Salahshoor, Ting-Yu Wang, Dina Malounda, George J. Lu, Tsui-Fen Chou, Mikhail G. Shapiro[#], and Grant J. Jensen[#]. *2023*. "Structure of Anabaena Flos-Aquae Gas Vesicles Revealed by Cryo-ET." *Structure*. https://doi.org/10.1016/j.str.2023.03.011

Abstract

Gas vesicles (GVs) are gas-filled protein nanostructures employed by several species of bacteria and archaea as flotation devices to enable access to optimal light and nutrients. The unique physical properties of GVs have led to their use as genetically-encodable contrast agents for ultrasound and MRI. Currently, however, the structure and assembly mechanism of GVs remain unknown. Here we employ cryo-electron tomography to reveal how the GV shell is formed by a helical filament of highly conserved GvpA subunits. This filament changes polarity at the center of the GV cylinder—a site that may act as an elongation center. Subtomogram averaging reveals a corrugated pattern of the shell arising from polymerization of GvpA into a β-sheet. The accessory protein GvpC forms a helical cage around the GvpA shell, providing structural reinforcement. Together, our results help explain the remarkable mechanical properties of GVs and their ability to adopt different diameters and shapes.

4.1 Introduction

A fundamental property of many living organisms is their ability to move within their environment, with single-celled organisms capable of swimming, swarming, and aligning with magnetic fields. The molecular machines underlying many of these motility functions have been characterized in detail (Komeili et al. 2006; Krause et al. 2018; Wadhwa and Berg 2022). Yet the structure underlying one of the oldest evolved forms of motility-flotation-remains more mysterious. Some cyanobacteria, heterotrophic bacteria, and archaea regulate their buoyancy in aquatic environments to access sunlight and nutrients using intracellular flotation devices called gas vesicles (GVs) (Walsby 1994a; Pfeifer 2012). These unique protein nanostructures consist of a gas-filled compartment, typically ~100 nm in diameter and ~500 nm in length, enclosed by a ~3 nm-thick protein shell (**Figure 4-1A**) that can withstand hundreds of kPa of applied pressure (Lakshmanan et al. 2017; Dutka et al. 2021). The interior of the shell is strongly hydrophobic, keeping out water while allowing gas molecules to diffuse in and out on a sub-millisecond timescale (Walsby 1994a; Pfeifer 2012).

In addition to their biological significance, GVs are a subject of intense interest for biotechnology. Analogously to fluorescent proteins, opsins and CRISPR nucleases, GVs' unusual biophysical properties can be harnessed for other purposes. The gaseous composition of GVs allows them to scatter ultrasound waves, enabling their use as genetically-encoded reporters and actuators of cellular function deep in tissues (Shapiro, Goodwill, et al. 2014; Bourdeau et al. 2018; Farhadi et al. 2019; Wu et al. 2019; Farhadi et al. 2020; Lakshmanan et al. 2020; Bar-Zion et al. 2021). Other applications take advantage of GVs' refractive index, gas permeability and susceptibility to

magnetic fields (Shapiro, Ramirez, et al. 2014; Lu et al. 2018, 2020).

GVs were discovered in the 19th century, but we still have limited knowledge of their structure and assembly. GVs adopt a cylindrical shape, with conical caps (**Figure 4-1A**). Their components are encoded in operons containing relatively few genes (8-23+, depending on the species) (Pfeifer 2012). One of these genes encodes the main structural protein, GvpA, a small (~8 kDa), highly hydrophobic protein that polymerizes to form the GV shell (Walsby 1994a). In some species, the gene cluster contains a secondary structural protein called GvpC, which binds to the exterior of the shell to provide mechanical reinforcement (Hayes, Buchholz, and Walsby 1992). The remaining genes encode proteins whose functions are not well understood, possibly including chaperones, assembly factors, and additional minor shell constituents. GVs are nucleated as bicones which then elongate into a cylindrical shape with low-pitch helical ribs (Pfeifer 2012; Offner et al. 1998), but their detailed molecular structure is not known.

Here, we apply state-of-the-art cryo-electron tomography (cryo-ET) and sub-tomogram averaging techniques to GVs from the cyanobacterium *Anabaena flos-aquae* (Ana). These GVs are among the best studied by biophysicists (Walsby 1994b; Maley et al. 2017; Cai et al. 2020) and the most commonly used in biotechnology applications (Lakshmanan et al. 2016, 2020; Hurt et al., n.d.). We show that the Ana GV shell is formed by a continuous helical filament of repeating GvpA subunits, giving rise to a corrugated cylindrical structure with terminal cones that taper over a conserved distance. Near the middle of the cylinder, the angle of corrugation is inverted, suggesting a potential elongation center for GV biosynthesis. The corrugated shell is externally reinforced by circumferential rods of GvpC. Combining our cryo-ET data with an atomic model

of the homologous *Bacillus megaterium* (Mega) GvpA protein determined in a complementary study (Huber et al. 2022), we build an integrative model of the Ana GV. This model explains the connection between the GV shell and GvpC and highlights the structural conservation of GVs between diverse species. Finally, we extend our study with biochemistry and computational modeling to corroborate our model and explore its implications for GV engineering.

4.2 Results

Molecular architecture of GVs

Ana GVs are long, cone-tipped cylinders with diameters of 85 ± 4 nm (Dutka et al. 2021) and lengths of 519 ± 160 nm (Lakshmanan et al. 2017) (**Figure 4-1A** and **B**). Although GVs have apparent helical symmetry, they are prone to deformations in thin ice (**Figure 4-S1**) and are therefore intractable for cryo-EM helical processing. For this reason, we decided to use cryo-ET. However, cryo-ET analysis of GVs presents its own challenges. We observed that GVs are highly sensitive to electron dose, losing high-resolution features quickly before deflating and shrinking (**Movie S1**). To mitigate this effect, we limited the total electron dose to ~45 electrons/Å^2 per tilt-series, which is ~2.5 times lower than typically used for high-resolution sub-tomogram averaging (Peukes et al. 2020; Metskas et al. 2022).

We started by examining large-scale structural features. While the diameter and length of GVs have been characterized (Walsby and Bleything 1988; Dutka et al. 2021), the conical ends and their connection to the cylindrical body are less studied. Close inspection of individual caps in our cryo-tomograms revealed a heterogenous morphology that deviated from simple conical structure (**Figure 4-1C** and **D**). We observed two elements in the majority of cones: a pointed closed tip,

and a rounded transition region between cone and cylinder (**Figure 4-1D**). The height of the conical caps was 59 ± 6 nm, independent of cylinder diameter (**Figure 4-1E**). The rounding of the base was more pronounced in GVs with larger diameters, so we also examined cryo-tomograms of *Bacillus megaterium* (Mega) GVs, whose average diameter is ~30 nm smaller than that of Ana GVs. However, Mega GVs showed similar rounding at the cap transition (**Figure 4-S2**), suggesting that this is a conserved feature of the structure independent of width.

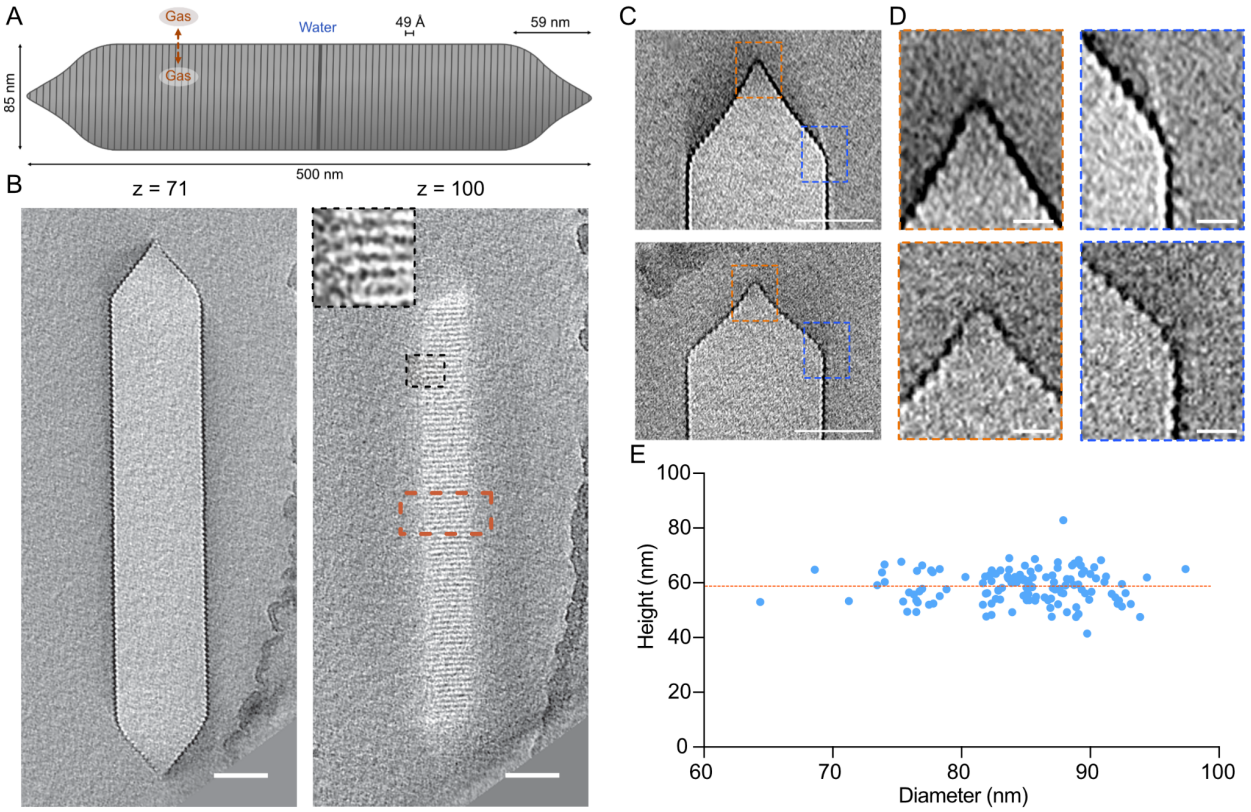


Figure 4-1. Molecular architecture of Ana GVs. (**A**) Schematic representation of an Ana GV with average dimensions annotated. (**B**) Representative slices at the indicated z-heights from a cryo-electron tomogram of an individual GV. Inset shows an enlargement of the area indicated by the black dashed box. Scale bars, 50 nm. (**C**) Central tomographic slices of two conical GV ends with different morphologies. Scale bars, 50 nm. (**D**) Enlarged views of the areas indicated by orange (apex) and blue (cone to cylinder transition) dashed boxes in **C**. Scale bars, 10 nm. (**E**) Distribution of the diameters and heights of conical GV ends, n= 132. Orange dashed line indicate average height of the cones (59 ± 6 nm).

The GvpA spiral reverses polarity in the middle of the cylinder

The GV shell consists of a low-pitch helix, running the length of the GV (**Figure 4-2A** and **B**). Near the middle of the GV, however, the angle of the helix abruptly inverted. Previously, Waaland and Branton (Waaland and Branton 1969) noticed that one rib in the middle of the GV cylinder appears to be thicker than the others and suggested that this could be the growth point, where new GvpA subunits are added. Indeed, this abnormal rib was clearly visible in our tomograms (**Figure 4-2A**). To obtain a better understanding of the rib architecture in that region, we applied subtomogram averaging, which revealed that the angle of corrugation is opposite above and below the central rib (**Figure 4-2B**). This polarity inversion occurs within one rib, and the continuity of the spiral is not broken (**Figure 4-2B** and **C**). We were unable to distinguish whether the polarity of GvpA subunits changed relatively gradually within the space of one helical turn, or abruptly from one monomer to the next. We also could not tell whether additional proteins are present at the inversion point.

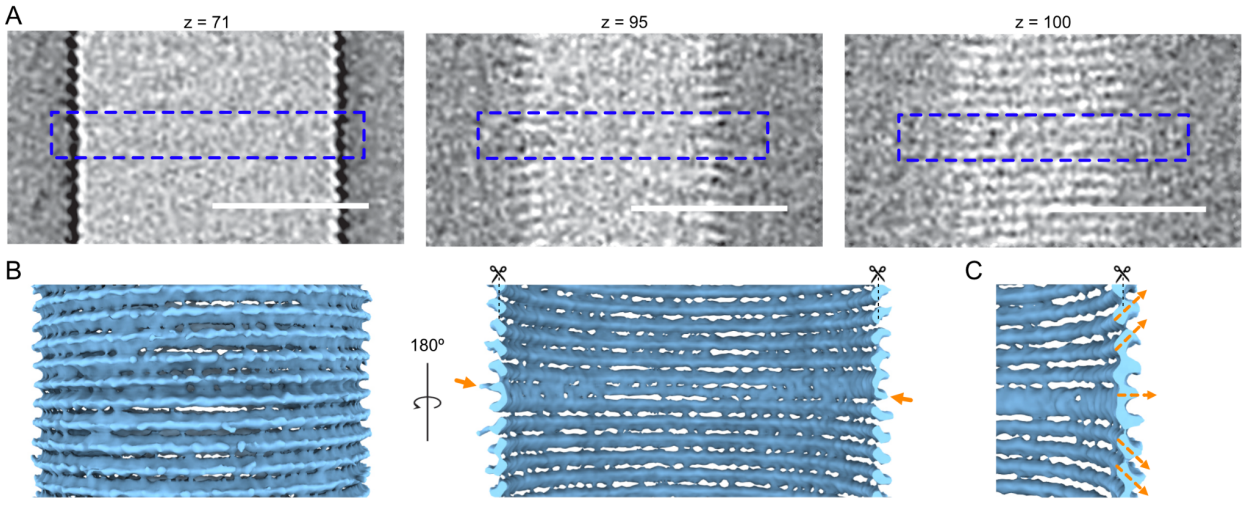


Figure 4-2. Polarity inversion point. (**A**) Enlargement of the tomographic slices from Figure 1B (indicated by the orange dashed box) at different z-heights. Blue dashed outline indicate section where polarity changes. Scale bars, 50 nm. (**B**) Subtomogram average of the middle region of the GV where the ribs reverse polarity. Arrows denote the rib where polarity is reversed. (**C**) Enlarged view of the subtomogram average in (**B**) highlighting the inversion of the helical assembly.

By inspecting hundreds of cryo-electron micrographs of GVs from different species (*A. flos-aquae*, *B. megaterium* and *Halobacterium salinarum*) we found that the polarity inversion point is a conserved feature (**Figure 4-S3**). Although in general the inversion point was near the middle of the cylinder, in some cases it was located closer to one end (**Figure 4-S3A**). If it is the nucleation point, this suggests that GvpA subunits are not always added symmetrically in both directions. Additionally, we observed some examples where a GV exhibited different diameters on either side of the inversion point (**Figure 4-S3B**). While we saw examples in all three species, it was most frequent, and most pronounced, in GVs from *H. salinarum* (Halo).

Sub-tomogram averaging of the GV shell

To understand the molecular details of the GV structure, we applied subtomogram averaging to the Ana GV shell, both in its native state and after biochemically removing the reinforcing protein GvpC to produce "stripped" (AnaS) GVs. Initially, we tried averaging tubular sections of the GVs. However, due to flattening and the low number of particles, the resolution of this approach was limited (**Figure 4-3A**). As an alternative, we decided to average only small sections of the shell with randomly seeded particle centers similar to an oversampling method (Peukes et al. 2020; Wan et al. 2020). This strategy produced a higher number of particles and allowed more rigorous 3D classification to remove distorted particles. With this method, we produced subtomogram averages of native Ana (**Figure 4-3B** and **4-S4**) and AnaS (**Figure 4-S5**) GV shells with global resolutions of 7.7 Å and 7.3 Å, respectively (**Table 4-T1** and **Figure 4-S4** and **4-S5**). Despite high global resolution, our maps manifested a certain degree of anisotropy with significantly lower resolution the in Y-direction (**Figure 4-S4D** and **4-S5D**). The particle poses after subtomogram averaging

indicate that all particles are oriented outward and consistent with a helical arrangement (**Figure 4-S6**). Typically, we observed one significant break in the particle poses per GV, which corresponds to the inversion point. However, due to the strong effects of missing wedge artifacts on tubular structures, such as GVs, they typically appear as two disconnected arches. As a result, we observed a fraction of misaligned particles in the direction of the missing wedge. Furthermore, flattening of the GV cylinder and small variability in diameters could lead to inaccurate alignment of some particles, resulting in blurring of the structure, particularly in the Y-direction, and limiting resolvability of the secondary structures. Although the GV corrugated structure has strong features in the X- and Z-directions, there are no features in the Y-direction that could aid subtomogram alignment. A visual examination of the maps revealed that despite the lower resolution, the map for the native Ana GV shell had higher quality (**Figure 4-S3F-G** and **4-S4C**). For this reason, we used the native GV shell map for further interpretation, and the AnaS map was only used to determine the position of GvpC.

The subtomogram average revealed a prominent pattern of beveled ribs, giving rise to the corrugated GV shell. The shell was ~4 nm wide at its thickest and only ~1 nm thick in the region between adjacent ribs (**Figure 4-3C**). We also observed pores in this region, at the interface between neighboring ribs of the spirals (**Figure 4-3B**), likely allowing gas to diffuse in and out of the GV. In contrast to the complex exterior face of the GV shell, the gas-facing interior appeared relatively smooth.

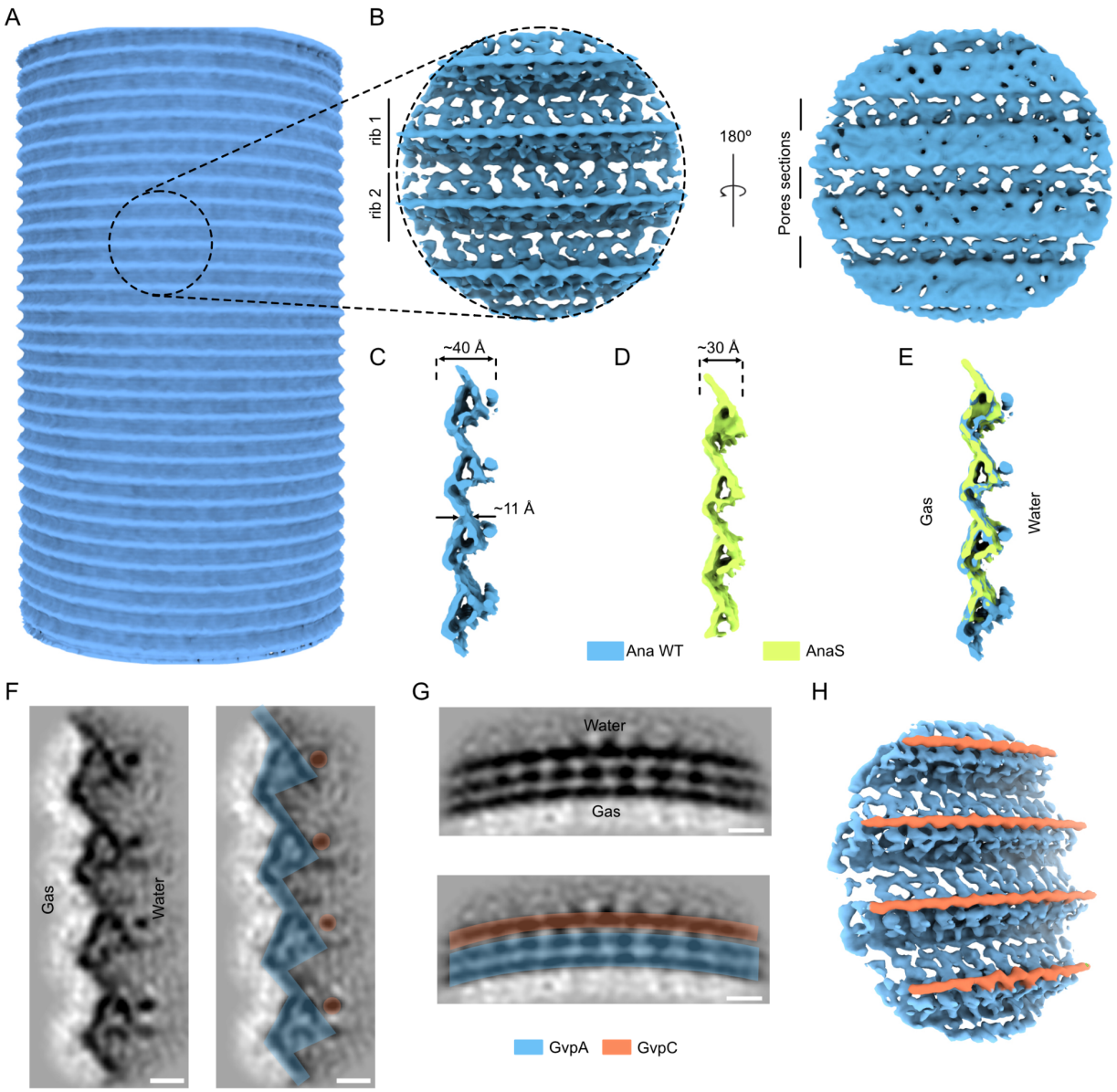


Figure 4-3. CryoET structure of the Ana GV shell. (**A**) Initial, low-resolution subtomogram average of a cylindrical GV segment. (**B**) Orthogonal views of a higher-resolution (7.7 Å) subtomogram average of the native Ana GV shell. (**C-E**) Cross-sections of the subtomogram averages of the GV shell: (**C**) native Ana GV, (**D**) AnaS GV, and (**E**) superimposed. (**F-G**) Projections trough the subtomogram average of the native Ana GV map. (**F**) Projection along the GV helical axis. In the right panel color-coded densities corresponding to GvpA and GvpC. (**G**) Projection trough the neighboring subunits forming GV helix. In the bottom panel color-coded densities corresponding to GvpA and GvpC. Scale bars, 2 nm. (**H**) Segmented density map of the native Ana GV indicating the locations of GvpC.

Comparing the maps of native Ana and AnaS GVs (lacking GvpC), we noticed a pronounced rod-like structure positioned along the GV ribs that is absent in AnaS (**Figure 4-3C-E**). Previously, various models for GvpC binding to the GV shell have been proposed (Buchholz, Hayes, and Walsby 1993), with most of the field favoring one in which GvpC spans longitudinally across GvpA ribs (Maresca, Lakshmanan, et al. 2018; Lakshmanan et al. 2020). Our structure shows instead that GvpC binds circumferentially to the thickest part of the GV shell, thereby creating a spiral cage around the GV cylinder (**Figure 4-3F-H**). We do not yet know whether the GvpC filament binds the central inversion rib or extends to the conical caps, where the decreasing radius of curvature might be prohibitive, or whether it is continuous, as the average would blur away gaps.

Conserved assembly of GvpA and its consequences on GV development and mechanics

The resolution of our Ana GV density map was sufficient for rigid-body fitting of a homology model of GvpA. Taking advantage of the high degree of conservation of the protein, we used the structure of GvpA2 from *B. megaterium* solved by helical reconstruction in a contemporaneous study (Huber et al. 2022). The only substantial difference between GvpA from Ana and Mega is an extended C-terminus in the latter (**Figure 4S7**), so our homology model was complete and fit well into our cryo-ET density map (**Figure 4-4A** and **4-S8**). After docking the model to our map, we observed that the fit of alpha helices is not perfect. It could be either because of limited resolution of our maps or these helices adopt slightly different conformation comparing with *B. megaterium* GvpA2. The GvpA spiral is formed by the polymerization of individual subunits, resembling the packing of amyloids. All domains of the small GvpA protein play a role in building the GV shell (**Figure 4-4B**), packing into a tight structure with only small pores contributing to the remarkable stability of GVs; we find purified GVs are stable for years at cool or ambient

temperature.

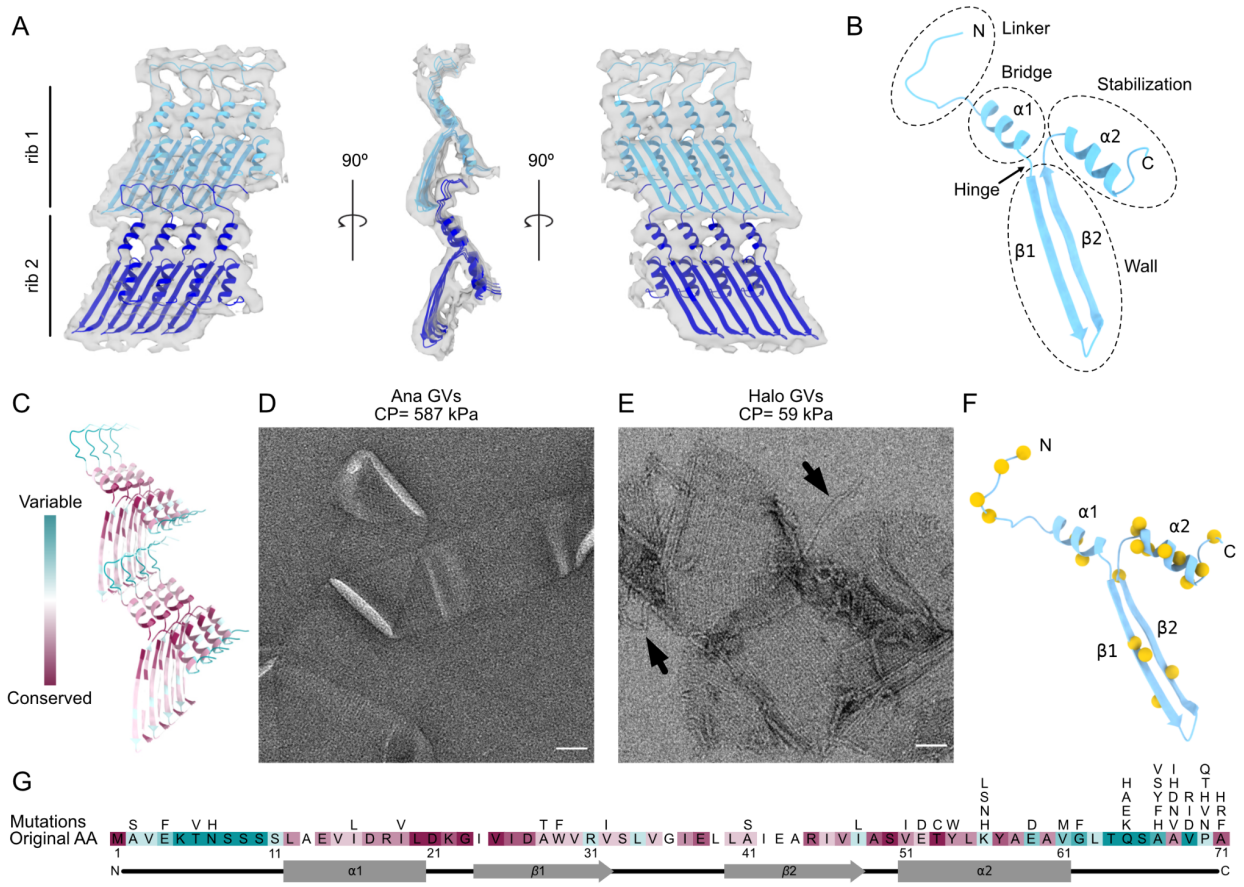


Figure 4-4. Conserved assembly of GV shell. (**A**) Segmented ~8 Å resolution structure of two adjacent GvpA ribs determined by subtomogram averaging (grey surface), fitted with a homology model of GvpA based on GvpA2 [PDB ID: 7R1C(Huber et al. 2022)]. (**B**) Domain annotation within an individual GvpA. (**C**) Conservation analysis of GvpA determined by ConSurf.(Ashkenazy et al. 2016) (**D**, **E**) Negative-stain EM images of collapsed GVs from (**D**) *A.flos-aquae* and (**E**) *H.salinarum*. Arrows indicate separated GvpA filaments. Collapse Pressure (CP) is indicated above. Scale bars, 50 nm. (**F**) Location of tolerated mutation sites (yellow spheres) in the GvpA structure (blue). (**G**) Map of all tolerated mutations in GvpA. Original sequence colored by conservation score as in **C**.

As mentioned above, the only major difference between *B. megaterium* GvpA2 and *A. flos-aquae* GvpA is the presence of an elongated C-terminus (**Figure 4-S7**). This C-terminus was not resolved in a recent structure solved by helical processing (Huber et al. 2022), presumably due to its

flexibility. In our cryo-tomograms of Mega GVs, we observed additional density on the surface of the shell that is absent from the structures of AnaS and native Ana GV shells (**Figure 4-S9**). The density was not highly regular but appeared connected. It may be that this extra density belongs to the C-terminus of GvpA2, which perhaps plays a role in stabilizing the GV shell.

The sequence of GvpA, the major structural protein, is highly conserved in all GV-producing species (Englert, Horne, and Pfeifer 1990; Griffiths, Walsby, and Hayes 1992) and we think it is likely that its structure is similarly conserved, as evidenced by a model from *B. megaterium* GvpA2 (Huber et al. 2022) and fitting into the density of *A. flos-aquae* GvpA. Remarkably, though, GvpA can assemble into GVs with varying diameters (**Figure 4-S10A**) (Dutka et al. 2021) and morphologies (**Figure 4-S10B** and **4-S10C**). For instance, the largest Halo GVs are ~7-times larger in diameter than the smallest Mega GVs. One key to understanding different morphologies may lie in what appears to be a hinge region located between helix α1 and strand β1 (**Figure 4-S4B**), where a conserved glycine resides (**Figure 4-4G** and **4-S7**). Small sequence differences in GvpA have been suggested to contribute to different morphologies of GVs (A. E. Walsby 1994b). *H. salinarum* contains two independent GV gene clusters, p-vac and c-vac (Pfeifer 2012). The sequences of the GvpA encoded by the two clusters are 94% identical (**Figure 4-S7**), yet these cluster can produce GVs with either a lemon shape (**Figure 4-S10B**) or more typical cylindrical shape with conical caps (**Figure 4-S10C**).

We used ConSurf (Ashkenazy et al. 2016) to visualize the evolutionary conservation of GvpA, revealing that the most conserved residues are located in the β-sheets and α-helices (**Figure 4-4C**). In contrast, the N-terminal domains of the protein responsible for interactions between neighboring

ribs showed the greatest variability (**Figure 4-4C**). Within the generally conserved β-strands, the most variable sites were those interacting with the N-terminus from the subunit below. This variability in amino acid composition in the domains responsible for holding adjacent ribs together might be one factor contributing to differences in the mechanical strength of GVs. Under hydrostatic pressure, GVs can collapse, forming flattened sacs (Dutka et al. 2021). The critical pressure required to collapse GVs varies greatly between species. For example, the hydrostatic collapse pressure threshold of Ana GVs is 587 kPa, while that of Halo GVs is 59 kPa, an order of magnitude lower (Lakshmanan et al. 2017). By EM imaging, we found that Ana GVs collapse without major disruptions to the rib structure (**Figure 4-4D**), while collapsed Halo GVs often exhibit major disruption of the rib structure and separation of the GvpA filament (**Figure 4-4E**). This supports the idea that the strength of connectivity between ribs varies between species.

To test the importance of conserved GvpA residues in GV assembly, we mapped tolerated mutations by screening a scanning site-saturation library of GvpA mutants in *Escherichia coli* engineered to express a hybrid gene cluster encoding the structural proteins GvpA and GvpC from the Ana GV gene cluster and the accessory proteins from the Mega GV gene cluster. GV-producing mutant clones were identified by nonlinear x-wave ultrasound (xAM) (**Figures 4-4F,G** and **4-S11**). The results largely correlated with observed evolutionary conservation, with the highest number of function-retaining mutations occurring in the evolutionarily variable C-terminal coil (**Figure 4-4G**). Interestingly, the only conserved region that tolerated mutations well was helix α2, which is not involved in interactions between monomers, but rather plays a crucial role in GvpC binding (see below).

GvpC forms a helical spiral around the GV shell

Having identified GvpC in our subtomogram average of the Ana GV shell (**Figure 3-3H**), we next investigated how GvpC binds to GvpA and how multiple GvpC proteins might cooperate to strengthen GVs. GvpC is predicted to form an amphipathic α-helical structure composed of a characteristic 33-residue repeating sequence (Walsby 1994b; Jumper et al. 2021; Mirdita et al. 2022) (**Figure 4-S12A**). *A. flos-aquae* GvpC consists of 5 such repeats, plus short N- and C-termini. To build a model of GV shell decorated with GvpC, we fitted a poly-alanine helix of a length corresponding to one repeating unit into our subtomogram average (**Figure 4-5A-B**).

We found that GvpC binds perpendicularly to the surface-exposed α2 helices of GvpA, directly above the hydrophobic pockets (**Figure 4-5B-5C**). Although there is insufficient density to anchor the helix, we predict that GvpC binds to GvpA with its hydrophobic side facing the shell. In addition to being amphipathic, GvpC also has an unequal distribution of charge (**Figure 4-12B**). In our model, GvpC binds directly above the negatively-charged C-terminus of GvpA (**Figure 4-S12C**). One 33-residue repeating sequence of GvpC interacts with approximately four GvpAs, indicating a GvpC to GvpA ratio of at least 1:20 if saturated. This is close to the previously calculated ratio of 1:25 (Buchholz, Hayes, and Walsby 1993).

Despite multiple rounds of 3D classification and application of different focus masks, we were unable to resolve the junctions between neighboring GvpC molecules. Instead, GvpC appeared as a continuous helical belt. To get a better understanding of GvpC-GvpC and GvpC-GvpA interactions, we performed chemical cross-linking coupled with mass spectrometry (XLMS) (**Table 4-T2**). Most of the cross-links we observed were between the N-terminus of GvpA and

apparently random locations on GvpC (**Figure 4-5D**), which is consistent with the close association between the N-terminus of GvpA and the GvpA α2 helix in the adjacent rib, where GvpC binds, in our structure (**Figure 4-5A**). However, we did not observe any cross-links between GvpC and helix α2, potentially due to the unfavorable orientation of the lysines. Among GvpC-GvpC cross-links, the most interesting was between K36 and K174 (**Figure 4-5D**). The distance between these residues is ~20 nm, too far for an intramolecular cross-link (Merkley et al. 2014), suggesting that GvpC termini are either closely packed or potentially interact head-to-tail (**Figure 4-S13**).

To quantify the effect of increasing GvpC occupancy on GV stabilization, we used solid mechanics simulations to estimate the applied pressure at which the GV shell starts to buckle–a parameter relevant to its ability to withstand hydrostatic pressure, as well as produce nonlinear signal in ultrasound imaging. We implemented several finite element models of a GV shell, each 500 nm in length and 85 nm in diameter, and with a custom density of GvpC molecules. From a continuous belt, representing 100% GvpC, we randomly removed GvpC-length (25 nm) segments of the helix to achieve the desired saturation for each model (**Figure 4-5E**). We subjected the outer surface of each GV shell to uniform normal stress, simulating hydrostatic or acoustic pressure, and obtained a critical buckling pressure by linear buckling analysis. We observed a simple linear dependence of buckling on scaffolding protein density (**Figure 4-5F**), consistent with previous experimental findings that GvpC level can be utilized to modulate the GV buckling threshold (Lakshmanan et al. 2016).

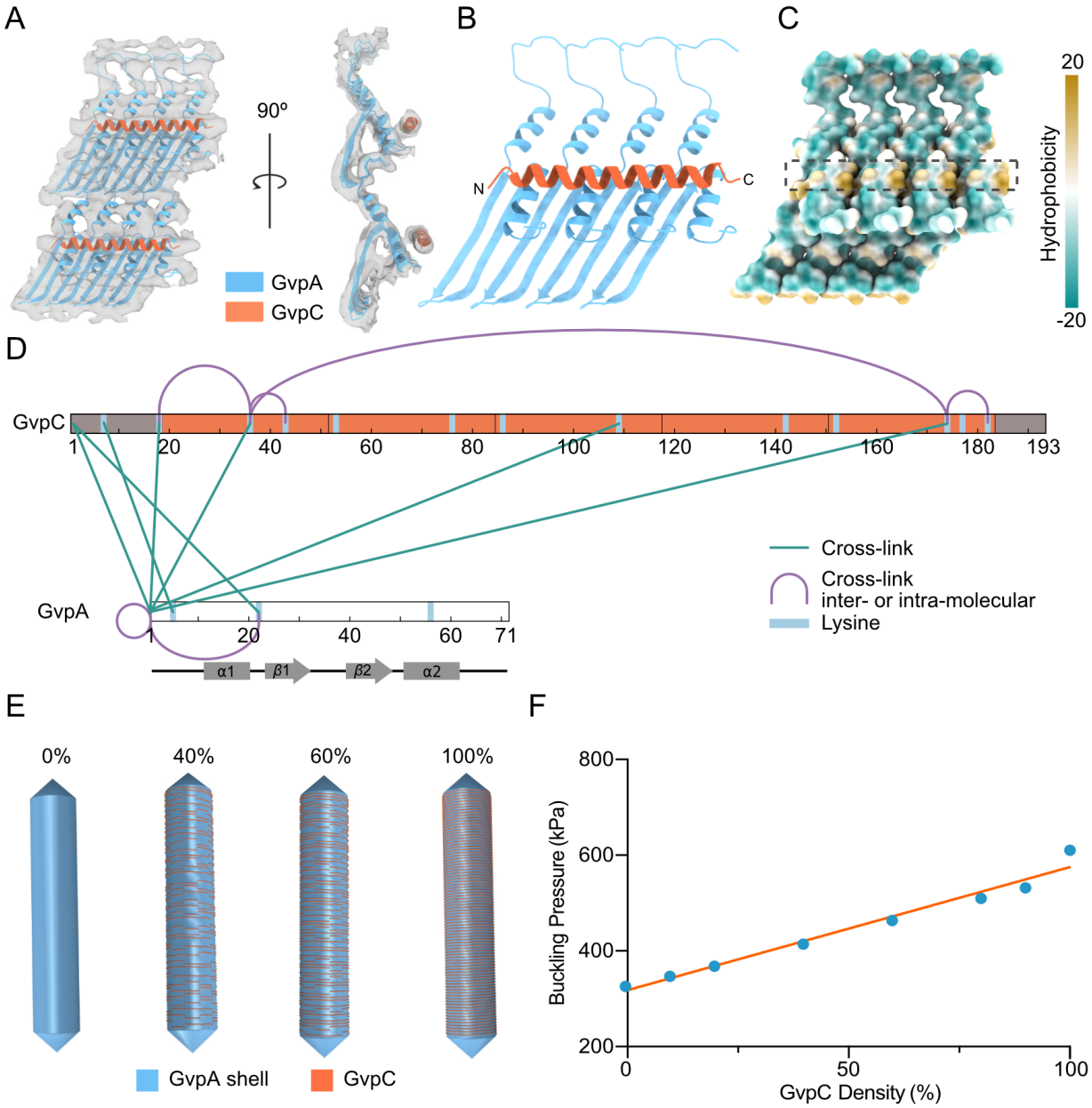


Figure 4-5. Mechanical reinforcement of the GV shell. (**A**) Segmented ~8 Å resolution subtomogram average of neighboring *A.flos-aquae* GvpA monomers connected by GvpC (grey surface) fitted with a model of GvpA and a poly-Ala chain corresponding in length to one repeating sequence of GvpC. (**B**) Resulting GvpC binding model. (**C**) GvpC binding site (dashed black box) at the hydrophobic pockets between α2 helices of GvpA. The surface of GvpA is colored by hydrophobicity. (**D**) Crosslinked sites between GvpA and GvpC identified by mass spectrometry. (**E**) Finite element shell models of a GV with a length of 500 nm and width of 85 nm and the indicated degree of GvpC saturation. (**F**) Buckling pressure as a function of GvpC density. The orange line represents a simple linear regression fit.

4.3 Discussion

The GV shell has remarkable mechanical properties: despite being only ~3 nm thick, it is highly stable and can withstand up to hundreds of kPa of pressure. This is achieved by tight packing of the GvpA subunits into a low-pitch helix that forms a corrugated cylinder. On the macroscopic level, corrugation is typically used when flexibility is important (*e.g.,* pipes) or to increase durability and strength (*e.g.,* unpressurized cans). One or both of these properties might be similarly important for GV function. Our data indicate that GV cylinders can be significantly deformed without collapsing the structure (Dutka et al. 2021). This elasticity of the GV shell may be crucial for adapting to pressure fluctuations *in vivo*, and enables GVs to be used as contrast agents in high-specificity nonlinear ultrasound imaging (Maresca et al. 2017). We noticed a highly-conserved glycine between helix α1 and strand β1 of GvpA. The single hydrogen in the side chain of glycine gives it much more flexibility than other amino acids (Huang and Huang 2018), suggesting that this region may act as a hinge that confers elasticity on the shell structure and lets it adapt to different geometries, such as those observed in terminal cones or the bodies of lemon-shaped GVs.

The primary contact between adjacent GvpA subunits is mediated by lateral interactions of antiparallel β-strands in an extended sheet, resembling the aggregation of β-amyloids (Liberta et al. 2019; Berhanu et al. 2015). Such assemblies are typically stabilized by an extensive network of backbone hydrogen bonding, conferring outstanding strength (Paul et al. 2016). Such strength is also observed in GVs from diverse species; individual GvpA monomers can only be dissociated from the polymer by harsh chemical treatment (Walker and Walsby 1983; Belenky, Meyers, and Herzfeld 2004). That backbone interactions are the main force driving subunit polymerization is

consistent with the wide range of diameters observed in different species (Dutka et al. 2021): as the curvature of the cylinder changes, the relative positions of backbone residues will be affected much less than those of side chains. We find that GvpA domains involved in forming the GV wall have a low tolerance for mutations, likely due to selective pressure to preserve the highly hydrophobic composition of the β-sheets and maintain interactions with the linker domain connecting subsequent coils of the helix. Our scanning mutagenesis data largely correlate with results obtained for Halo GVs (Knitsch et al. 2017). Interestingly, however, Halo GVs appear to be more tolerant to mutations in the conserved regions, possibly because unlike Ana or Mega GVs they are synthesized without turgor pressure in the cells.

Stacked ribs of the continuous GvpA polymer are joined by interactions of the coiled N-termini from one row of subunits with the β-strands of the subunits in the next. We observe that the strength of these inter-rib interactions varies between species, likely related to evolutionary variability in the N-terminal linker. It was previously observed that the critical collapse pressure of Mega GVs is much higher than that of Ana or Halo GVs (Lakshmanan et al. 2017), likely due to the narrower diameter of Mega GVs (Beard et al. 1999, 2000; Salahshoor et al. 2022). However, we note that the C-terminus of Mega GvpA is longer than in other species and in our tomograms of Mega GVs, we observed extended irregular surface densities connecting ribs. We suggest that these extra densities correspond to the extended C-termini of *B. megaterium* GvpA2 and may confer additional mechanical strength.

Other mechanisms also enhance the strength of the GV shell. Almost all GV gene clusters encode an additional, minor structural protein, GvpC, that binds to the GvpA helical spiral and reinforces

the shell (Walsby and Hayes 1988; Lakshmanan et al. 2016); we find that GvpC binds to the surface-exposed α2 helix of GvpA. In our mutational analysis, this helix was relatively mutation-tolerant, suggesting that it has a minimal role in GvpA shell integrity and instead acts primarily as an adapter for GvpC. In contrast to previous models of GvpC spanning ribs, we find that GvpC instead tracks along ribs, forming a spiral cage around the GV cylinder. Our XLMS results indicate close conjunction of GvpC molecules and, even with multiple masking and 3D classification strategies, we never observed discontinuity in the GvpC rod in our subtomogram averages. Although we could not resolve interactions between GvpC N- and C-termini, we previously showed that their removal leads to a significant drop in critical collapse pressure of Ana GVs (Lakshmanan et al. 2016). Here, we used finite element simulations to quantify the reinforcing effect of GvpC density on GV buckling and find that the degree of strengthening is directly proportional to the amount of GvpC bound. However, full GvpC occupancy is not required for full strengthening, and small gaps in the GvpC cage have a negligible effect on collapse pressure. Even though our work focused on Ana GVs, it is possible that the GvpC binding model is conserved between different species of GVs. Previously, the interaction between GvpA and GvpC was studied in *H. salinarum* by split-GFP assay (Jost and Pfeifer 2022) providing similar results to those obtained in our XLMS analysis.

In the initial stage of assembly, GVs grow as bicones until reaching their target diameter; at that point, growth elongates the central section, producing cylinders which can reach several micrometers in length (Pfeifer 2012; Farhadi et al. 2019). The trigger for this transition is unclear. Our data show that the height of mature cones is relatively constant, regardless of GV diameter, indicating that the number of helical turns/height is the measured quantity, rather than the number

of GvpA subunits. Our observation of a polarity inversion near the middle of the GV suggests that this is the site of cylinder elongation, with individual subunits being incorporated in both directions. In some cases, we observed that the elongation center was located closer to one end of the GV, suggesting a mechanism that does not require GvpA subunits to be added symmetrically in both directions. Although GV cylinders typically exhibit a uniform diameter, we documented some examples with different diameters on either side of the elongation center. We observed variations in the shape of conical ends, both within and between GVs. This hints that mismatches in GV geometry might arise in the initial bicone growth stage, but further investigation is needed to fully dissect the mechanism of GV morphogenesis.

Currently, the method of choice for solving the structure of helical assemblies is helical reconstruction (Egelman 2015; He and Scheres 2017). However, the large and nonuniform diameter of Ana GVs (~85 nm) and their susceptibility to deformation during cryopreservation present challenges for this approach. Cryo-ET and subtomogram averaging can circumvent these limitations by focusing on smaller, and therefore more uniform, 3D sections of the object of interest. Subtomogram averaging can reach high resolution in certain favorable cases such as for large (Tegunov et al. 2021) or symmetrical (Schur et al. 2016; Metskas et al. 2022) proteins, but for most targets, resolution has remained limited. Here we show that even with a fairly challenging target, recent developments in cryo-ET data collection and subtomogram averaging methods combined with integrative modeling make it possible to obtain a sufficient resolution to dock an atomic model. Our work, together with a complementary study of Mega GVs (Huber et al. 2022). advances our understanding of the molecular architecture of GVs and may inform further engineering of GVs to serve as genetically-encoded contrast agents and biosensors.

4.4 Limitations of the Study

Using subtomogram averaging we determined the structure of the Ana GV protein shell, providing insight into GV morphogenesis and explaining their unusual mechanical properties. Because of the high conservation of GvpA, we were able to build an integrative model of the Ana GV shell using the homologous structure of *B. megaterium* GvpA2 (Huber et al. 2022). However, the limited resolution of our map allowed us only for a rigid-body fitting. Despite the high homology of GvpA, there might be a subtle difference between the structure of GvpAs from different organisms reflecting unique proprieties of each GV type. Additionally, we are not able to discern if there are any conformational changes caused by GvpC binding. Future higher-resolution studies will be necessary to allow for flexible fitting of GvpA models to extend our knowledge on GV evolution and mechanics. Additionally, to better understand how GVs are assembled will require biochemical and structural work focusing on the GV initiation and elongation process.

4.5 STAR Methods

Resources availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Grant J. Jensen (grant_jensen@byu.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The unprocessed tilt series used for the data analysis are available upon request. Representative tomograms for Ana, Mega, and Halo GVs have been deposited in the Electron Microscopy Data Bank (EMDB) under accession codes EMDB: EMD-29922, EMD- 29925, EMD-29924, EMD-29923. Subtomogram averages for native Ana and AnaS GV shell have been deposited in EMDB under accession codes EMD-29921 and EMD-29916, respectively. The integrative model of Ana GvpA/GvpC has been deposited in the Protein Data Bank (PDB): PDB 8GBS. The XLMS data have been deposited to the ProteomeXchange Consortium with the dataset identifier PXD038631. The code for ultrasound data collection and processing is available upon request.

Experimental model and subject details

GVs were produced either in native sources, *Anabaena flos-aquae* (Ana) and *Halobacterium salinarum* NRC1 (Halo), or expressed heterologously in Rosetta 2(DE3)pLysS *Escherichia coli*, *Bacillus megaterium* (Mega). We followed previously published protocols by Lakshmanan et al. (Lakshmanan et al. 2017) describing in details bacterial growth conditions specific for production of each GV type investigated here.

Methods details

GV preparation

GVs were isolated as previously described (Lakshmanan et al. 2017). In the final steps of buoyancy purification, the sample buffer was exchanged for 10 mM HEPES, pH 7.5. To obtain GVs stripped of GvpC (AnaS), 6 M urea solution was added to purified native GVs and two additional rounds of buoyancy purification were performed. AnaS GVs were subsequently dialyzed in 10 mM HEPES, pH 7.5. Concentrations were measured by optical density (OD) at 500 nm using a spectrophotometer (NanoDrop ND-1000, Thermo Scientific).

Cryo-ET

A freshly purified GV sample was diluted to OD_{500} = ~20 (Ana and Halo), ~3 (AnaS), or ~1 (Mega) and mixed with 10 nm BSA-coated gold beads. A 3 µL volume of sample was applied to C-Flat 2/2 - 3C grids (Protochips) that were freshly glow-discharged (Pelco EasiGlow, 10 mA, 1 min). GV samples were frozen using a Mark IV Vitrobot (FEI, now Thermo Fisher Scientific) (4°C, 100% humidity, blot force 3, blot time 4 s).

Tilt-series were collected on a 300 kV Titan Krios microscope (Thermo Fisher Scientific) equipped with a K3 6k × 4k direct electron detector (Gatan). Multi-frame images were collected using SerialEM 3.39 software (Mastronarde 2005) using a dose-symmetric tilt scheme. Super-resolution movies were acquired at a pixel size of 0.8435 Å (53,000× magnification) with varying defocus from - 1.0 to - 3.5 µm. Tilt-series of Halo and Mega GVs were collected from -60° to 60° with 3° increments. Tilt-series of native Ana GVs were collected in two sessions. The first set was collected from -60° to 60° with 3° increments and the second from -44° to 44° with 4° increments. For AnaS GVs, data were collected from -45° to 45° with 3° increments. Due to the rapid shrinking

of GVs during exposure to the electron beam (**Movie S4.1**), the total accumulated dose in all cases was limited to 45 electrons/Å^2. Data collection parameters are summarized in **Table S4.1**.

Raw movies were binned by a factor of 2 and gain- and motion-corrected on-the-fly using Warp (Tegunov and Cramer 2019). Assembled tilt-series were exported to Dynamo (Castaño-Díez et al. 2012) for automated alignment using *autoalign_dynamo* (Burt et al. 2021). Aligned tilt-series were CTF corrected and full tomograms were either reconstructed in Warp with a pixel size of 10 Å or manually aligned and reconstructed using Etomo (Mastronarde and Held 2017).

Subtomogram averaging – inversion point

Sub-volume extraction, alignment, and averaging were performed using the Dynamo software package (Castaño-Díez et al. 2012). Particles for subtomogram averaging of the inversion site were manually selected from GVs with a diameter of ~85 nm, yielding a total of 68 particles. Sub-volumes were extracted from 4x binned tomograms with a final pixel size of 6.748 Å and 180x180x180 box size. The initial reference for particle alignment was generated by averaging segments with azimuth-randomized orientations. Due to the low number of particles, subtomogram averaging was not performed according to a gold standard. Instead, convergence of the structure was analyzed by changes in particle shifts and cross-correlation scores. During the final rounds of refinement, a soft cylindrical mask was applied to the central 40% of the GV tube.

Subtomogram averaging – GV shell

Subtomogram averaging was carried out using Dynamo,(Castaño-Díez et al. 2012) Warp (Tegunov and Cramer 2019), Relion-3.1 (Zivanov et al. 2018), and M (Tegunov et al. 2021)

software packages. Data transfer between Dynamo and Warp/M was carried out using a set of tools provided by *warp2catalogue* and *dynamo2m* (Burt et al. 2021). Particle selection and initial reference generation were performed using the Dynamo package. Orientations and positions of shell sections were determined using geometrical tools for particle picking in Dynamo (Castaño-Díez, Kudryashev, and Stahlberg 2017). Initial estimates of positions and orientations on the GV shell were generated with an interparticle distance of ~150 Å (~3 ribs). Particles were extracted in Dynamo with a pixel size of 10 Å and averaged. After removal of duplicated particles, data was transferred to Warp and subtomograms were reconstructed with a pixel size of 5 Å based on the alignment information from Dynamo. Subtomograms were subsequently refined in RELION, re-reconstructed at 2.5 Å/pixel and 3D classified without alignment in RELION. After 3D classification, several additional rounds of 3D refinement were carried out in RELION. Finally, subtomograms were reconstructed at 1.687 Å/pixel and iteratively refined in RELION and M using a soft-edged mask around ~3 or 4 adjacent ribs. Although we did not see a resolution boost after iterative refinement of the tilt-series parameters in M, subsequent refinement in RELION produced a better-quality reconstruction when applied to particles reconstructed after M refinement. Final maps were post-processed in RELION. The resolution was estimated using a soft-edged mask around ~3-4 adjacent ribs in 3DFSC program (Tan et al. 2017). The final results are summarized in **Figures 4-S4, 4-S5,** and **Table 4-T1**.

Model building and validation

Although the density map for AnaS reached a higher overall resolution, individual features were better resolved in the map of native Ana GVs (**Figure 4-S4**), so all model building was performed using this map. To build the GvpA model, a high-resolution cryo-EM structure of the homologous

GvpA2 from *B. megaterium* (PDB: 7R1C) (Huber et al. 2022) was fitted into the segmented cryo-ET density map corresponding to an individual subunit in UCSF Chimera (Pettersen et al. 2004). The GvpA amino acid sequence was rebuilt by manual replacement of mismatched residues in Coot (Emsley et al. 2010). The *A. flos-aquae* GvpA model was subsequently refined by rigid-body fitting using the Phenix real-space refinement tool (Adams et al. 2010). The refined GvpA model was used to populate a larger section of the cryo-ET map in UCSF Chimera (Pettersen et al. 2004). The multimeric GvpA model was further refined by rigid-body fitting in Phenix to maximize fit into the density map. The GvpC model was built as a poly-Ala chain in Coot. The poly-Ala chain corresponds in length to a single 33-residue repeating sequence of GvpC and spans across four subunits of GvpA.

The quality of the fit was analyzed by visual inspection and fitting scores from UCSF Chimera (**Figure 4-S8**). We roughly placed four GvpA subunits at the height of one rib and performed a global search using "fitmap" command in Chimera. Subsequently, we analyzed scores for cross-correlation and fraction inside density for each fit. The three best results with similar fitting scores all fit our density map very well and are only different in that they shift by one subunit along Y (the are essentially all the same "fit"). We obtained similar results with a starting point at the height of other ribs.

Negative stain EM

To prepare collapsed GV samples, the purified GV sample was diluted to OD_{500}~ 0.5 and pressurized in a sealed syringe until the solution turned transparent. Three microliters of the target sample was applied to a freshly glow-discharged (Pelco EasiGlow,15 mA, 1 min) Formvar/carbon-

coated, 200 mesh copper grid (Ted Pella) for 1 min before blotting. Afterward, the sample was incubated for 1 min with a 0.75% uranyl for-mate solution before blotting and air-dried. Image acquisition was performed using a Tecnai T12 (FEI, Thermo Fisher Scientific) EM at 120 kV, equipped with a Gatan Ultrascan 2 k×2 k CCD.

Cross-linking mass spectrometry (XLMS)

The cross-linking procedure was carried out according to the manufacturer's instructions (Thermo Fisher). In brief, a freshly purified sample of native Ana GVs in 10 mM HEPES, pH 7.5 was mixed with an excess of cross-linker: either DSSO or BS3 (Thermo Fisher). The sample was incubated for 1h at room temperature and subsequently the reaction was quenched with Tris buffer at a final concentration of 20 mM.

The crosslinking samples were digested in an S-Trap mcrio spin column (Protifi, USA) according to the manufacturer's instructions. For trypsin digestion, an additional aliquot of trypsin was added after 24 hours on the S-trap column and the digestion continued for another 24 hours. After elution and drying, peptides were suspended in LCMS-grade water containing 0.2% formic acid and 2% acetonitrile for further LC-MS/MS analysis. LC-MS/MS analysis was performed with an EASY-nLC 1200 (Thermo Fisher) coupled to a Q Exactive HF hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher). Peptides were separated on an Aurora UHPLC Column (25 cm × 75 μm, 1.6 μm C18, AUR2-25075C18A, Ion Opticks) with a flow rate of 0.35 μL/min for a total duration of 43 min and ionized at 1.7 kV in the positive ion mode. The gradient was composed of 6% solvent B (2 min), 6-25% B (20.5 min), 25-40% B (7.5 min), and 40–98% B (13 min); solvent A: 2% ACN and 0.2% formic acid in water; solvent B: 80% ACN and 0.2% formic acid. MS1 scans were acquired at a resolution of 60,000 from 375 to 1500 m/z, AGC target 3e6, and a maximum injection time of 15 ms. The 12 most abundant ions in MS2 scans were acquired at a

resolution of 30,000, AGC target 1e5, maximum injection time 60 ms, and normalized collision energy of 28. Dynamic exclusion was set to 30 s and ions with charges +1, +7, +8, and >+8 were excluded. The temperature of the ion transfer tube was 275°C and the S-lens RF level was set to 60. For cross-link identification, MS2 fragmentation spectra were searched and analyzed using Sequest and XlinkX node bundled into Proteome Discoverer (version 2.5, Thermo Scientific) against *in silico* tryptic digested *Dolichospermum-flos-aquae* GvpA from the Uniprot database. The maximum missed cleavages were set to 2. The maximum parental mass error was set to 10 ppm, and the MS2 mass tolerance was set to 0.05 Da. Variable crosslink modifications were set DSS (K and protein N-terminus, +138.068 Da) for BS3 crosslink and DSSO (K and protein N-terminus, +158.004 Da) for DSSO crosslink, respectively. For BS3 crosslink, the dynamic modifications were set to DSS hydrolyzed on lysine (K, +156.079 Da), oxidation on methionine (M, +15.995 Da), protein N-terminal Met-loss (-131.040 Da), and protein N-terminal acetylation (+42.011 Da). For the DSSO crosslink, the dynamic modifications were set to DSSO hydrolyzed on lysine (K, +176.014 Da), DSSO Tris on lysine (K, +279.078 Da), oxidation on methionine (M, +15.995 Da), protein N-terminal Met-loss (-131.040 Da) and protein N-terminal acetylation (+42.011 Da). Carbamidomethylation on cysteine (C, +57.021 Da) was set as a fixed modification. The false discovery rate (FDR) for crosslinked peptide validation was set to 0.01 using the XlinkX/PD Validator Node and crosslinks with an Xlinkx score greater than 30 were reported here. The raw data have been deposited to the ProteomeXchange Consortium (Deutsch, Bandeira, and Sharma, n.d.) via the PRIDE (Perez-Riverol et al. 2019) partner repository.

Scanning site saturation library generation and screening

The scanning site saturation library was constructed via a Gibson assembly-based version of cassette mutagenesis as previously described (Ravikumar et al. 2018). Briefly, the *A. flos-aquae*

GvpA coding sequence was divided into sections that tiled the gene, and oligos were designed to have a variable middle region with flanking constant regions against which PCR primers with Gibson overhangs were designed. The variable region was designed to sequentially saturate each residue with every amino acid other than the WT at that position, plus a stop codon to produce truncation mutants (*i.e.*, the size of such libraries is 20 * [# of amino acids in the protein]). Oligos were synthesized as a pool by Twist Biosciences, and were amplified by 10 cycles of PCR (both to make them double-stranded and to add overhangs for Gibson assembly) using Q5 polymerase (according to the manufacturer's protocol, but with 5 μM of each primer) and assembled with the rest of the GV gene cluster (*i.e.,* Ana GvpC and Mega GvpR-GvpU) into a pET28a vector via Gibson assembly using reagents from New England Biolabs. Assembled libraries were electroporated into NEB Stable *E. coli* and grown in Lennox LB with 100 ug/mL kanamycin and 1% glucose (Ammar, Wang, and Rao 2018). Plasmid DNA was miniprepped (Econospin 96-well filter plate, Epoch Life Science) and verified by Sanger sequencing. Ultrasound-based phenotyping of mutants was performed in BL21-AI (Thermo Fisher) as previously described (Hurt et al., n.d.), and all screened mutants were sequenced using the evSeq pipeline (Wittmann et al. 2022).

Finite element simulation

We first developed a finite element model of a single stripped GV isolated from *A. flos-aquae* (AnaS). The geometry, adapted from the cryo-EM images, comprises a cylindrical shell with conical ends, with height and diameter, respectively, of 500 nm and 85 nm. The protein wall was idealized as a continuum shell with a thickness of 2.8 nm and a shell density of 1350 kg/m^3. The rib-like structure of the gas vesicle wall was mirrored in the computational model by an elastic anisotropic material model, with elastic moduli across and along the principal axis of the GV of 0.98 GPa and 3.92 GPa, respectively (Maresca et al. 2017). In order to simulate the nearly

incompressible nature of the protein shell, we assigned a Poisson's ratio of 0.499. We note that the material parameters were not obtained from direct experimental measurements, but rather chosen such that, in addition to falling within a range of parameters consistent with those of protein-based biological materials (Gosline et al. 2002), they effectively replicated the buckling pressures observed experimentally.

We next added a helical rod that spirals around the cylindrical portion of the GV shell, modeling the GvpC molecules. We modeled the GvpC rod as a shell of radius 0.6 nm. The helical structure was generated by assigning a pitch of 4.9 nm. The finite element model of the resultant wild-type GV was obtained by discretizing the entire geometry with quadrilateral shell elements of effective side length 1 nm with reduced integration (*i.e.*, S4R elements) in Abaqus (Dassault Systemes Simulia, France). These general-purpose shell elements with only one integration point within each element are capable of capturing both tensile and in-plane bending, and, with a sufficiently fine mesh size, are computationally cost-effective. We subjected the interior surfaces of the GV to an initial pressure of 101 kPa, modeling the inner gas pressure. We further subjected the vertices at both the top and bottom conical ends of the GV to a zero-displacement Dirichlet boundary condition, which prevented rigid body translations and rotations of the entire GV structure.

In order to investigate the effect of GvpC density on the buckling pressure, we first computed the total length of the helix where N, D, and z are the total number of turns, the perimeter of the GV cross-section, and the pitch of the helix, respectively. Given the pitch and the length of the cylindrical segment of the GV model, 416.5 nm, the total number of turns was computed as 85. We thus computed the total length of the helix as 22.702 micrometers. Given that the length of

GvpC is ~25 nm, about 908 GvpC molecules constituted the helix in our model. We generated six additional finite element models with distinct GvpC saturation levels of 90%, 80%, 60%, 40%, 20%, and 10%, for which we randomly removed about 90, 180, 360, 540, 720, and 810 GvpC units, respectively.

We conducted linear buckling analysis (LBA) and solved the corresponding eigenvalue problem to obtain the threshold buckling pressures for each model. We solved this problem using the Lanczos algorithm and obtained the first ten modes of buckling. Unlike the buckling modes (*i.e.*, eigenvectors), which were virtually identical at different levels of GvpC saturation, the buckling pressures (*i.e.*, eigenvalues) were remarkably dependent on the GvpC density, with an almost linear monotonic relation, where decreasing the saturation level decreases the buckling pressure. **Figure 4-S14** depicts the buckling modes and pressures for 100%, 60%, 20%, and 0% GvpC saturations.

Bioinformatics and visualization

Protein sequence alignment was carried out using Clustal Omega (Sievers et al. 2011) and visualized with Jalview (Waterhouse et al. 2009). Protein conservation analysis was performed using ConSurf (Ashkenazy et al. 2016). Data were visualized using GraphPad Prism, IMOD (Kremer, Mastronarde, and McIntosh 1996), Chimera (Pettersen et al. 2004), and ChimeraX (Goddard et al. 2018), Identified crosslinks were visualized using xiNET (Combe, Fischer, and Rappsilber 2015).

Quantification and statistical analysis

The heights of the GVs' conical ends (**Figure 4-1E)** were manually measured from cryo-electron tomograms using ImageJ. The average height is calculated from 132 conical ends and reported as mean±SD.

The subtomogram averages were determined using software listed in the Key Resources Table. Details of the data processing are displayed in **Figure 4-S4** and **4-S5**, and **Table 4-T1**. The resolution anisotropy and final FSC curves (**Figure 4-S4D** and **4-S5D**) were determined using the 3DFSC package.

The ultrasound data (**Figure 4-4F** and **4-S11**), XLMS analysis (**Figure 4-5D** and **Table 4-T2**), and finite element simulation (**Figure 4-5E-F** and **4-S14**) were analyzed using software listed the Key Resources Table.

No other statistical analyses were performed.

4.6 Acknowledgments

The authors are grateful to Catherine Oikonomou for helpful editorial comments. We thank Songye Chen for assistance with tomography data collection. Electron microscopy was performed in the Beckman Institute Resource Center for Transmission Electron Microscopy at Caltech. The Proteome Exploration Laboratory (PEL) is supported by the Beckman Institute and NIH 1S10OD02001301. This work was supported by the National Institutes of Health (grant R01-AI127401 to G.J.J. and R01-EB018975 to M.G.S.) and the Caltech Center for Environmental

Microbial Interactions (CEMI). Related research in the Shapiro Laboratory is supported by the Packard Foundation, the Chan Zuckerberg Initiative and the Heritage Medical Research Institute.

4.7 Supplementary Data

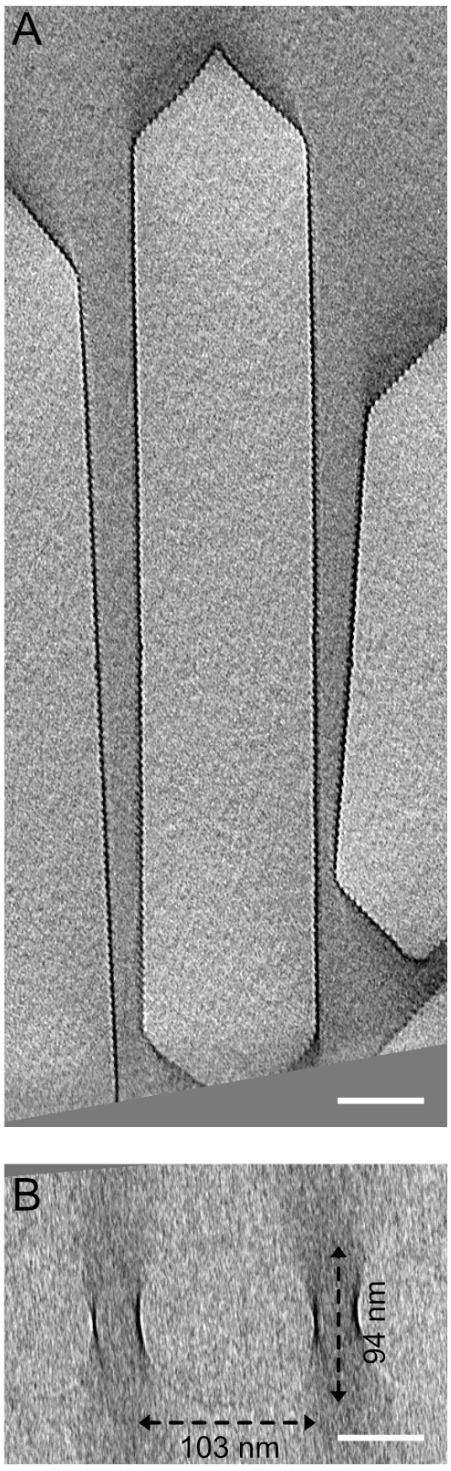


Figure 4-S1. GV flattening in the thin ice. (**A**) XY and (**B**) XZ tomographic slices of the deformed Ana GV. Scale bars, 50 nm.

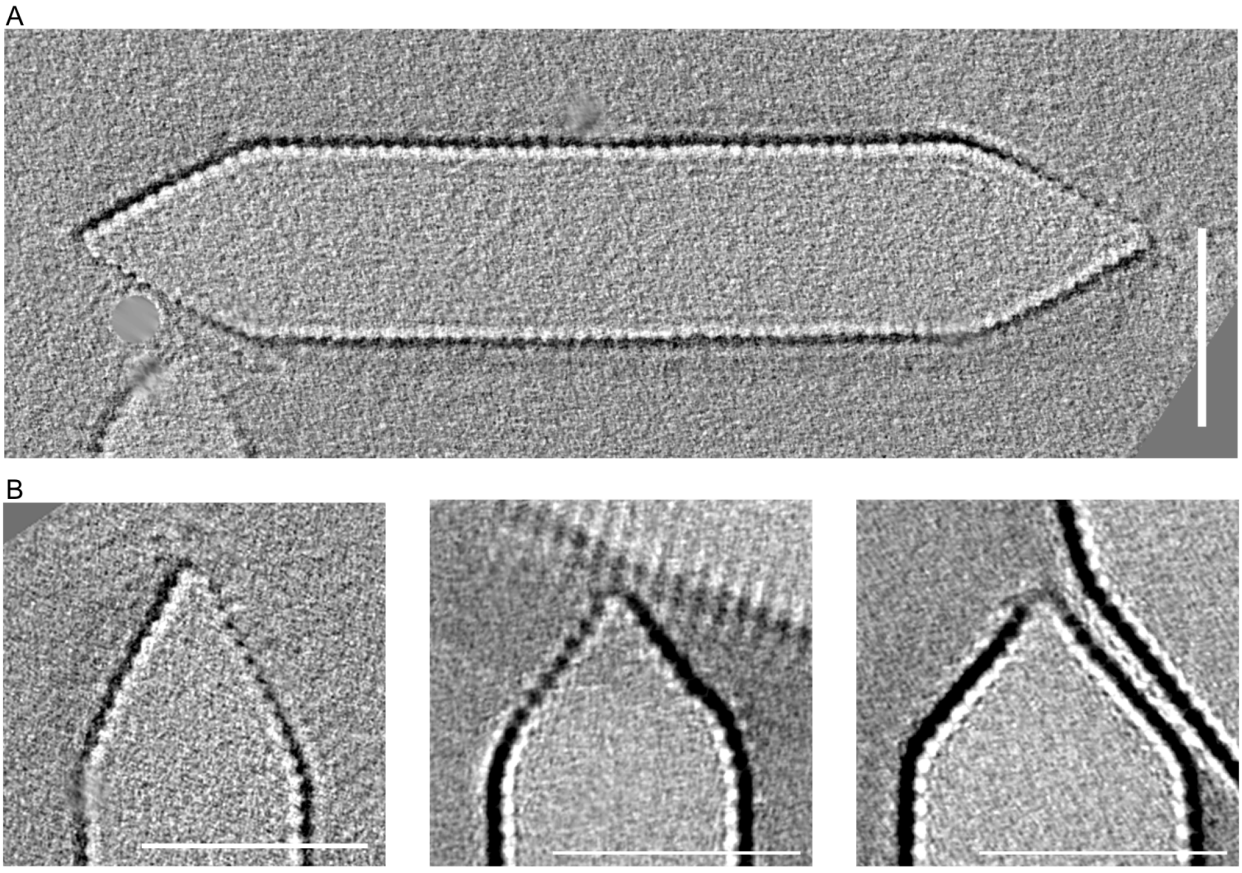


Figure 4-S2. The architecture of Mega GVs. (**A**) Representative central slice from cryo-electron tomogram of individual Mega GV. (**B**) Central tomographic slices of the Mega GV conical ends with slightly different shapes. Scale bars, 50 nm.

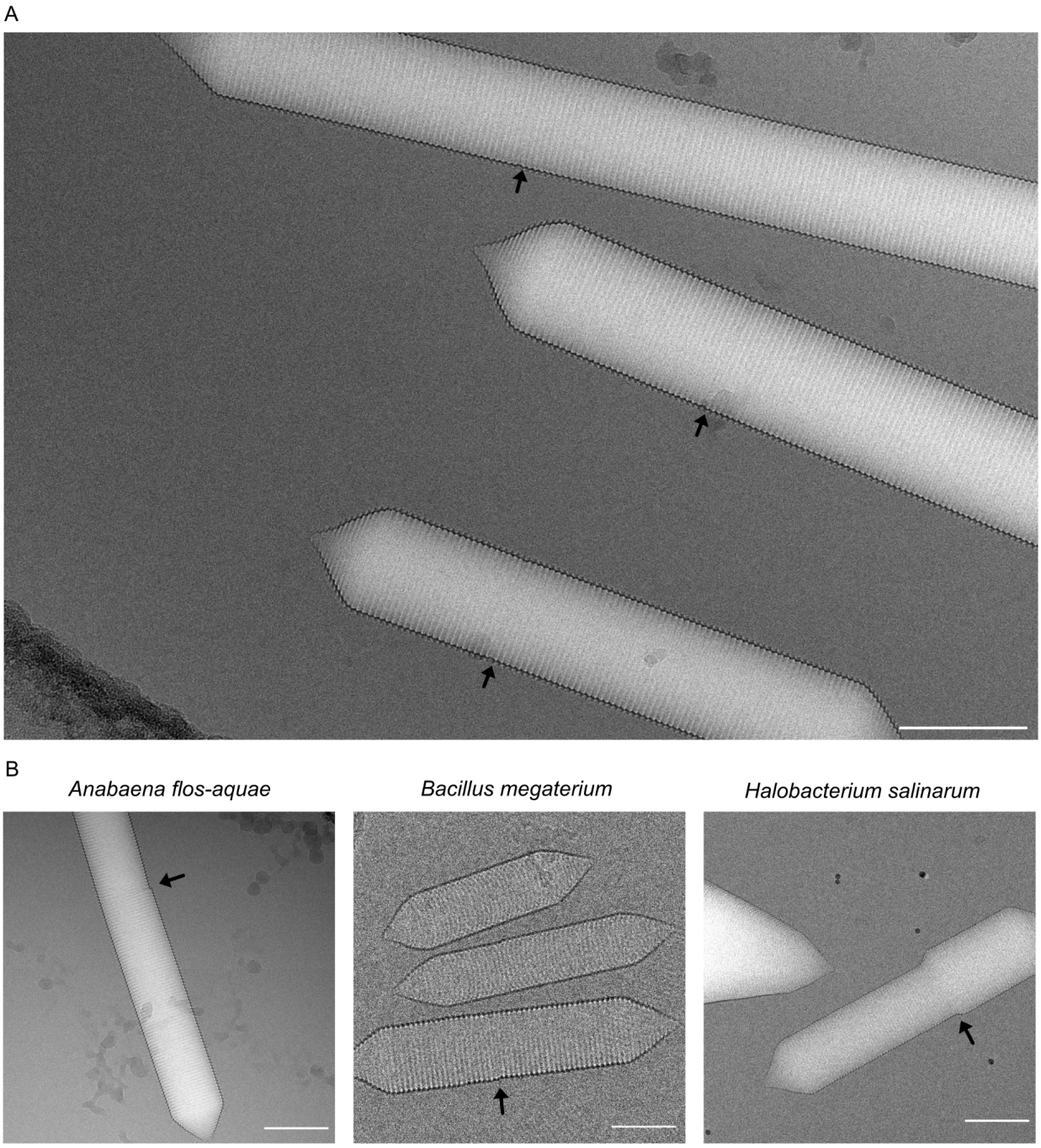


Figure 4-S3. GV Polarity inversion point. (**A**) Location of the polarity inversion point. (**B**) GVs can have different diameters on either side of the inversion point. The black arrows indicate the location of the inversion point.

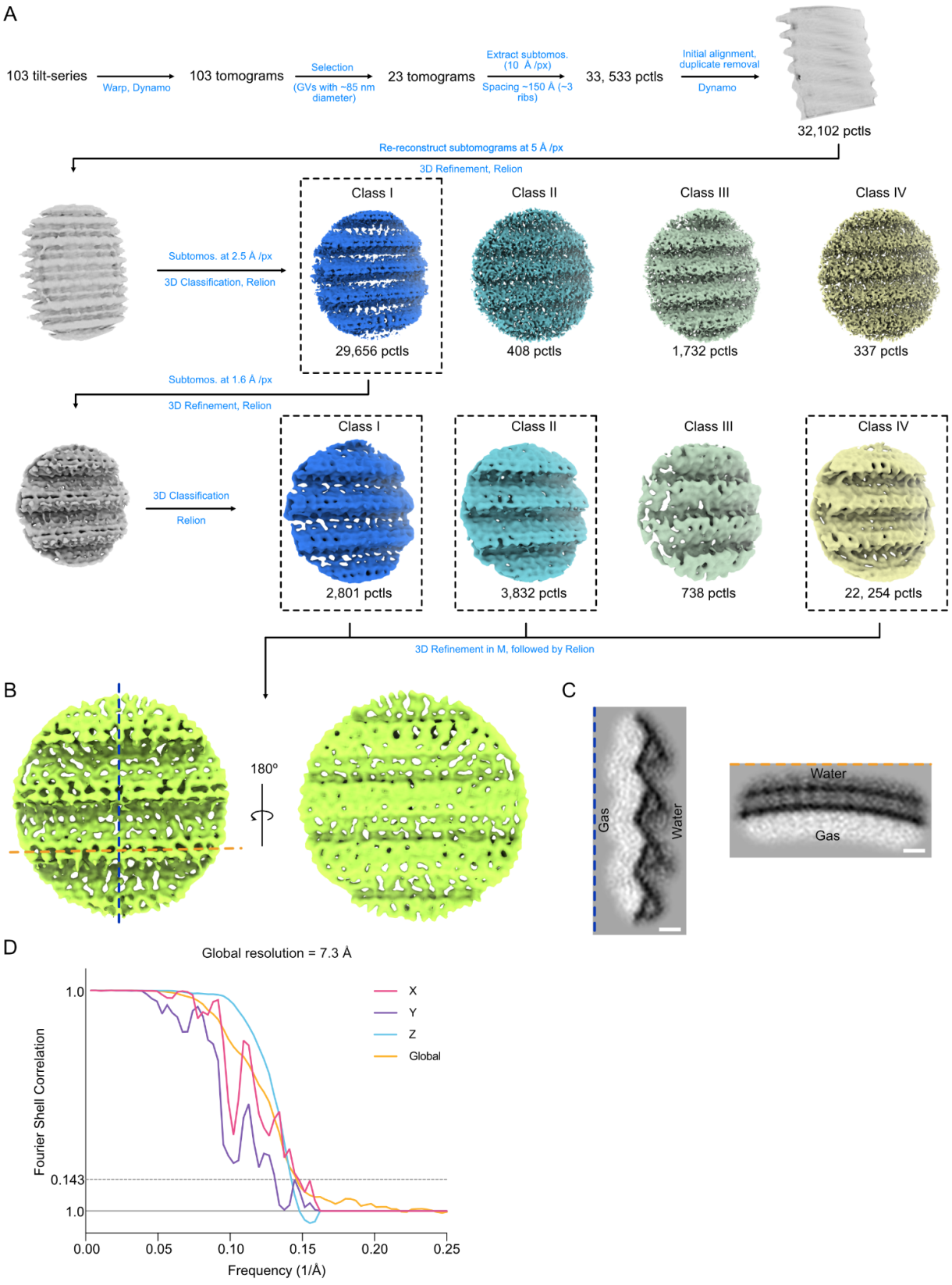

A
103 tilt-series
Warp, Dynamo
103 tomograms
Selection
(GVs with ~85 nm diameter)
23 tomograms
Extract subtomos. (10 Å /px)
Spacing ~150 Å (~3 ribs)
33, 533 pctls
Initial alignment, duplicate removal
Dynamo
32,102 pctls
Re-reconstruct subtomograms at 5 Å /px
3D Refinement, Relion
Subtomos. at 2.5 Å /px
3D Classification, Relion
Class I
29,656 pctls
Class II
408 pctls
Class III
1,732 pctls
Class IV
337 pctls
Subtomos. at 1.6 Å /px
3D Refinement, Relion
3D Classification
Relion
Class I
2,801 pctls
Class II
3,832 pctls
Class III
738 pctls
Class IV
22, 254 pctls
3D Refinement in M, followed by Relion
B
180°
C
Gas
Water
Water
Gas
D
Global resolution = 7.3 Å
Fourier Shell Correlation
1.0
0.143
1.0
0.00
0.05
0.10
0.15
0.20
0.25
Frequency (1/Å)
X
Y
Z
Global

Figure 4-S4. CryoET data processing for the native Ana GV shell. (**A**) Simplified schematic of the subtomogram averaging workflow highlighting crucial steps in the pipeline. (**B**) Orthogonal views of the final average. (**C**) Gray-scale orthographic slices of the native Ana GV shell in the positions indicated by the green and yellow dashed lines in **B**. Orange and blue arrows point to GvpC rod and GvpA subunits, respectively. (**D**) Global FSC curve for AnaS GV shell (yellow), and map anisotropy analysis by FSC curves in X (pink), Y (purple) and Z (blue) directions.

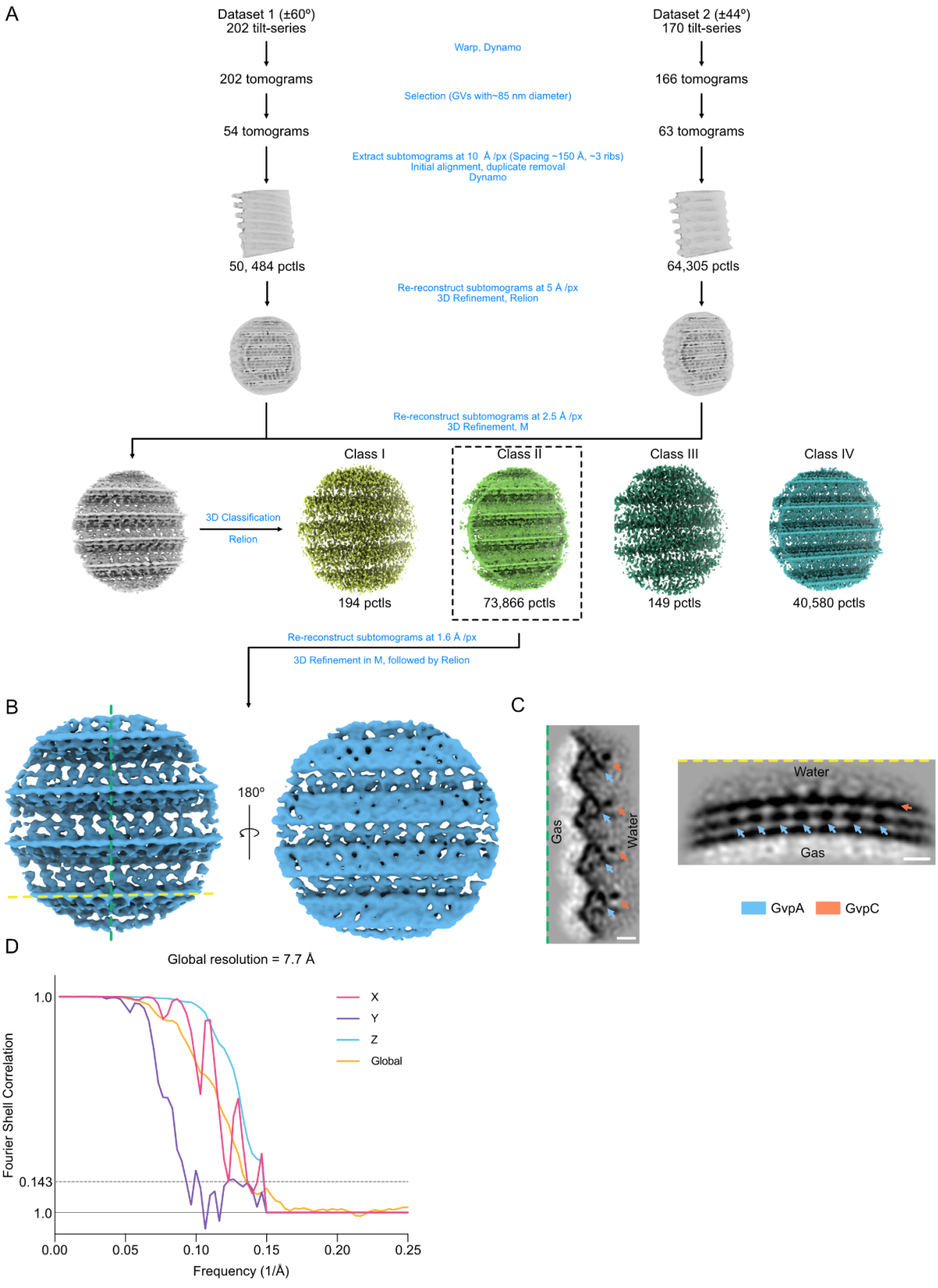

A
Dataset 1 (±60°)
202 tilt-series
202 tomograms
54 tomograms
50, 484 pctls
Dataset 2 (±44°)
170 tilt-series
166 tomograms
63 tomograms
64,305 pctls
Warp, Dynamo
Selection (GVs with~85 nm diameter)
Extract subtomograms at 10 Å /px (Spacing ~150 Å, ~3 ribs)
Initial alignment, duplicate removal
Dynamo
Re-reconstruct subtomograms at 5 Å /px
3D Refinement, Relion
Re-reconstruct subtomograms at 2.5 Å /px
3D Refinement, M
3D Classification
Relion
Class I
Class II
Class III
Class IV
194 pctls
73,866 pctls
149 pctls
40,580 pctls
Re-reconstruct subtomograms at 1.6 Å /px
3D Refinement in M, followed by Relion
B
180°
C
Gas
Water
Water
Gas
GvpA
GvpC
D
Global resolution = 7.7 Å
Fourier Shell Correlation
1.0
0.143
1.0
0.00
0.05
0.10
0.15
0.20
0.25
Frequency (1/Å)
X
Y
Z
Global

Figure 4-S5. CryoET data processing for AnaS GV shell. (**A**) Simplified schematic of the subtomogram averaging workflow highlighting crucial steps in the pipeline. (**B**) Orthogonal views of the final average. (**C**) Gray-scale orthographic slices of AnaS GV shell in the positions indicated by the blue and orange dashed lines in **B**. (**D**) Global FSC curve for AnaS GV shell (yellow), and map anisotropy analysis by FSC curves in X (pink), Y (purple) and Z (blue) directions.

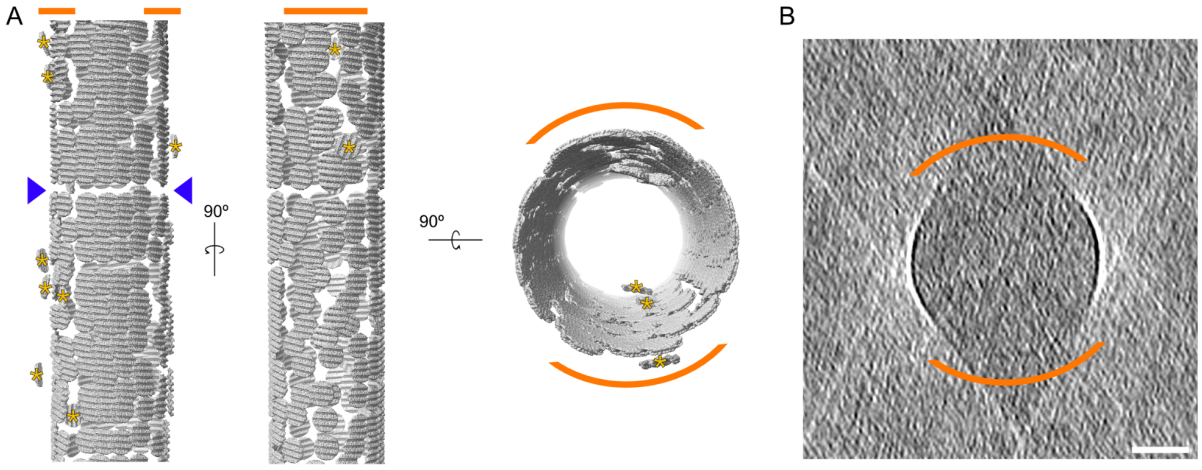


Figure 4-S6. Particle poses. (**A**) Orthogonal views of the particle orientations after subtomogram averaging for a representative tomogram of native Ana GV. Blue arrows indicate the position of the inversion point. Misaligned subtomograms are marked with yellow asterisks. (**B**) Tomographic slice showing that the missing wedge orientation corresponds with the most distorted particles. Scale bar, 25 nm. Orange lines indicate position of the missing wedge.

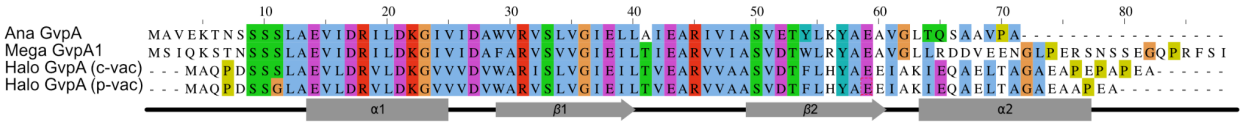


Figure 4-S7. Protein sequence alignment. Sequence alignment among homologs of the major structural protein (GvpA) from Mega, Ana, and Halo (p-vac and c-vac gene clusters).

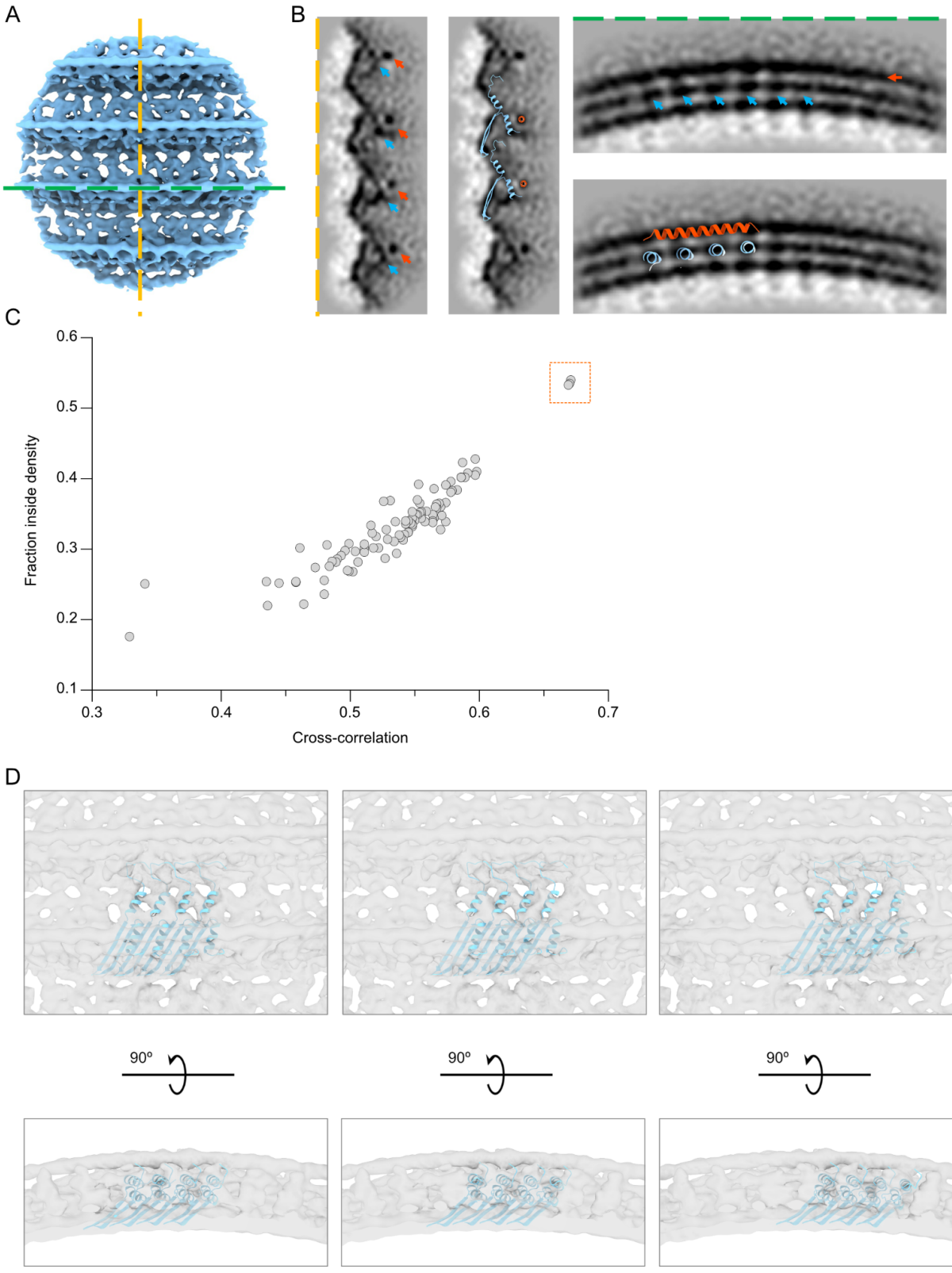

A
B
C
Fraction inside density
0.6
0.5
0.4
0.3
0.2
0.1
0.3
0.4
0.5
0.6
0.7
Cross-correlation
D
90°
90°
90°

Figure 4-S8. Model fitting validation. (**A-B**) Visual inspection of secondary structures matching with cryoET density. (**A**) Isosurface rendering. Dashed lines indicate positions of the orthographic slices visualized in **B**. (**B**) Gray-scale orthographic slices of the native Ana GV shell. Secondary structures of GvpA and GvpC are overlayed with the cryoET density visible in orthographic slices. Orange arrows indicate position of the GvpC and blue arrows positions of the GvpA fragments that bind GvpC. (**C**) Results of the rigid-body fitting of four GvpA molecules in Chimera using "fitmap" command. Three best scoring results are highlighted by the orange dashed box. (**B**) The three best results with similar fitting scores (orange dashed box in panel A) all fit our density map. Each fit from left to right are only different in that they shift by one subunit along Y-axis.

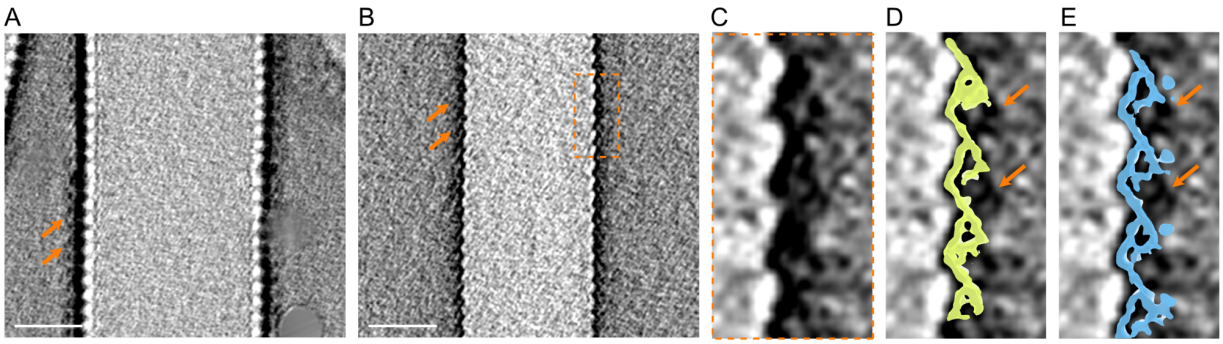


Figure 4-S9. Additional densities on the surface of Mega GVs. (**A**,**B**) Slices from cryo-electron tomograms of individual Mega GVs show additional density on the surface. Defocus values: (**A**) -5 μm and (**B**) -1 μm. (**C**) Enlarged section form **B** as outlined by orange dashed box. (**D, E**) Superimposition of subtomogram averages (Figure 3C and D) for AnaS and Ana GV shell. Orange arrows indicate extra densities. Scale bars, 20 nm.

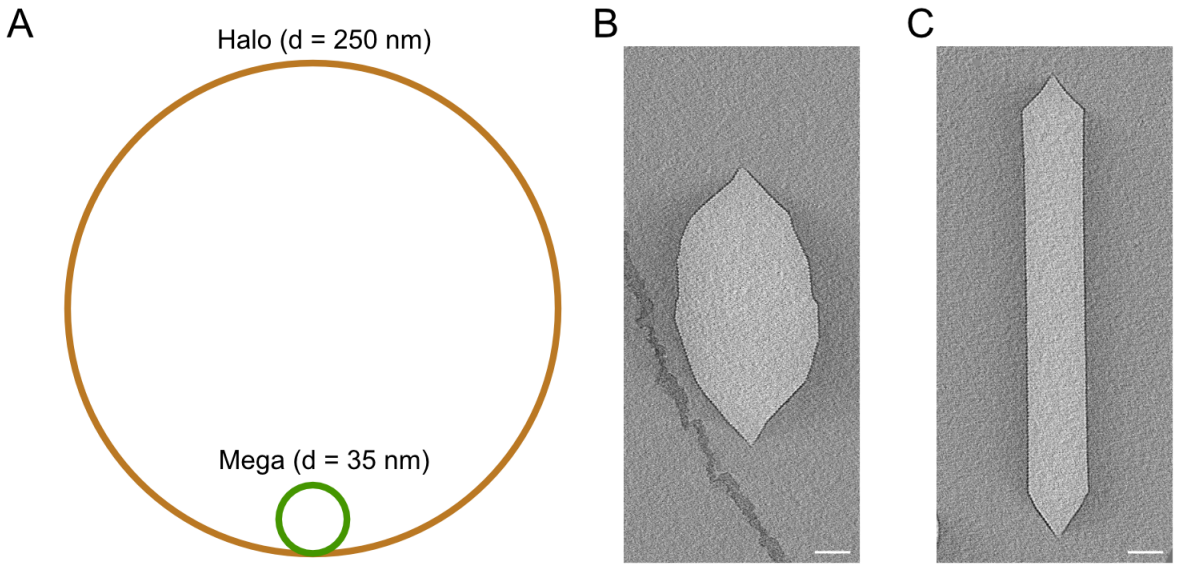


Figure 4-S10. GVs adopt a wide range of diameters and different morphologies. (**A**) Schematic showing difference in rib curvature between smallest (Mega) and largest (Halo) measured diameter(Dutka et al. 2021). (**B**,**C**) Representative central slices from cryo-electron tomograms of individual Halo GVs encoded by (**B**) p-vac and (**C**) c-vac gene clusters. Scale bars, 50 nm.

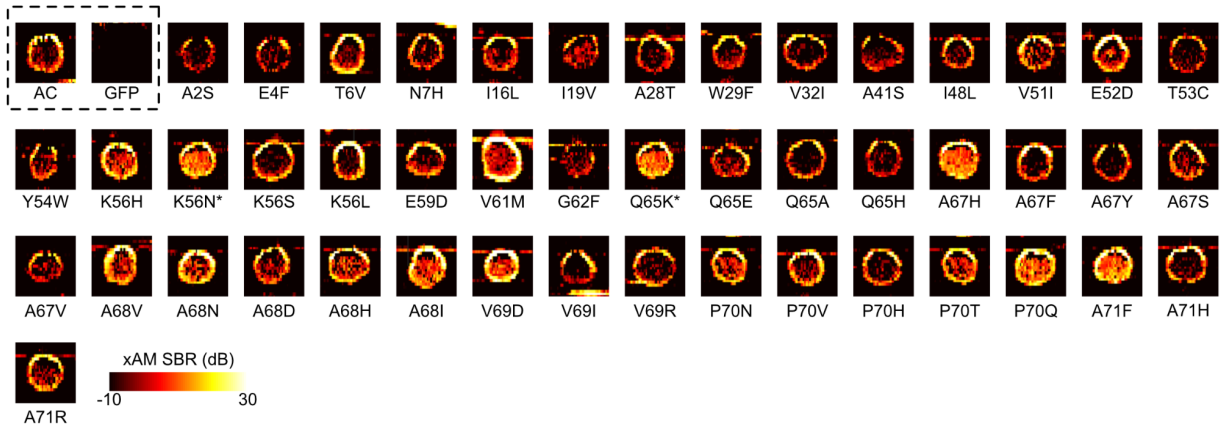


Figure 4-S11. Ultrasound images of *E. coli* clones expressing select GvpA mutants. Pre-minus-post-collapse nonlinear xAM images of clones of *E. coli* expressing GVs with the indicated mutations in GvpA. All the shown mutants display clear non-zero contrast and therefore successfully form GVs. Wild type GvpA (AC) and GFP are included as positive and negative controls, respectively. Color map corresponds to SBR, the signal-to-background ratio. *Mutations K56N and Q65K occurred in the same clone. GV expression is more pronounced on the edges of the patches because of those cells' increased access to nutrients.

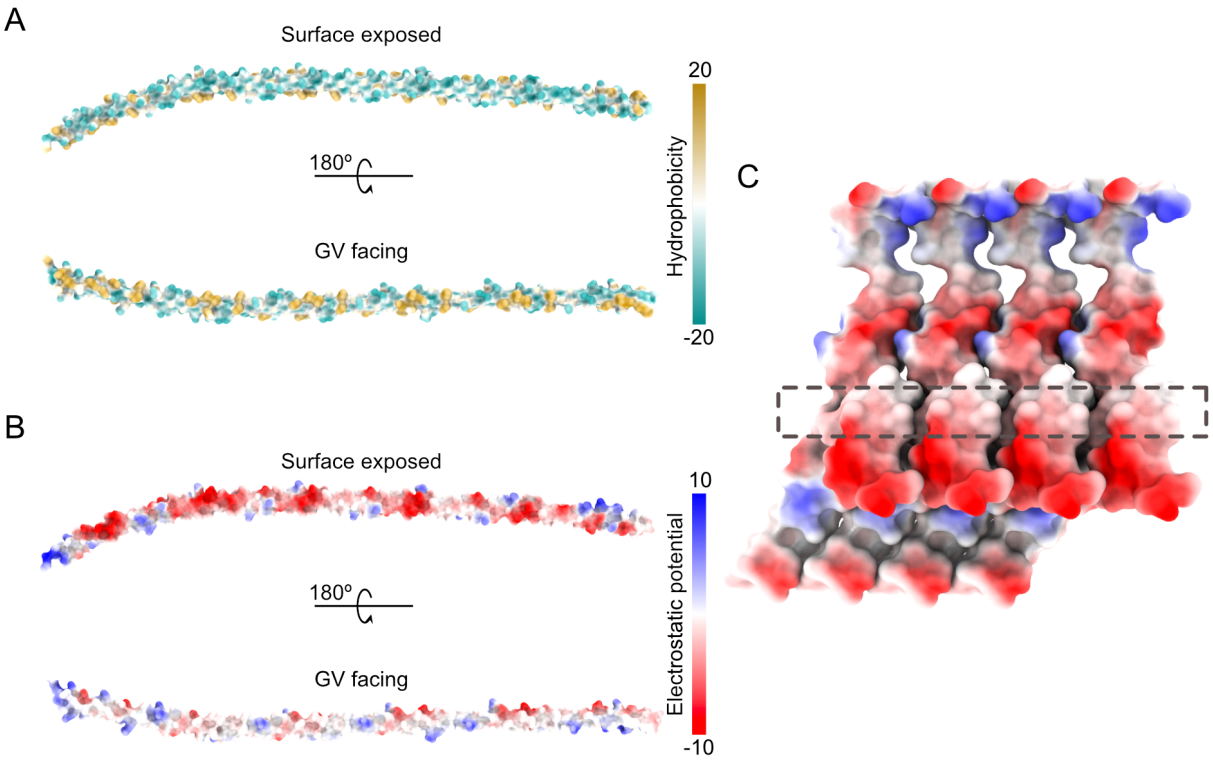


Figure 4-S12. Hydrophobicity and charge distribution on GvpC surface. (**A**,**B**) AlphaFold2 predicted models of full-length GvpC**.** (**A**) Hydrophobicity of the GvpC surface. (**B**) Distribution of the electrostatic potential of the GvpC surface. (**C**) Distribution of the electrostatic potential on the GvpC binding model.

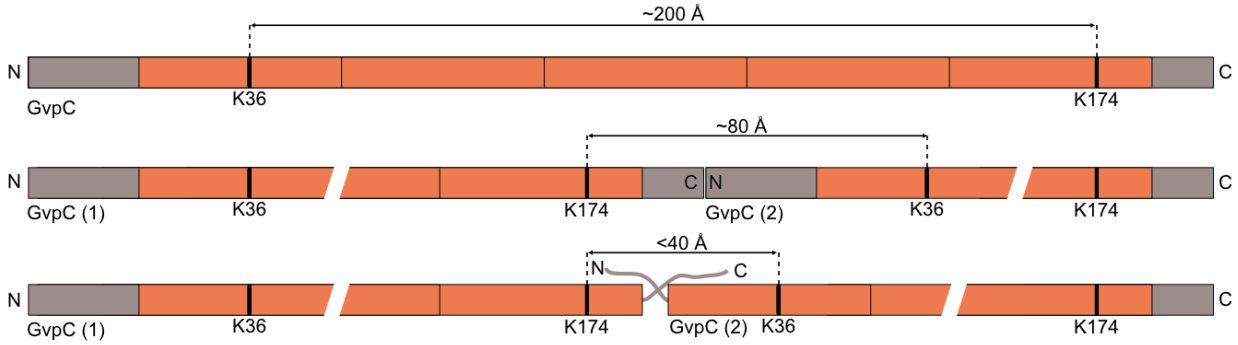


Figure 4-S13. Distances for different scenarios of Lys cross-linking between GvpC molecules.

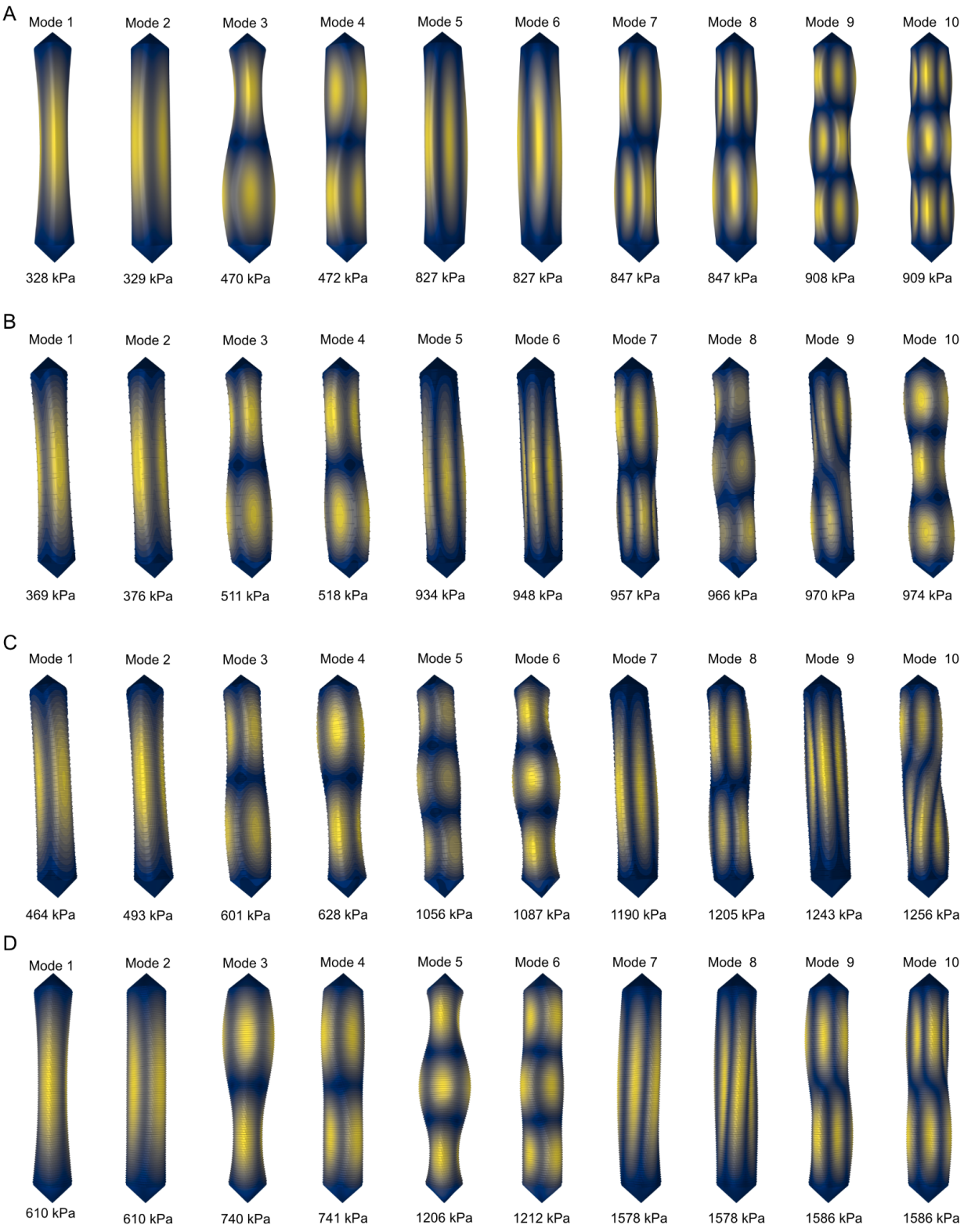


Figure 4-S14. Buckling modes of GVs with different degrees of GvpC saturation. The first ten buckling modes and pressures were obtained from linear buckling analysis for GV with distinct

saturation levels of GvpC. Rows from top to bottom represent GvpC densities of (**A**) 0%, (**B**) 20%, (**C**) 60%, and (**D**) 100%.

Table 4-T1. Data collection and processing parameters for GV shell.

	Native GVs	AnaS GVs
Magnification	53,000×	53,000×
Voltage (keV)	300	300
Energy Filter	Yes	Yes
Slit width (eV)	20	20
Pixel size (Å)	0.8435	0.8435
Defocus range (μm)	1.5 to 3.5	1.5 to 3.5
Defocus step (μm)	0.5	0.2
Tilt range (min/max, step)	-60/60°,3° or -44/44°,4°	-45/45°,3°
Tilt scheme	Dose-symmetric	Dose-symmetric
Total dose (electrons/Å^2)	~45	~45
Frame number	10 or 5	10
Tomograms used/acquired	127/368	28/103
Number of cylinders	136	32
Final subtomograms (no.)	73,866	28,887
Symmetry	C1	C1
Map resolutions (FSC = 0.143)	7.7 Å	7.3 Å

Table 4-T2. List of validated cross-linked peptides.

Cross-linker type	Protein 1	Sequence 1	Protein 2	Sequence 2	Cross-linked residues	Score
Cross-links						
BS^3	GvpC	**M**ISLMAK	GvpA	**M**AVEK	0-0	31.72
BS^3	GvpA	MAVE**K**TNSSSSLAEVIDR	GvpC	ISLMA**K**IR	5-7	53.19
DSSO	GvpC	QEHQSIAE**K**VAELSLETR	GvpA	**A**VEK	18-1	65.91
DSSO	GvpC	EFLSVTTA**K**R	GvpA	**A**VEK	36-1	60.36
DSSO	GvpC	IAQAE**K**QAQELLAFYQEVR	GvpA	**A**VEK	109-1	58.99
DSSO	GvpC	TAQA**K**EQK	GvpA	**A**VEK	174-1	43.9
Cross-links (inter- or intra- molecular)						
DSSO	GvpC	QEHQSIAE**K**VAELSLETR	GvpC	EFLSVTTA**K**R	18-36	102.98
DSSO	GvpC	TAQA**K**EQK	GvpC	ESLL**K**FR	174-182	86.63
DSSO	GvpC	EFLSVTTA**K**R	GvpC	TAQA**K**EQK	36-174	84.36
DSSO	GvpC	QEQAE**K**QAQELQAFYK	GvpC	EFLSVTTA**K**R	43-36	71.03
DSSO	GvpA	**M**AVEK	GvpA	**A**VEK	0-1	49.79
DSSO	GvpA	ILD**K**GIVIDAWVR	GvpA	**A**VEK	22-1	58.99

Chapter 5

STRUCTURE AND FUNCTION OF THE Dot/Icm T4SS

Adapted from:

Przemysław Dutka*, Yuxi Liu*, Stefano Maggi*, Debnath Ghosal, Jue Wang, Stephen D. Carter, Wei Zhao, Sukhithasri Vijayrajratnam, Joseph P. Vogel, Grant J. Jensen#. 2023. "Structure and Function of the Dot/Icm T4SS". *bioRxiv*, 2023. https://doi.org/10.1101/2023.03.22.533729

Abstract

The *Legionella pneumophila* Dot/Icm type IV secretion system (T4SS) delivers effector proteins into host cells during infection. Despite its significance as a potential drug target, our current understanding of its atomic structure is limited to isolated subcomplexes. In this study, we used subtomogram averaging and integrative modeling to construct a nearly-complete model of the Dot/Icm T4SS accounting for seventeen protein components. We locate and provide insights into the structure and function of six new components including DotI, DotJ, DotU, IcmF, IcmT, and IcmX. We find that the cytosolic N-terminal domain of IcmF, a key protein forming a central hollow cylinder, interacts with DotU, providing insight into previously uncharacterized density. Furthermore, our model, in combination with analyses of compositional heterogeneity, explains how the cytoplasmic ATPase DotO is connected to the periplasmic complex via interactions with membrane-bound DotI/DotJ proteins. Coupled with *in situ* infection data, our model offers new insights into the T4SS-mediated secretion mechanism.

5.1 Introduction

Bacterial type IV secretion systems (T4SS) are among the largest macromolecular complexes known and are present in most classes of bacteria (Backert and Grohmann 2018). T4SSs are used to translocate DNA and protein substrates across bacterial cell envelopes, often into other target cells in a contact-dependent manner (Li et al. 2019). For example, conjugative T4SSs drive bacterial conjugation by facilitating DNA transfer, an important process in the spread of antibiotic resistance (Dubnau and Blokesch 2019).

Legionella pneumophila is a Gram-negative bacterial pathogen that causes a form of pneumonia called Legionnaires' disease. There have been 10,000 reported cases of Legionnaires' disease in the USA so far with a 10% fatality rate (Dooling et al. 2015). *L. pneumophila* causes disease by surviving and replicating within alveolar macrophages (Hubber and Roy 2010; Qiu and Luo 2017). Intracellular survival is mediated by a specialized T4SS known as Dot/Icm (defective in organelle trafficking/intracellular multiplication). The Dot/Icm system exports over 300 effector proteins into host cells that modulate the host's cell biology, including its histone modification, transcriptome, proteome, and vesicle trafficking (Kubori and Nagai 2016). This results in the prevention of phagosome-lysosome fusion and the establishment of a vacuole permissive for bacterial replication.

In situ (Ghosal et al. 2017; Chang et al. 2018; Ghosal et al. 2019; Kaplan et al. 2022; Chetrit et al. 2018; Park et al. 2020) and *in vitro* (Chung et al. 2019; Durie et al. 2020; Sheedlo et al. 2021; Macé et al. 2022) structural studies of T4SSs have revealed their complexity and elegance. Our early cryo-electron tomography (cryo-ET) studies (Ghosal et al. 2017, 2019), and those of others

(Chetrit et al. 2018; Park et al. 2020), revealed the overall architecture of the intact Dot/Icm T4SS *in situ* and began to dissect the molecular organization of this complex (Ghosal et al. 2019; Chetrit et al. 2018). Single particle reconstructions then resolved the outer membrane complex (OMC) and periplasmic ring (PR) to near-atomic resolution, identifying new components and symmetry mismatches between the OMC and PR (Durie et al. 2020; Sheedlo et al. 2021).

To provide a more comprehensive understanding of the system, here we collected an order of magnitude more data and performed extensive subtomogram averaging with a focus on mobile connections within the system. Exploiting the improved quality of our maps, new structural and biochemical data, and advances in protein structure predictions, we construct a nearly complete molecular model of the *L. pneumophila* Dot/Icm T4SS. Our model incorporates over 460,000 atoms, more than double the number of atoms in any previous model of the system (Sheedlo et al. 2021; Ghosal et al. 2019; Chetrit et al. 2018; Park et al. 2020). Together with an analysis of compositional heterogeneity and infection *in situ*, our model sheds light on the T4SS-mediated secretion mechanism. We particularly highlight the role of IcmF as a primary component that forms a central hollow cylinder. We show that IcmF is a flexible protein spanning the inner membrane (IM) with a Ras-like domain on the cytosolic site. IcmF, and its interacting partner DotU, play a crucial role in polar targeting (Ghosal et al. 2019), biogenesis (Ghosal et al. 2019) and secretion of effectors (VanRheenen et al. 2004). Furthermore, we show that the cytosolic ATPases connect with the periplasmic complex via a linker domain formed by DotI:DotJ complexes and the collar density, connections we propose form an important signaling pathway between the cytoplasm and the OMC.

5.2 Results

Compositional Heterogeneity of the Dot/Icm T4SS

To further characterize the T4SS beyond our previous work (Ghosal et al. 2019), we collected a much larger dataset of approximately 900 tomograms, resulting in ~5,700 particles (**Figure 5-S1**). Because our previous work had shown that a fusion of superfolder green fluorescent protein (sfGFP) to the C-terminus of DotF (DotF-sfGFP) stabilizes the complex (Ghosal et al. 2019), we again imaged this strain. We also imaged a derivative strain with an additional DotB (E191K) mutation, which locks the cytoplasmic ATPase DotB onto membrane-associated DotO ATPase (Sexton et al. 2005; Ghosal et al. 2017, 2019; Chetrit et al. 2018). Subtomogram averages were produced using data combined from both strains.

One notable feature of the Dot/Icm T4SS and related systems is the unique architecture of the cytoplasmic ATPases (Chetrit et al. 2018; Macé et al. 2022). Typically, AAA+ ATPases assemble into hexameric structures, but in T4SS, DotO and its homologs (e.g. VirB4) form a unique hexamer of dimers. Each copy of DotO in a dimer is in a different conformation (Chetrit et al. 2018; Macé et al. 2022). In our previous work, the structure of the cytoplasmic complex was unclear, possibly due to compositional heterogeneity (Ghosal et al. 2017, 2019). Here we conducted extensive 3D classification to gain a better understanding of the composition and heterogeneity of the inner membrane complex (IMC) *in situ* (**Figure 5-S2**). After testing several programs, we obtained the best results using the I3 subtomogram averaging package (Winkler 2007).

We performed 3D classification using a simple cylindrical mask enclosing the entire T4SS. This approach allowed us to classify a fraction of the particles that lack the IMC (14%, Class I) (**Figure**

5-S2A). We then used the remaining particles to perform focused refinement and classification on the IMC. This approach revealed two additional classes of particles. In the largest class (72%, Class II), the arrangement of the ATPases was ambiguous. Although we could resolve a central barrel surrounded by external densities, their connectivity and symmetry was unclear. This class might reflect either a partially assembled state or misalignment of particles. The final class, which contained only 14% (Class III) of the original data set, exhibited the previously described hexamer-of-dimers arrangement (Chetrit et al. 2018). Notably, the presence of the hexamer of dimers was associated with a change in the IM curvature. However, no difference in the distance between the IM and outer membrane (OM) was observed, regardless of the cytoplasmic complex composition.

To confirm our classification results we analyzed individual particles imaged in thin *L. pneumophila* cell lamella prepared by cryogenic focused-ion-beam milling (cryo-FIB-milling). Substantial heterogeneity in the cytoplasmic complex was observed (**Figure 5-S2B**). We identified particles without IMC, with asymmetric IMC density, and with characteristic V-shaped IMC density connecting to the collar via the linker domain, confirming our 3D classes.

After performing focused refinement on the IMC in particle classes II and III, 13-fold symmetry was no longer apparent in the OMC, indicating that the periplasmic and cytoplasmic densities are not rotationally linked. To analyze the potential stabilizing effect of the linker connecting with the collar, we performed another round of focused refinement with a mask around the periplasmic complex and examined how much the particles moved. We found that particles shifted substantially more in class II where the linker and outer ATPases were not well-resolved (**Figure 5-S2C**) compared to class III where the linker, outer ATPases and collar were better resolved

(**Figure 5-S2D**). In both cases clear 13-fold symmetry was restored in the OMC. These results suggest that the linkers connect the periplasmic and cytoplasmic domains and stabilize the complex. This is consistent with previous cryo-ET data showing that the linker domain is only present after association of the IMC (Park et al. 2020).

Domain Architecture of the System

To obtain a structure of the complete T4SS, we performed a series of focused refinements. First we carried out focused refinement on the previously well-characterized periplasmic region (Ghosal et al. 2017; Chetrit et al. 2018; Ghosal et al. 2019). The resulting structure at ~21 Å resolution (**Figure 5-S3**) matched previously-published *in situ* subtomogram averages well in that region (Ghosal et al. 2017, 2019; Chetrit et al. 2018). The only notable difference compared with other subtomogram averages of that region was the presence of "elbow" density connecting the OMC and periplasmic ring (PR), which is likely due to the presence of sfGFP at the C-terminus of DotF. The collar domain appears to be connected to both the PR and elbow (**Figure 5-1A and 5-S4**). In *Helicobacter pylori*, the collar is also connected to the PR (**Figure 5-S4C**). Park et al. (2020) however suggested that the Dot/Icm T4SS collar connects directly to the OMC through a contact near the DotF-corresponding density (**Figure 5-S4D**).

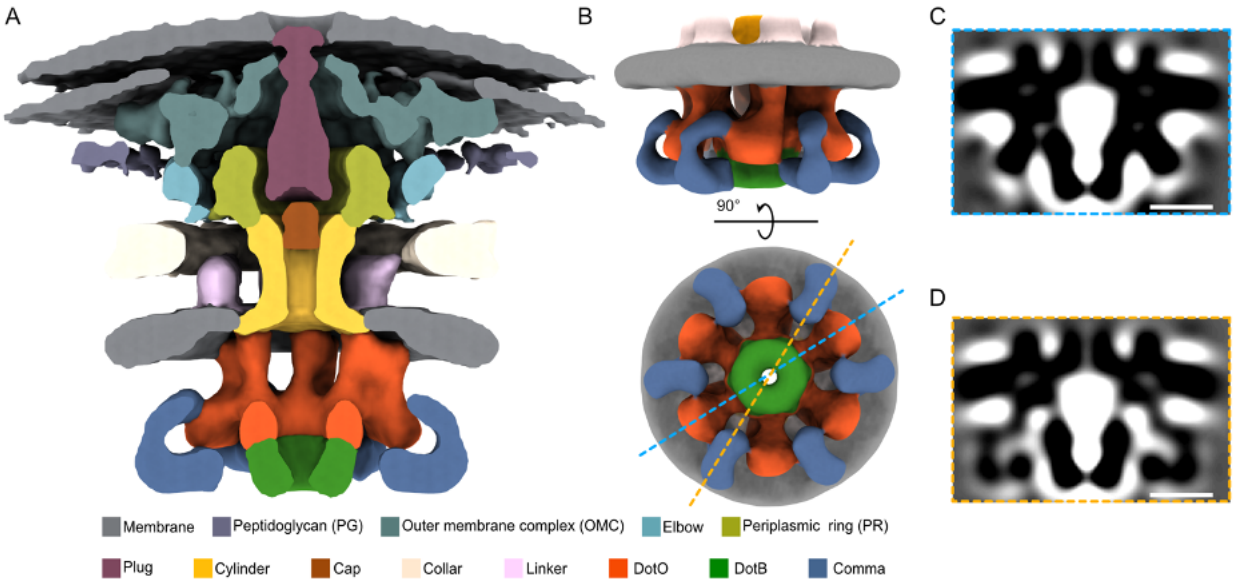


Figure 5-1. Architecture of the Dot/Icm T4SS. (**A**) Segmented side view of the composite cryo-ET map. (**B**) 3D rendering of the density from focused refinement on the IMC from Class III after applying 6-fold symmetry during post-processing highlighting the novel "comma" density (blue). Densities outside the IMC are masked out. (**C,D**) Central tomographic slices through the symmetrized IMC reconstruction as indicated by corresponding, color-coded, dashed lines in **B**. Scale bars, 10 nm.

Second, we performed focused refinement on the region containing the collar, linker, and a portion of IMC (**Figure 5-S5**). A characteristic feature of the T4SS and related systems are symmetry mismatches between individual components. The best characterized mismatches are in the OMC and PR (Durie et al. 2020; Sheedlo et al. 2021). More recently, another symmetry mismatch was proposed between the collar and linker domain of Dot/Icm T4SS (Park et al. 2020). In our average, we observe the transition from a 13-fold symmetry in the collar to 6-fold symmetry in the linker (**Figure 5-S5B-C**). The linker domain emerges from the IM and bends towards the collar density (**Figure 5-S5A**). Our ability to resolve different symmetries within the same refinement suggests that the elements may be structurally linked.

Third, we performed focused refinement on the IMC in Class III and noticed previously undescribed additional density between the outer ATPases (**Figure 5-1B-D**). In our initial 3D classification, we did not apply any symmetry and the initial alignment used particles with randomized orientations along the symmetry axis of the particle. Even in the asymmetric reconstruction, the uncharacterized density is highly regular (**Figure 5-S6A**). To obtain a higher-quality average, we applied 6-fold symmetry to the map (**Figure 5-1B-D**). The uncharacterized density emerges near the bottom of the outside DotO protomers and bends downward in the shape of a comma. While this density has not been described before, we observed a weaker but related structure in the prior map of the IMC from the Liu group (EMD-7612) (Chetrit et al. 2018) (**Figure 5-S6B-C**). In that map, however, the comma density disappears quickly with increasing threshold whereas it does not in our map. One possible explanation is that the presence of the DotF-sfGFP fusion stabilized complete particles and/or allowed better classification (Ghosal et al. 2019). Finally we merged the maps, using the best-resolved detail at each position, to produce the composite cryo-ET map of the system (**Figure 5-1A** and **5-S1**).

Integrative Model of the Dot/Icm T4SS

The improved quality of our cryo-ET maps, newly-available high-resolution structures of some Dot/Icm T4SS components (Sheedlo et al. 2021), and the recent structure of a related conjugative T4SS (Macé et al. 2022), combined with advances in protein structure prediction (Jumper et al. 2021) (**Figure 5-S7**) offered us the opportunity to update and extend the molecular model of the *L. pneumophila* Dot/Icm T4SS (**Figure 5-2** and **Movie 5-S1**). We began by fitting previously known and characterized components of the T4SS. First, we docked the single-particle reconstruction of the OMC and PR (Sheedlo et al. 2021) into the corresponding densities in our

cryo-ET map (**Figure 5-S8**), accounting for nine components of the system: DotC, DotD, DotF, DotG, DotH, DotK, Dis1, Dis2, and Dis3. As mentioned above, the elbow is most likely a result of the DotF-sfGFP fusion. Consistent with this prediction, we were able to fit 26-copies of sfGFP into the elbow density, resulting in a ring sandwiched by the DotF copies facing in opposite directions (**Figure 5-S9**). In the single-particle reconstruction, only the C-terminus of DotF was resolved in both OMC and PR (Sheedlo et al. 2021). Since DotF is known to self-interact (Vincent et al. 2006), we built homodimers in ColabFold. We predict that the N-terminal domains of DotF dimerize, forming a coiled-coil structure, and then penetrate the IM with the transmembrane domain, with potentially flexible N-termini exposed in the cytoplasm (**Figure 5-S9D**). We show this detail in **Figure 5-S9D**, but because corresponding density is not seen in our cryo-ET map, allowing us to place these structures confidently (likely because of flexibility), we did not include the N-terminal domains of DotF in the deposited model shown in **Figure 5-2**.

Previously, we identified IcmX as a major component of the plug density by comparing difference maps between wild-type and various T4SS mutants (Ghosal et al. 2019). We ran structure predictions for several different oligomeric states of IcmX, with the best results for a tetramer or pentamer (**Figure 5-S10A-C**). The IcmX pentamer had a higher modeling score, as determined by ColabFold (Mirdita et al. 2022), and occupied a larger volume fraction in our density map. Additionally, we noticed a similarity in shape between the IcmX plug density and the stalk density from the conjugative T4SS (Macé et al. 2022) (**Figure 5-S10D**). Notably, the stalk density is generated by VirB5, which is also a pentamer. In addition, IcmX and VirB5 occupy similar "on-axis-in-the-periplasm" positions in the Dot/Icm and conjugative T4SSs, respectively, further indicating their structural and functional relationship. Taking all of this into account, we modeled

IcmX as a pentamer into our final model (**Figure 5-2** and **5-S10E-F**).

Between the PR and the IM there is a hollow cylinder. In our previous research, we proposed that DotG may be a crucial component of this cylinder (Ghosal et al. 2019). No cylinder was seen, however, in the single-particle reconstructions of the purified OMC and PR complex/complexes, which clearly contained DotG (Sheedlo et al. 2021; Durie et al. 2020). Later cryo-ET data showing different assembly stages of the Dot/Icm T4SS revealed one stage where the OMC, with a fully-formed dome structure consisting of DotG also had no signs of a cylinder (Park et al. 2020). Moreover, the N-terminus of DotG is formed by a variable number of 10 amino-acid repeating sequences (**Figure 5-S11A**) (Gomez-Valero et al. 2019). The location of the repeats in DotG, along with their variable number, suggests that DotG does not form the cylinder. Instead, we believe it extends from the PR to form the collar and then penetrates the IM (**Figure 5-S11B**). We show this detail in **Figure 5-S11**, but the predicted structure for the N-terminus of DotG is low confidence, so we did not dock it into the collar region of our cryo-ET map.

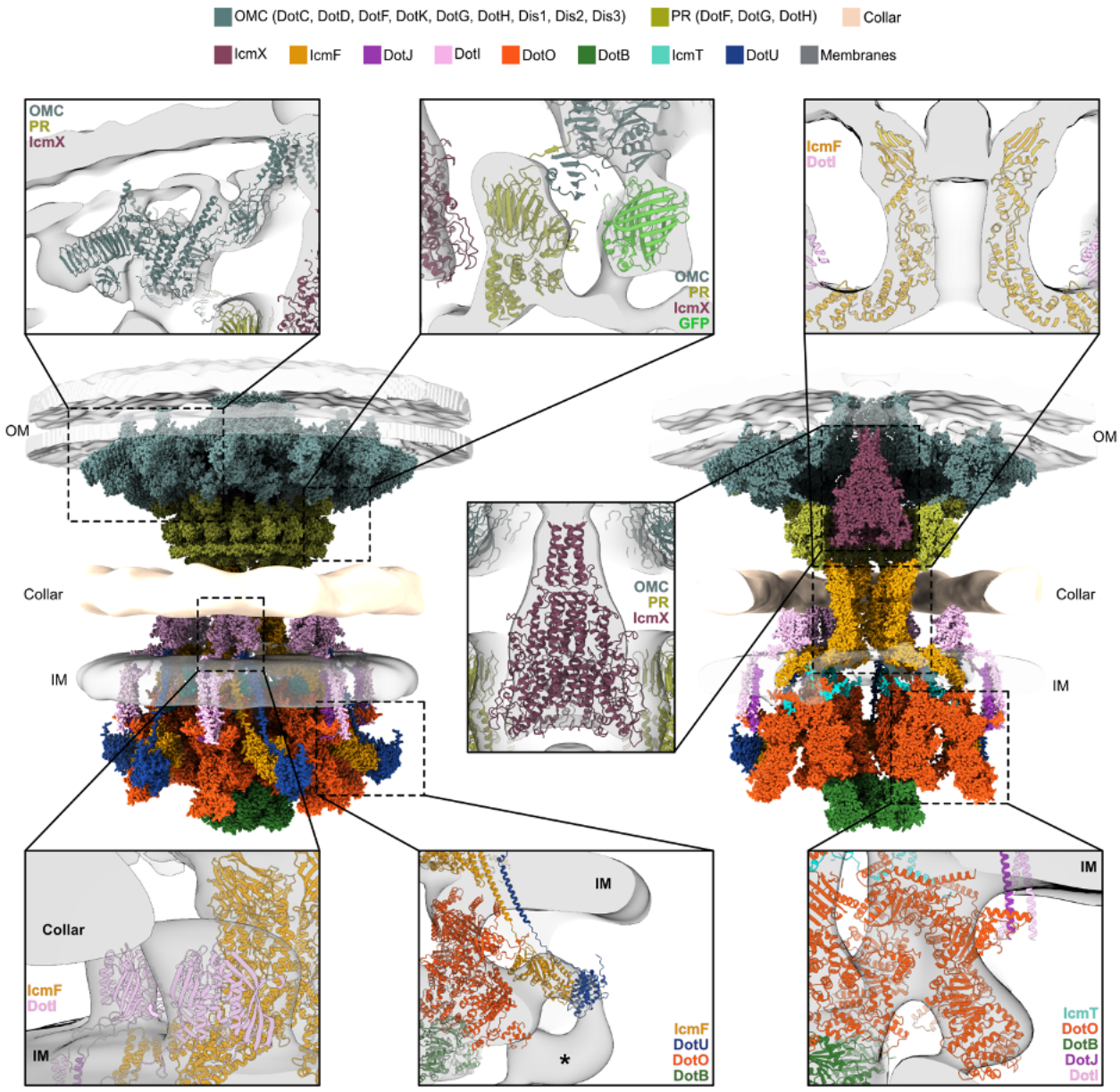


Figure 5-2. Molecular model of *L. pneumophila* Dot/Icm T4SS. Integrative model constructed by docking experimental and predicted protein structures into the *in situ* cryo-ET density maps. The OMC and PR structures are from PDB 7MUS (Sheedlo et al. 2021), while the DotB structure is from PDB 6GEB (Prevost and Waksman 2018). The remaining atomic models are AlphaFold/ColabFold (Jumper et al. 2021; Mirdita et al. 2022) predictions. OM – outer membrane; IM – inner membrane; * – unmodeled portion of the uncharacterized comma density in the IMC. The OM, IM, and collar are illustrated as isosurfaces.

Two additional candidates that could form the cylinder are DotA and IcmF. Given that DotA is a predicted polytopic membrane protein, and assembles into ~10 nm hollow rings upon secretion by

planktonic bacteria (Nagai and Roy 2001), it is unlikely to be a major component of the cylinder. IcmF shares sequence homology with TssM, which forms a decameric (pentamer of dimers) channel within the type VI secretion system (T6SS) (**Figure 5-S12A-B**) (Ghosal et al. 2019). The AlphaFold prediction for the structure of the IcmF monomeric structure has three domains: a cytoplasmic domain, a transmembrane domain, and a periplasmic domain (**Figure 5-S12C-D**). The transmembrane domain for IcmF was not annotated in UniProtKB, but DeepTMHMM (Hallgren et al. 2022) confirmed its presence (**Figure 5-S12C**). Six subunits of the C-terminal domain of IcmF fit well into the cylinder density of our cryo-ET map (**Figure 5-2** and **5-S12E-F**). In T6SSs, TssM (homolog of IcmF) and TssL (homolog of DotU) interact and function to recruit T6SS baseplate proteins to the T6SS membrane complex (Ma et al. 2012; Logger et al. 2016; Zoued et al. 2016). In the T4SS DotU and IcmF are co-dependent on each other for stability (Sexton et al. 2004) and ColabFold's $IcmF_{cyto}$:DotU dimer model predicts the proteins also interact within their transmembrane helices and cytosolic portions (**Figure 5-S13A-E**). Modeling in these transmembrane domains, the $IcmF_{cyto}$:DotU complex is well-positioned to occupy the upper portion of the comma density (**Figure 5-2** and **5-S13F**).

The IMC is primarily composed of two ATPases, DotO and DotB. DotO is a homolog of VirB4 and when associated with the T4SS complex exists in a unique hexamer of dimer formation at the inner membrane (Chetrit et al. 2018; Macé et al. 2022). DotB is a homolog of VirB11 and is recruited to the DotO dodecamer in an ATP-dependent manner (Sexton et al. 2005). Since DotO does not possess a transmembrane domain, it likely requires another protein to recruit it to the membrane. In the conjugative T4SS, VirB4 is anchored to the IM via interactions with VirB3 and VirB8. Six VirB3 molecules bind to the inner six subunits of VirB4, whereas trimers of VirB8 interact with the outer six subunits of VirB4 (Macé et al. 2022). VirB8 trimers further extend to

the periplasm via a transmembrane helix where they form periplasmic arches, stabilized by interactions of the C-terminal domain (Macé et al. 2022). In the Dot/Icm T4SS, we found a similarity between IcmT and VirB3 (**Figure 5-S14A**). DotI/DotJ were previously recognized as homologs of VirB8 (**Figure 5-S15A**) (Ghosal et al. 2019).

Unfortunately, the resolution of our map is not sufficient to model a protein as small as IcmT on its own. As a result, we built a model of the DotO:IcmT subcomplex with ColabFold and fit it to the density map (**Figure 5-S14**). The predicted model shares high structural similarity with the homologous structure from the conjugative T4SS. The completed model of the inner DotO:IcmT hexamer fits the cryo-ET map well (**Figure 5-2** and **5-S14J**). In the conjugative T4SS, each VirB8 trimer creates a periplasmic arch with its N-terminal helices and then elongates a cytoplasmic domain that binds the VirB4 ATPase (**Figure 5-S15B**) (Macé et al. 2022). Since biochemical data indicates DotI and DotJ exist as a heterotrimer with a 2:1 DotI:DotJ stoichiometry (Kuroda et al. 2015), we first built a model of a DotI:DotJ heterotrimer (**Figure 5-S15**). We then used it to model the DotI:DotJ:DotO subcomplex, thus allowing us to optimize its position within the cryo-ET map. Finally, we positioned the crystal structure of DotB (Prevost and Waksman 2018) beneath the inner DotO hexamer (**Figure 5-2** and **5-S16**), guided by structural and biochemical data (Park et al. 2020; Sexton et al. 2005).

Several transmembrane domains of the Dot/Icm proteins appear to be artificially located below the IM in our final model (**Figure 5-S17A**). In our cryo-ET map of class III (fully assembled complex) (**Figure 5-S2A**), the IM appears to be less curved than in many individual particles (**Figure 5-S2B**). It is likely that the discrepancy we observed is due to variability in the curvature of the IM

among individual particles, resulting in the flattening and blurring of the IM in the subtomogram average. As a result, the leaflets of the bilayer are not resolved (as they are in the OM) (**Figure 5-S17A**). As we have previously noted, the curvature of the IM is correlated with the resolvability of the linker domain. To investigate this further, we evaluated the effect of the linker-forming proteins (DotI:DotJ) on the membrane curvature prediction. In their absence, the predicted membrane is flat (**Figure 5-S17B**), similar to the class I and II averages (**Figure 5-S2A**). However, the addition of DotI:DotJ (**Figure 5-S17C**) results in a membrane curvature resembling that of the class III average (**Figure 5-S2A**). These results are consistent with changes in the IM curvature being associated with the docking of the DotO ATPase via the DotI:DotJ linker.

Functional adaptations of the IcmF/TssM protein family

TssM has been extensively studied in the context of T6SSs (Rapisarda et al. 2019; Yin, Yan, and Li 2019). Analysis of the cytoplasmic domain revealed the presence of nucleotide binding motifs, and biochemical data confirmed nucleotide hydrolysis activity (Sexton et al. 2004; Ma et al. 2012). Further detailed sequence analysis revealed that IcmF/TssM belongs to the P-loop NTPase family and contains Walker A and Walker B motifs, typically associated with GTPase activity (Ma et al. 2012). By searching for structural homologs of $IcmF_{cyto}$, we discovered a remarkable similarity with eukaryotic small GTP-binding proteins (Ras) (**Figure 5-3A-C**), a class of proteins that act as molecular switches to quickly turn on and off signaling pathways. Specifically, the AlphaFold model of $IcmF_{cyto}$ shares strong similarity with Arf1, a small G protein that regulates membrane traffic in eukaryotic cells (Renault et al. 2003). Notably, the homology includes highly conserved crucial residues involved in magnesium and GDP binding (**Figure 5-3B-C**). Given this similarity, combined with the observation that TssM is capable of hydrolyzing GTP (Ma et al. 2012), we propose that the cytoplasmic domain of IcmF/TssM proteins contain a Ras-like domain.

Another potentially crucial structural adaptation of IcmF is the presence of hinge regions. In the cryo-EM structure of TssM, a hinge region was identified in the middle section of the channel, allowing the protein to rotate and adopt two distinct conformations (Durand et al. 2012; Rapisarda et al. 2019). A second hinge region was proposed at the base of the periplasmic domain of TssM (Rapisarda et al. 2019). Both regions are conserved in IcmF (**Figure 5-3D-F** and **5-S12A**). IcmF flexibility might play a crucial role in substrate transport through the cylinder and/or contribute to the small cap-like density observed on top of the cylinder in our cryo-ET map (**Figure 5-1**).

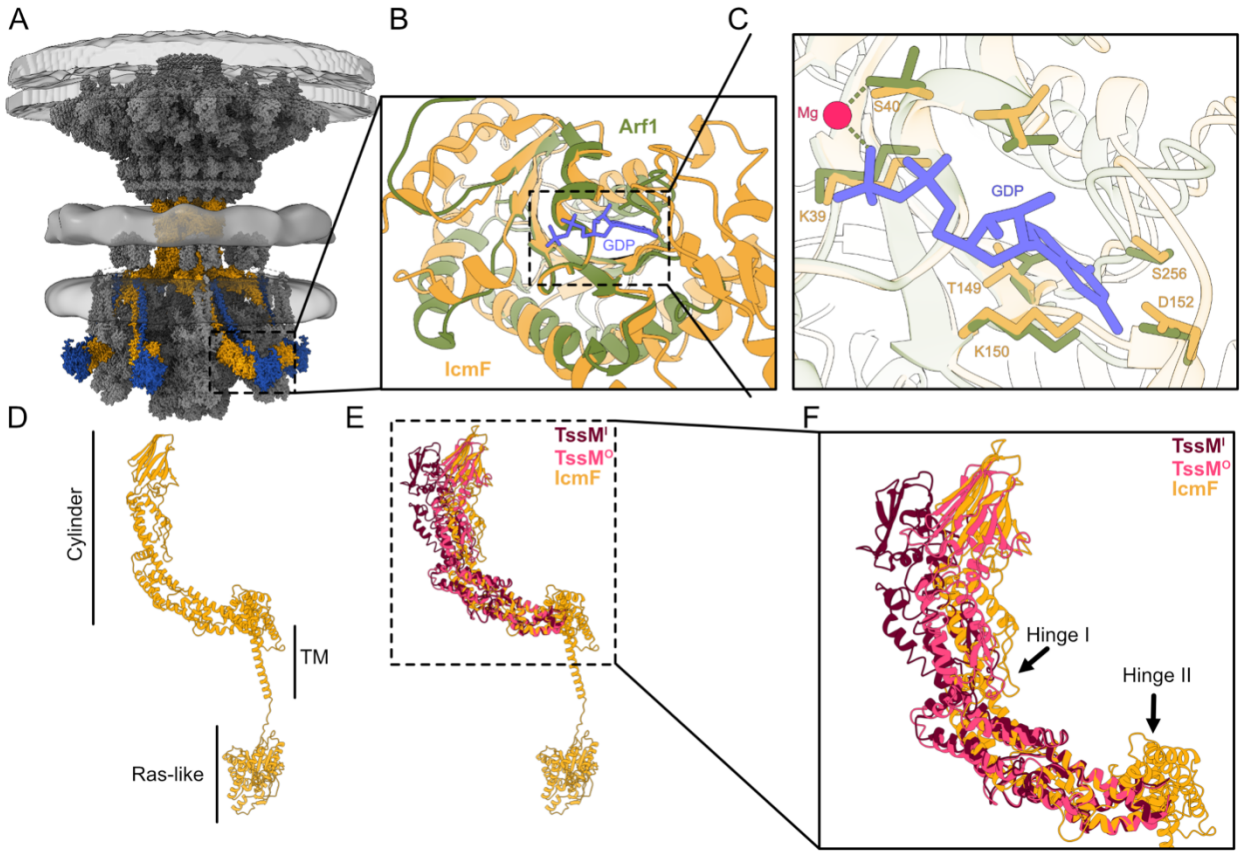


Figure 5-3. IcmF is a flexible protein with a Ras-like domain. (**A**) Composite model of the T4SS with IcmF and DotU highlighted. (**B**) Cytoplasmic domain of IcmF, which is superimposed with the crystal structure of eukaryotic small G-protein Arf1 (PDB 1R8Q, shown in green)(Renault, Guibert, and Cherfils 2003). (**C**) Structural preservation of the nucleotide-binding site in the IcmF predicted model. A GDP molecule bound to Arf1 and residues critical to nucleotide

binding are shown in sticks following the same coloring scheme. (**D**) IcmF domain annotation. (**E**) Structure alignment of IcmF model with cryo-EM structures of TssM from T6SS in two conformations. $TssM^{I}$ and $TssM^{O}$ denote molecules forming the inner and outer parts of the T6SS core complex channel, respectively. (**F**) Enlarged view of the cylinder/channel-forming portion of the IcmF and TssM indicating potential hinge regions.

T4SS tethers to the host membrane during early infection

Our atomic model is based on inactive T4SSs expressed in planktonic bacterial cells. We therefore went on to investigate the architecture of actively secreting *L. pneumophila* T4SSs within infected hosts. For host cells, we used human U937 monocytic cells, as they serve as an established model system for the study of *L. pneumophila* macrophage interaction (*34*). In order to slow substrate translocation, and potentially stabilize actively secreting Dot/Icm T4SS complexes that contain a substrate, we challenged U937 cells with *L. pneumophila* cells expressing a chimera of the T4SS substrate RalF with a tightly folded version of ubiquitin (V76G). *Legionella* containing vacuoles (LCV) could be observed associating with mitochondria and ER, typical of infections with wild-type cells (**Figure 5-4** and **5-S18**) (Escoll et al. 2017). Ninety-five percent of the T4SS particles found (49/52) exhibited a similar conformation as planktonic T4SS. However, 3 particles were observed in a clearly different state with tethers between the bacterial OM and the membrane of the LCV (**Figure 5-4** and **5-S18**). Similar sites were previously reported in *L. pneumophila* infections of amoeba, suggesting the mechanism of translocation is conserved in the mammalian system (Böck et al. 2021). Notably, the IM in the 3 particles with tethers adopted a sharp curvature similar to that seen in class III planktonic T4SS (**Figure 5-S2A-B**), suggesting that the IMC is fully assembled when the T4SS is tethered to the LCV. In contrast to the planktonic class III particles, additional density could be observed between the OMC and PR of the three particles with tethers (**Figure 5-4** and **5-S18**), likely representing T4SS substrates.

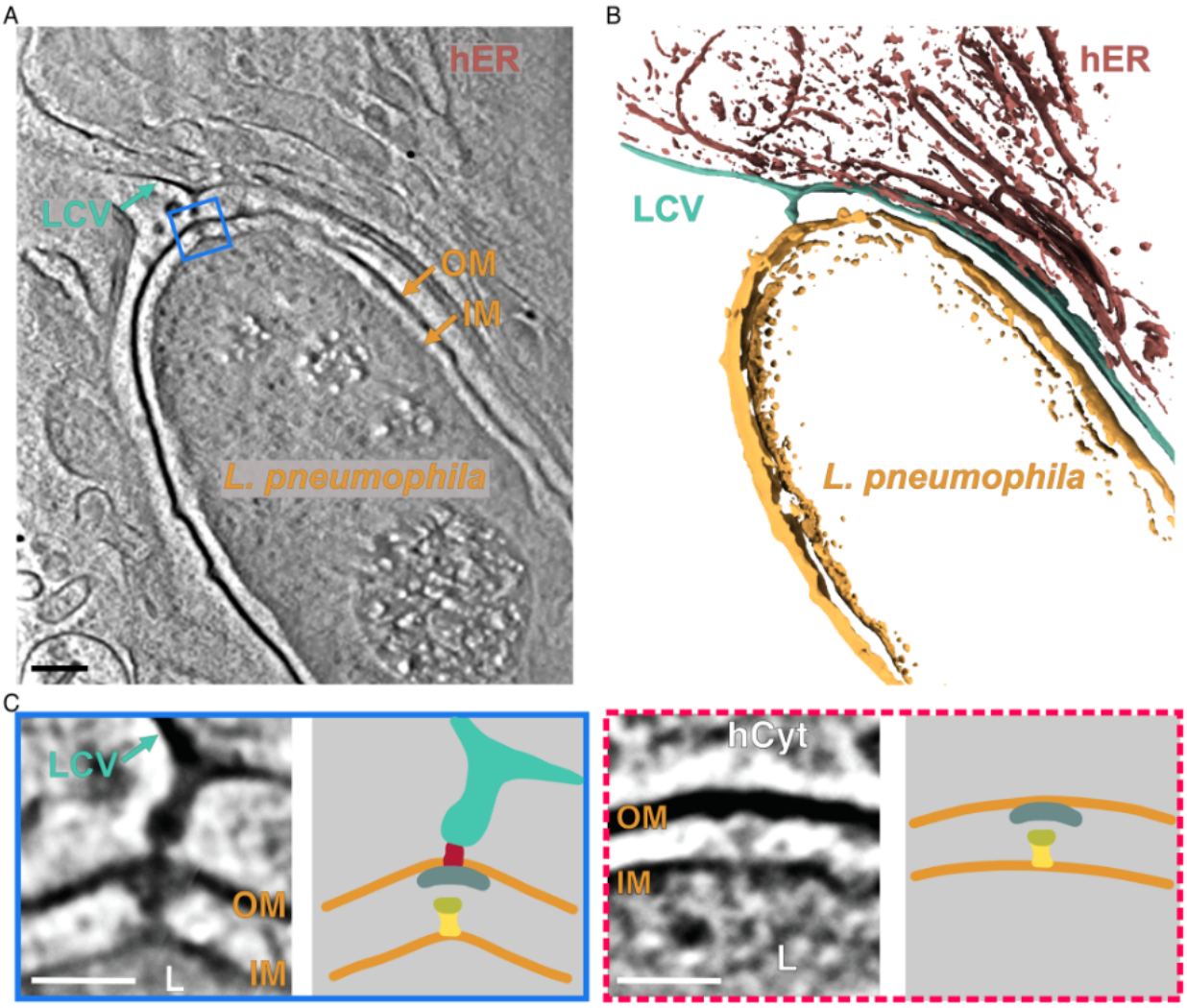


Figure 5-4. *L. pneumophila* T4SS tethers to the host membrane during early infection. Representative tomographic slice (**A**) and segmented model (**B**) illustrating a T4SS observed inside *L. pneumophila* cells infecting human U937 cells at 1 h post-infection. OM (orange) – *L. pneumophila* outer membrane; IM (orange) – *L. pneumophila* inner membrane; LCV (cyan) – *L. pneumophila* containing vacuoles. L (orange) – *L. pneumophila* cells; hER (brown) – Endoplasmic reticulum of U937 cells. Scale bar, 100nm. (**C**) Enlarged T4SS particles identified in A (blue frame) and a different cell shown in Figure 5-S17C (red dashed frame). The particle tethered to the host membrane is enclosed by the blue frame; the particle without tethering site is indicated by the red dashed frame. The panels on the right depict the individual particles making the OMC, PR, and the cylinder schematically. Scale bars, 50nm.

4.3 Discussion

The *L. pneumophila* Dot/Icm T4SS is a large macromolecular machine composed of ~30 proteins. Twenty-two of the proteins form the core system and 8 are part of a distinct membrane-associated complex, the Type IV Coupling Protein (T4CP) complex (Meir et al. 2020; Lockwood et al. 2022; Sheedlo et al. 2022). Of the 22 proteins forming the core system, our model contains 17 proteins, with over 460,000 total atoms, comprising one of the largest and most complex macromolecular structures modeled to date. Compared to the structure of the conjugative T4SS published last year (Macé et al. 2022), our model revealed four major novelties, which are discussed individually below (**Figure 5-5**).

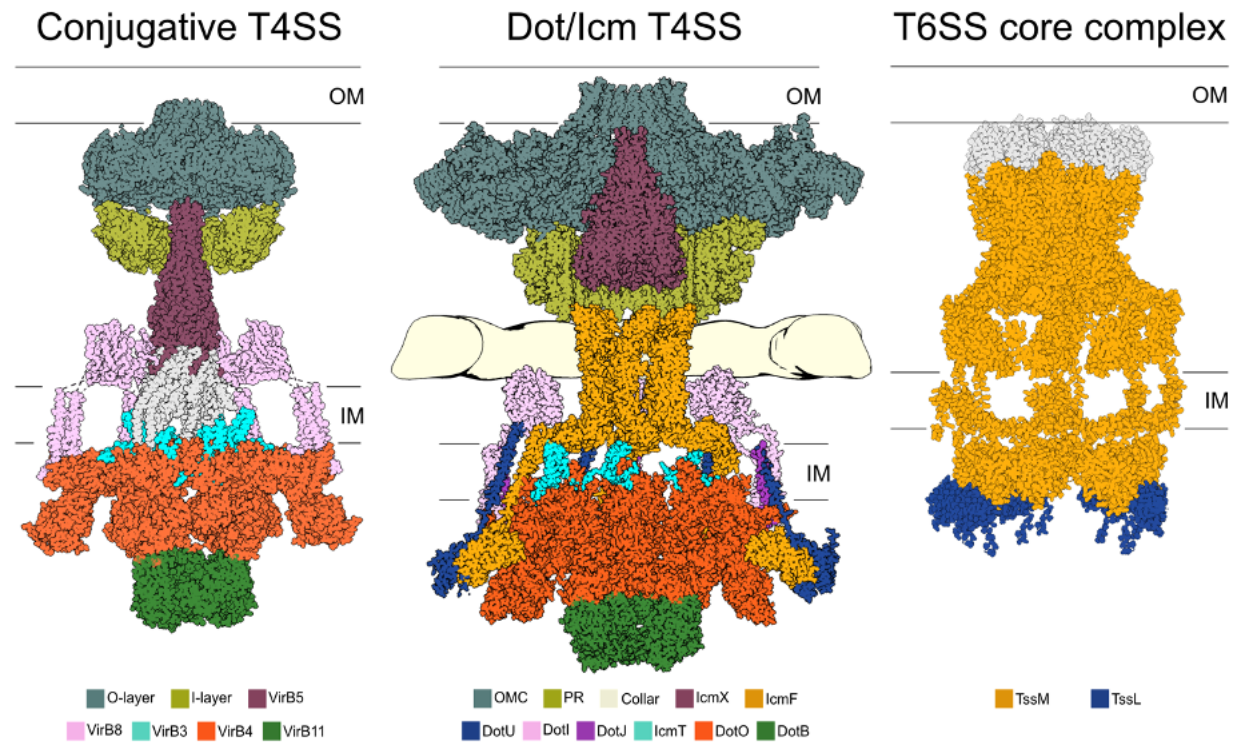


Figure 5-5. *L. pneumophila* Dot/Icm T4SS is a hybrid system. Left panel – atomic model of the conjugative T4SS (PBD 7O3J, 7O3T, 7O3V, 7O41) (Macé et al. 2022) with VirB11 model (PDB 2GZA, shown in green) (Hare et al. 2006) placed as proposed by Mace et al. (2022). Central panel – integrative model of the *L. pneumophila* Dot/Icm T4SS (this study). Right panel – composite model of T6SS core complex. Periplasmic domain (TssJ/TssM) is from the single particle cryo-EM reconstruction (PDB 6IXH) (Yin et al. 2019), the cytosolic domain of TssM (yellow) is an

AlphaFold prediction, and TssL is from a crystal structure (PDB 3U66) (Durand et al. 2012). Cytosolic domains of T6SS are placed based on our previous structure of the T6SS (EMD-8600) (Chang et al. 2017). Homologous components between the systems are marked with the same colors.

IcmF cylinder

The first difference is the presence of IcmF forming a central hollow cylinder in the Dot/Icm T4SS in place of the conjugative T4SS component VirB6. In the conjugative system, VirB6 forms a pentameric, conical platform on the axis of the system adjacent to the IM (Macé et al. 2022). Mace et al. (2022) proposed that VirB6 scaffolds the assembly of the pilus, which is made up of VirB2 subunits. In contrast, the Dot/Icm system does not appear to encode a VirB6 homolog or any pilin subunits, nor has a pilus structure been observed. Although the Dot/Icm T4SS appears to have been derived from an IncI plasmid conjugation system, additional adaptations were apparently required to secrete the wider variety of substrates (over 300) exported by *L. pneumophila* into host cells (Gomez-Valero et al. 2019; Gomez-Valero and Buchrieser 2017). These changes included the incorporation of new components such as secretion chaperones and DotU/IcmF, which are homologs of the T6SS factors TssL/TssM (Nguyen et al. 2018; Durand et al. 2012). Since the Dot/Icm system does not contain a VirB6 homolog, and the periplasmic portion of IcmF is located in a similar place as VirB6 in the conjugative system (**Figure 5-5**), we propose the VirB6 platform in conjugative T4SSs was replaced with IcmF in the Dot/Icm T4SS. In T6SSs, TssM forms a channel just above the IM through which the Hcp rod of the T6SS (a large cargo) is projected (Wang et all 2019); in the Dot/Icm T4SS, IcmF appears to form a central hollow cylinder through which substrates traverse from the IMC to more distal portions of the T4SS complex in the periplasm. In addition, both IcmF and TssM contain hinge regions, which may be critical to the export of larger sized cargo.

Comma density, IcmF$_{cyto}$ and GTP hydrolysis

The second major difference observed in our model is the presence of a cytoplasmic density which we called the comma. We assigned the upper portion of this density to the cytoplasmic domains of IcmF and DotU. TssL/TssM, homologs of DotU/IcmF, function to stabilize the T6SS complex and recruit an assembly platform of proteins known as the baseplate (Ma et al. 2012; Logger et al. 2016; Zoued et al. 2016). Similarly, DotU/IcmF stabilize the T4SS complex as part of their role in mediating optimal Dot/Icm substrate export (Sexton et al. 2004). In contrast to TssL/TssM, however, DotU/Icm perform an additional function as they are necessary and sufficient to target the Dot/Icm T4SS to the bacterial poles (Ghosal et al. 2019). Based on the interaction of TssL/TssM's cytoplasmic domain with other T6SS baseplate components, it is likely that DotU/IcmF's cytoplasmic domains bind another factor(s) and this may account for the density in the lower portion of the comma. Potential candidates for this factor(s) include a polar-localizing determinant, an as-yet-unidentified Dot/Icm component, and/or possibly an early-translocated Dot/Icm substrate.

We further found that IcmF$_{cyto}$ consists of a Ras-like domain, but the role of GTP hydrolysis remains unclear. One possibility is that nucleotide hydrolysis by IcmF$_{cyto}$ is required for the proper targeting and/or assembly of the Dot/Icm T4SS complex. Alternatively, nucleotide hydrolysis could mediate conformational changes in the cytoplasmic ATPases and cylinder, allowing substrates to engage with the complex and/or cause the IcmF cylinder to open. Finally, secretion through the T4SS is contact-dependent (Alvarez-Martinez and Christie 2009; Cascales and Christie 2003; Christie 2016), and our *in situ* data shows T4SS particles making contact with the

host membrane. Another possibility is therefore that host contact induces a signal that is transmitted through the periplasmic components of the Dot/Icm T4SS to $IcmF_{cyto}$, resulting in GTP hydrolysis and recruitment of other factors to the complex.

Collar

The third major difference is the presence of the periplasmic density we called the collar. We propose that a portion of the N-terminus and/or the middle domain of DotG containing the repeats forms the collar for several reasons. First, in the *Helicobacter pylori* Cag T4SS, the resolved fragment of CagY, the DotG homolog, in the PR faces the collar, which hints that the unresolved repeating fragment could extend further to form the collar (Sheedlo et al. 2020). This is further suggested by a well-resolved connection between the PR and collar in our earlier cryo-ET structure of the Cag T4SS (Chang et al. 2018). Second, we note that there is no similar repeat-rich, middle domain of VirB10, the DotG homolog in conjugative systems, and no collar has been seen in structures of those systems (Macé et al. 2022; Khara et al. 2021; Hu et al. 2019). Finally, there are 13 copies of DotG in the OMC, and we found that the collar also has 13-fold symmetry. Unfortunately, the low resolution of the collar density and the lack of confident structural predictions for these domains of DotG prevented us from modeling this region. The purpose of the collar remains unclear. One intriguing possibility is that the repeat-rich middle domain of DotG allows the T4SS cylinder to expand in either width or height to accommodate large conformational changes during substrate export or penetrating a host cell membrane.

Plug

The fourth major difference is the plug, composed of IcmX in the Dot/Icm system. Interestingly, both IcmX and VirB5 form pentameric cone-shaped structures on-axis in the periplasm. VirB5 is later found at the tip of extended pili (Macé et al. 2022), however, and the Dot/Icm system is not thought to build a pilus. The IcmX pentamer is also located closer to the OM than VirB5. Other differences include the Dot/Icm T4SS being only partially assembled in an Δ*icmX* mutant, with no cylinder and an altered conformation of the PR (Ghosal et al. 2019), and IcmX being released from planktonically grown bacteria (Matthews and Roy 2000; Segal, Feldman, and Zusman 2005). Based on these observations, we hypothesize that IcmX functions either as a scaffold for assembly of the Dot/Icm T4SS and/or as a plug in non-secreting T4SSs, waiting to be expelled at the onset of substrate export.

DotI:DotJ linkers

One similarity between the conjugative and Dot/Icm systems is the presence of a hexameric set of linkers spanning the IM. Previously, we reported that the linker/collar region (formerly referred to as "wings") appears flexible (Ghosal et al. 2019). The linker formed by DotI:DotJ complexes provides a docking site at the IM for the outer six subunits of the DotO dodecamer. Based on their location, another potential role of DotI:DotJ could be sensing conformational changes from the outer DotO subunits during substrate transport, propagating them through the collar to the OMC/PR, thereby allowing for plug release and secretion of effector proteins.

Secretion pathway

While here we have emphasized the novel role of IcmF in forming a hollow cylinder through which substrates pass, the entire Dot/Icm system appears to be an extended channel. Effectors likely first

interact with the T4CP DotL ATPase, then pass through the opening of the dodecameric DotO ATPase, across the IM, through IcmF and finally the PR/OMC. Although data suggest that unfolded T4SS substrates are preferred (Sheedlo et al. 2022; Amyot et al. 2013), we could not identify any obvious hydrophobic patches along the secretion path that would protect hydrophobic fragments of unfolded targets from refolding. Considering the wide variety of effector proteins secreted through the T4SS (Lockwood et al. 2022), it is possible that some substrates are secreted in folded or partially-refolded forms. In support of that notion, expulsion of IcmX would create a chamber of sufficient size to allow small proteins to refold.

Integrative modeling

Much of cell biology is driven by large, multi-subunit flexible complexes that are challenging or impossible to purify or reconstitute, so may only be available for study within intact cells. Here we have demonstrated an integrative approach to characterize such structures. New cryo-ET instrumentation, methods, and software (Chreifi et al. 2021; Chreifi et al. 2019; Mastronarde 2005; Winkler 2007) made it possible to image the T4SS directly *in situ* (Oikonomou and Jensen 2017) and overcome stoichiometric variance by collecting and classifying thousands of particles. Focused refinement mitigated the effects of flexibility. Recent developments in single-particle cryo-EM methods (Durie et al. 2020; Sheedlo et al. 2021) and structure prediction with AlphaFold (Jumper et al. 2021; Mirdita et al. 2022; Evans et al. 2022) provided atomic models of individual components and sub-complexes. The docking of these structures into our cryo-ET maps illustrates what is becoming possible in a new era of integrative modeling.

5.3 Materials and Methods

L. pneumophila strains preparation

L. pneumophila strain (JV9082) expressing a DotF-sfGFP from the chromosome was constructed by allelic replacement into strain Lp02 using the suicide plasmid pJB7255 (DotF-sfGFP). pJB7255 was constructed by cloning two PCR products (generated by PCR using primers JVP2973/JVP2990 and JVP2992/JVP2993, respectively) into the R6K suicide plasmid pSR47S (**Table 5-T1**). The PCR products were digested with BamHI/NotI and NotI/SacI, respectively, and ligated into BamHI/SacI-digested pSR47S.

L. pneumophila strain (JV9298) expressing DotB(E191K) and DotF-sfGFP from the chromosome was constructed by allelic replacement into strain JV9082 using the suicide plasmid pJB7487. This plasmid was constructed by subcloning a ~1.2 kb fragment (BamHI/SalI) containing DotB (E191K) from plasmid pJB2920 (Sexton et al. 2005) into BamHI/SalI-digested pSR47S.

Sample preparation for cryo-ET

L. pneumophila strain (JV9082 and JV9298) was cultured as previously described (Ghosal et al. 2017). In short, *L. pneumophila* were grown to stationary phase in CYE media supplemented with 150 ug/mL ferric nitrate, 400ug/mL of L-cysteine and 100ug/mL of thymidine. Bacteria cells were collected and resuspended in the same growth media to OD_{600} = 6.0. Resuspended culture was mixed with 10 nm gold beads as fiducial markers and then applied onto glow-discharged copper R2/2 200 mesh Quantifoil holey carbon grids (Electron Microscopy Sciences) in Vitrobot Mark IV (Thermo Fisher Scientific). Grids were plunge frozen with liquid ethane/propane mixture (Tivol et al. 2008) and stored at liquid nitrogen temperature.

Cryo-ET data collection

Tilt series were collected at 300 keV at pixel sizes 2.65 Å/pix on a Titan Krios (Thermo Fisher) equipped with a K2 or K3 camera. One-hundred micron objective aperture was inserted during collection and a 20 eV slit energy filter (GATAN EFTEM) was used to improve contrast. For whole cell tomograms, either cell pole was targeted for collection. Tilt series were collected bi-directionally from 0 deg to 60 deg on each side, with 2 or 3 deg steps. A total of 100 electrons/Å^2 was distributed evenly across the tilt series. Tilt images were collected either as a single image or a stack of 10 frames in the super resolution mode.

Tomograms reconstruction and subtomogram averaging

Raw image stacks were motion corrected with MotionCor2 (Zheng et al. 2017) and assembled into tilt series with the IMOD newstack program (Kremer et al. 1996). For subtomogram averaging in EMAN2, tilt series alignment and subsequent tomogram reconstruction was performed automatically within the EMAN2 tomography pipeline (Chen et al. 2019). EMAN2 reconstructed 4* binned tomograms were used in the IMOD 3dmod program to manually pick particles. An effort was made to include as many top views as possible. Particle coordinates were picked to be centered visually at the point where the cylinder meets the IM; an orientation was also recorded to have the particles aligned as closely as possible to what is shown in **Figure 5-1A**. The center and orientational information were then converted to particle records in EMAN2. Subsequent subtomogram averaging was performed using the manual alignment information as the starting point in EMAN2. We used 2* binned particles for the initial round of averaging, focusing on the density in the periplasmic space with C13 symmetry enforced. Then aligned particles were allowed to move locally with a cylindrical mask containing the whole particle. The alignment information was used to extract unbinned particles and the same local refinement was repeated. Next, sub-tilt

refinement was performed to obtain the best StA map for the whole particle.

For subtomogram averaging in Dynamo (Castaño-Díez et al. 2012) and I3 (Winkler 2007), tilt series alignment was performed either manually or automatically in Etomo on 2* binned tilt images. We used tomo3d (Agulleiro and Fernandez 2015) to reconstruct 2* binned weighted back-projection tomograms from the aligned tilt series. Then we used an in-house script to transfer the aligned particle coordinates into a Dynamo compatible form (script available: https://doi.org/10.5281/zenodo.7659617) and performed StA in Dynamo with a mask enclosing the entire particle and C13 symmetry enforced. The aligned particles were extracted again from the tomograms. These sub-tomograms were treated as individual tomograms containing one particle in I3. The particle alignment orientation was also transferred into I3 with an in-house script (script available: https://doi.org/10.5281/zenodo.7679253). Initially I3 StA was carried using a soft-edge cylindrical mask enclosing the entire system to exclude particles without assembled IMC using classification. The rotational angle around the C13 symmetry axis of the remaining particles were randomized and the particles were further classified using a soft-edge cylindrical mask around the IMC. The class with the fully assembled IMC complex was used for further focus refinement on either the IMC or connection between the cytoplasm and periplasm. At any step of data processing in I3 we did not apply symmetry to the reference map. Final reconstructions with well-resolved symmetry were eventually symmetrized to aid model building. Crucial steps in the data processing pipeline are summarized in **Figure 5-S1**. The resolution estimation is performed on the unfiltered, unsymmetrized maps using ResMap (Kucukelbir et al. 2014) (**Figure 5-S3**). The composite map was created in ChimeraX using *vop maximum* command (Goddard et al. 2018).

Subtomogram averaging – data analysis

To test symmetry of the collar, the cryo-ET map from focused refinement on collar/IM connection was further masked out to only include the collar ring. Rotational cross-correlation was calculated using *dynamo_symmetry_scan* function in Dynamo (Castaño-Díez et al. 2012) (**Figure 5-S5B**).

Using an in-house python script (script available: https://doi.org/10.5281/zenodo.7679388), particle shift between two iterations was calculated as the total shift of the particle center as recorded in the I3 output after each iteration .

Cryo-FIB milling

L. pneumophila strain (JV9298) was cultured similarly as for the whole cell tomography. Grid preparation follows the same procedure as described above, except that bacterial cells were resuspended to OD_{600} = 20.0 and applied onto glow-discharged copper R2/2 200 mesh Quantifoil holey carbon grids. Excess material on the grid was removed by manual blotting from the back before plunge-freezing. Frozen grids were mounted into Autogrids (Thermo Fisher Scientific) and transferred into an FEI Versa 3D equipped with a Quorum PP3010T Cryo-FIB/SEM preparation system (Quorum Technologies LLC, East Sussex, UK) operating at cryogenic temperatures. Samples were sputter coated with a thin layer of platinum (15 mA, 60 sec) and then imaged with the SEM at 10 keV. Identified areas with multiple layers of bacterial cells were further subject to stepwise milling with 30keV gallium ions, starting from 0.3 nA beam current and gradually stepping down to 10 pA, to generate final lamellae approximately 150-300 nm thick (Lam and Villa 2021). Tomograms were reconstructed in IMOD (Mastronarde and Held 2017) and processed with IsoNet (Liu et al. 2022).

Infection assay for cryo-ET

Human U937 cells (ATCC CRL-1593.2) were grown in RPMI1640 supplemented with 10% FBS and L-glutamine for several passages and then subjected to a 72-h incubation with 12-O-tetradecanoylphorbol-13-acetate (TPA) for differentiation. Differentiated U937 cells became adhesive and were seeded onto gold London Finder grids (Au H2 R2/2, Quantifoil) coated with vitronectin (25 μg/mL in 1xPBS, pH 7.4), allowing for a 48 h incubation at 37˚C. *L. pneumophila* Lp02 expressing CyaA tagged RalF attached to an irremovable ubiquitin chain (Lp02 CyaA-RalF::V76G) (Amyot et al. 2013) was induced at OD_{600} = 2 to 2.5 with 250 μM IPTG for 4-5 h. U937 cells were infected at an MOI of 100 for either Lp02 CyaA-RalF::V76G or JV9082 and vitrified after 1 h in the presence of 10 nm gold beads as fiducials. Tilt-series were collected bi-directionally at 33 000 x equipped with the Volta Phase Plate from -54˚ to 54˚ with an increment of 3˚. A total of 100 electrons/A^2 was distributed evenly across the tilt-series. Each tilt-image was collected in the super resolution mode either as a single image or a stack of frames. Tilt-series were phase corrected and reconstructed with SIRT in IMOD (Zheng et al. 2017). Tomograms were further processed with IsoNet (Liu et al. 2022) and manually segmented with Amira (ThermoFisher).

Modelling

All the maps were segmented using the Segger package in ChimeraX (Goddard et al. 2018; Pintilie et al. 2010; Pintilie and Chiu 2012). Available experimental structures for OMC/PR (Sheedlo et al. 2021) and DotB (Prevost and Waksman 2018) were fitted in the map using the *fitmap* command. The atomic models of the remaining components were predicted using AlphaFold/ColabFold (**Figure 5-S7**) (Jumper et al. 2021; Evans et al. 2022; Mirdita et al. 2022). We used the R388 T4SS structure (PDB 7OIU) (Macé et al. 2022) as scaffold for the positioning of DotO, IcmT and

DotJ:DotI$_2$ atomic models. Detailed information on placing Dot/Icm T4SS components into our density maps are described below.

OMC/PR

The single particle cryo-EM of the OMC/PR (Sheedlo et al. 2021) was fitted as a rigid body into the corresponding density in the cryo-ET map from focused refinement on the periplasmic complex (**Figure 5-S8**). Fitting was guided by the 13-fold symmetry of the OMC.

DotF/GFP

DotF consists of a TM in position 53-75 followed by a coiled-coil region (135-176). The C-terminal portion of DotF has been solved (208-269) and is occupying two different positions: a C13 ring beneath the Dis3 protein of the OMC and a C18 ring in the PR (Sheedlo et al. 2021). DotF IM localization and its ability to self-interact has been reported before (Vincent et al. 2006). Considering the coiled-coil region, we build a DotF dimer in ColabFold. The model dimerizes in this region with long flexible linkers connecting to the N- and C- terminal. We positioned the coiled-coil region of this model halfway between two DotF C-terminal domain solved structure and the TM helix in density corresponding to the IM. We build the linkers in MODELLER (Sali and Blundell 1993) and symmetrized it with the respective symmetry of the PR and OMC (**Figure 5-S9**).

Given the DotF-sfGFP high local concentration the sfGFP may dimerize or form high molecular weight aggregates (Stepanenko et al. 2016). To our knowledge the only available structure of a native GFP dimer is the extra-superfolder GFP (PDB 5B61) (Choi et al. 2017) a close homolog of

the sfGFP superimposing with a RMSD of 1.8 Å (Choi et al. 2017). We fitted the structure in the collar by rigid docking and applied C13 symmetry (**Figure 5-S9**).

IcmX

IcmX as a main component of the plug was identified based on the analysis of the T4SS mustants (Ghosal et al. 2019). However, symmetry of the IcmX is unknown. For that reason, we tested structure predictions for multiple symmetries as generated by ColabFold. The best results were obtained for tetramer and pentamer with overall scores for pentamer being higher. Thus, we fitted IcmX pentamer as a rigid body into the plug density in ChimeraX (**Figure 5-S10**).

DotG

DotG repetition were identified using the MOTIF search tool of the Japanese GenomeNet service (Kanehisa 1997). As previously reported (Gomez-Valero et al. 2019) five copies of the Pentapeptide_4 repeat motif was found. Given the low score ranging from 3.1e-06 to 6.8e-12 and the Alphafold DotG model not showing the classic right-handed β-helical structure typical of the pentapeptide repeat motif we investigated the repetition sequence. The GenomeNet identified repetition were aligned in MEGA using MUSCLE (Edgar 2004a, 2004b). For this multiple alignment we were able to identify a 10 amino acid pattern, as previously described(Vincent et al. 2006), starting with a Lys or Arg and ending with a Leu or Iso. We therefore extracted those sequence and built a sequence logo using WebLogo3 (Crooks et al. 2004) to identify the motif sequence. This motif ([KR]-X-A-G-[FY]- [STND]-X-X-X-[LI]) is different form the pentapeptide repeat sequence ([STAV][DN][LF][STR][G]) and explain the discrepancy of the DotG AlphaFold model with the solved structures. In order to investigate the fold of this repetition we scanned the PDB database with this motif but only the DotG protein was found.

IcmF:DotU

IcmF and DotU proteins are present in both T4SS and T6SS (TssM and TssL, respectively) (Durand et al. 2012). The T6SS proteins have been extensively characterized and the structure of both proteins has been solved (Yin et al. 2019; Rapisarda et al. 2019; Robb et al. 2012; Ruiz et al. 2020; Chang and Kim 2015; Durand et al. 2012). The TssM monomeric structure is characterized by the presence of three domains spanning across the inner membrane. The periplasmic portion consists of four helix-bundle domains connecting a β-sheet rich domain, involved in TssJ binding. The periplasmic domain is connected by a TM helix to the cytoplasmic ATP binding domain (Ma et al. 2012). TssL:TssM interaction involves a 110-aa TssM extension of the nucleotide binding domain (Logger et al. 2016) and a cleft on the TssL structure (Asp74, Glu75, Gly137 and Phe138) (Zoued et al. 2016).

We closely inspected IcmF and DotU AlphaFold models for T4SS and T6SS shared features. In the IcmF Uniprot page there is no TM helix annotated, but DeepTMHMM (Hallgren et al. 2022) predicted, with high confidence, a TM in position 317-338 (**Figure 5-S12C**). While a C-terminal TM helix is predicted for DotU. Sequence alignment of DotU with TssL revealed conservation of the residue involved in TssM binding (**Figure 5-S13A**). Given the high conservation of IcmF and TssL predicted interacting regions we run AlphaFold multimer. In the $IcmF_{Cyto}$:DotU dimer model the proteins are interacting with the cytosolic portion of IcmF and DotU with the two transmembrane helices in a good orientation. We built the IcmF cylinder by manually placing it in the map region forming the central cylinder and applying C6 symmetry (**Figure 5-S12**). We then manually placed the $IcmF_{cyto}$:DotU in the newly identified comma density and manually rebuilt the IcmF linker and reoriented the DotU TM helix to cross the membrane density in our map (**Figure 5-S13**).

DotO:IcmT

The DotO biological assembly consists of DotO dimers arranged in a hexamer (Park et al. 2020; Macé et al. 2022). Interaction between DotO and IcmT allows DotO-membrane tethering. We therefore built a $DotO_2$:IcmT trimer atomic model. ColabFold modeled the DotO dimer in the isoform involved in hexamer assembly rather than the dimer formation. Therefore we used R388 T4SS structure (PDB 7OIU) (Macé et al. 2022) as a template to position the outer DotO subunits. Superimposition of the DotO:IcmT with VirB4(TrwK):VirB3(TrwM) revealed a good RMSD between the two structures and similar positioning of IcmT in comparison to TrwM (**Figure 5-S14F**). A DotO monomer was further aligned with its VirB4 (TrwK) counterpart to generate the dimer. The $DotO_2$:DotT trimer was fit into the map from focused refinement on the IMC using Powerfit (Zundert and Bonvin 2015) and C6 symmetrized in ChimeraX. This generated a clash between DotO C-terminal domains in the inner protomers. In the Colabfold DotO dimer this region is predicted to fit in a cleft region of the other subunit (**Figure 5-S14G**). We used this prediction to remodel the clash.

DotI:DotJ

To model the DotI:DotJ subcomplex we submitted to ColabFold a 2:1 DotI:DotJ assembly based on the available biochemical data (Kuroda et al. 2015). This model can be divided into two regions: a transmembrane helix bundle and the periplasmic part connected by a flexible loop. The TM helix trimer superimposes well with the VirB8 (TrwG) triple helix (RMSD = 0.501), preserving the position of the TM segment. Therefore, we used this conformation for the positioning of the helix bundle in DotO proximity. The periplasmic portion consists of the two DotI globular domains modeled with the alpha helices facing the interior of the complex. This orientation is the same as one of the conformations resolved in the DotI crystal (**Figure 5-S15G**) (Kuroda et al. 2015). This

configuration is also adopted by VirB8 (TrwG) for the formation of the arches region in the conjugative T4SS (Macé et al. 2022). We manually positioned the DotI periplasmic part in the linker density and applied C6 symmetry. The flexible linkers were built using MODELLER (Sali and Blundell 1993).

DotB

The DotB structure has been solved and characterized by mutagenesis study. *In vivo* cryo-ET studies revealed its ability to induce a conformational change in DotO upon binding (Sexton, Yeo, and Vogel 2005; Prevost and Waksman 2018; Park et al. 2020). We used the available mutation data (Sexton et al. 2005) to place DotB into the cryo-ET map below the inner DotO hexamer (**Figure 5-S16**).

Bioinformatics and visualization

Membrane simulation. Since the IM in our STA is bend, we used PPM3.0 web server (Lomize, Todd, and Pogozheva 2022) to investigate proteins orientation. This server allows protein-induced membrane deformations using membrane systems reflecting the properties of gram-negative inner membrane.

Protein sequence alignment was performed using Clustal Omega (Sievers et al. 2011) and visualized with Jalview (Waterhouse et al. 2009). Data were visualized using GraphPad Prism, IMOD (Kremer, Mastronarde, and McIntosh 1996), and ChimeraX (Goddard et al. 2018).

5.5 Acknowledgements

Electron microscopy was performed in the Beckman Institute Resource Center for Transmission Electron Microscopy at Caltech and Pacific Northwest Center for Cryo-EM (PNCC). We thank the Liu lab for sharing data and expertise in using I3 to resolve the DotO hexamer of dimers. Y.L. was supported by a Jane Coffin Childs Memorial Fund for Medical Research post-doctoral fellowship. This work was supported by the National Institutes of Health (grant R01-AI127401 to G.J.J.).

5.6 Supplementary Data

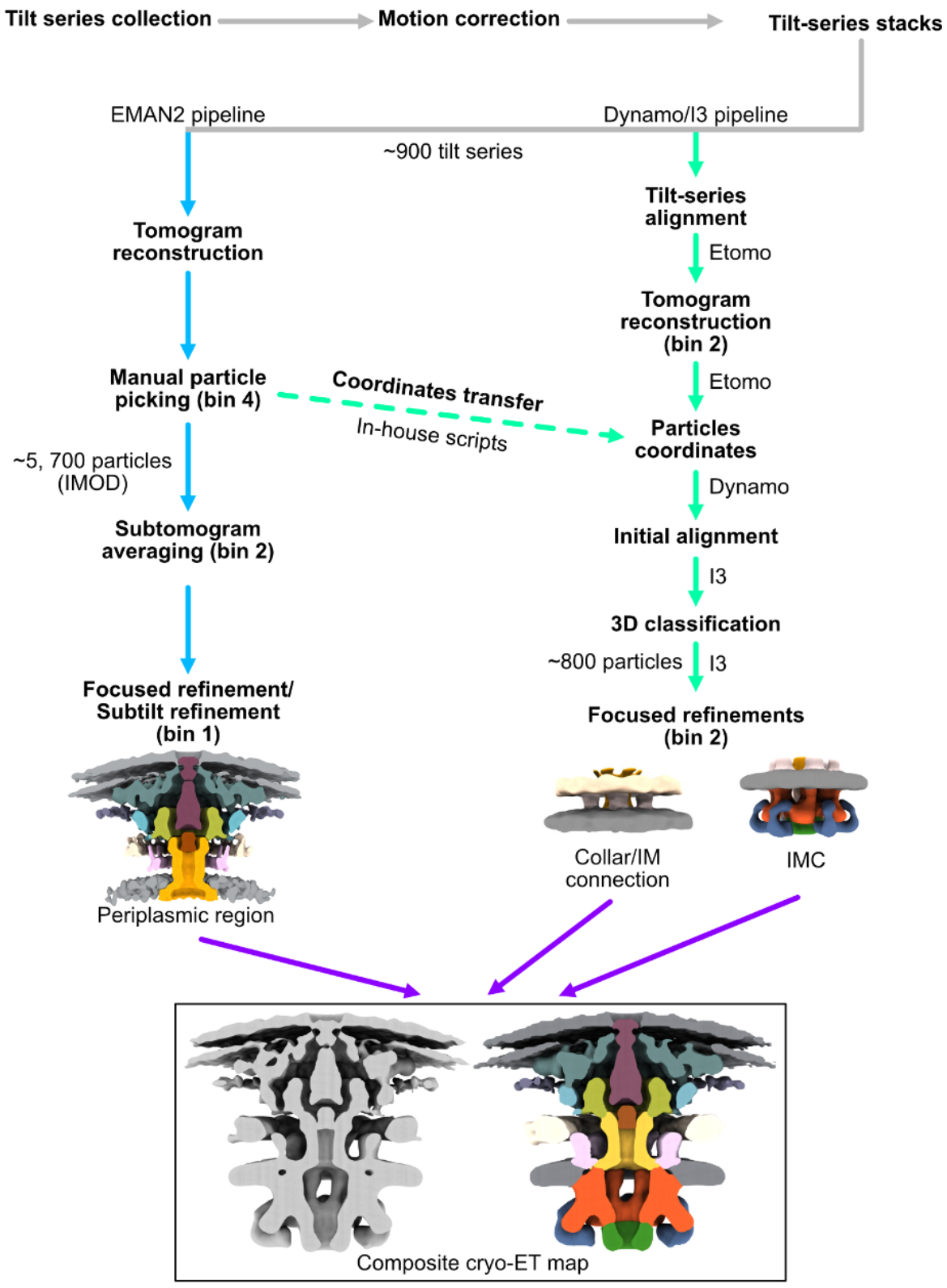


Figure 5-S1. CryoET data processing. Simplified schematic of the data processing workflow highlighting crucial steps in the pipeline.

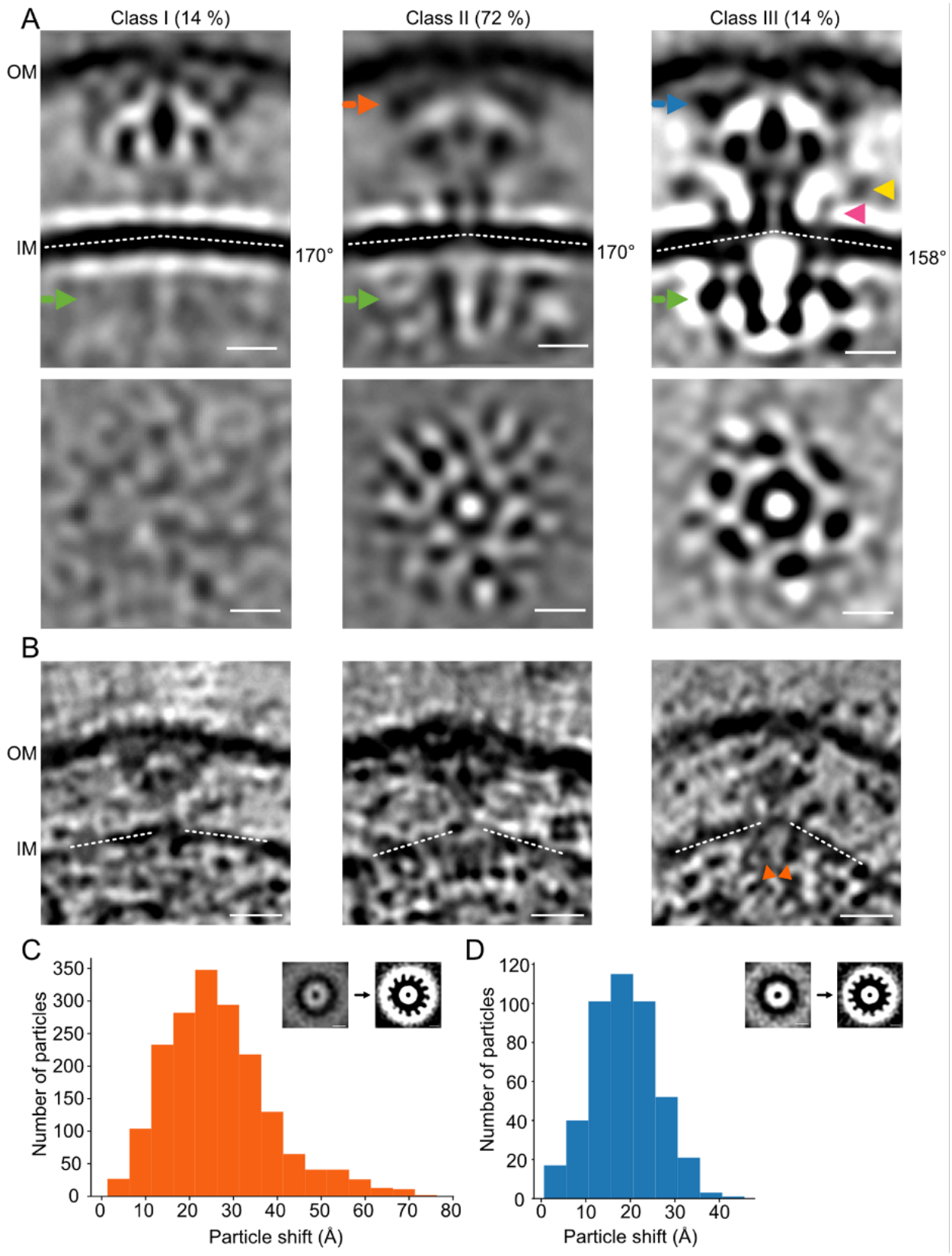


Figure 5-S2. Compositional Heterogeneity of the inner membrane complex (IMC) *in situ*. (**A**) 3D class averages of T4SS with different compositions of the IMC. Top panel, side view. Bottom

panel, cross-section at the position indicated by the green arrows in the top panel. Percentages indicate the number of particles within each class. OM – outer membrane; IM – inner membrane. Yellow and pink arrows indicate the collar and linker densities, respectively. Scale bars, 10nm. (**B**) Individual T4SS particles from FIB-milled samples witch matching composition. Scale bars, 20 nm. Orange arrows indicate the cytosolic ATPases. (**A**,**B**) IM curvature is marked by white dashed lines. (**C**,**D**) Histograms showing total shifts required to restore 13-fold symmetry in the OMC for particles included in Class II (**C**) and Class III (**D**) after focused refinement on the IMC. Insets show a cross-section at the position indicated by orange (**C**) and blue (**D**) arrows in (**A**), respectively. The left and right subpanels represent OMC structure before and after symmetry recovery. Scale bars, 10 nm.

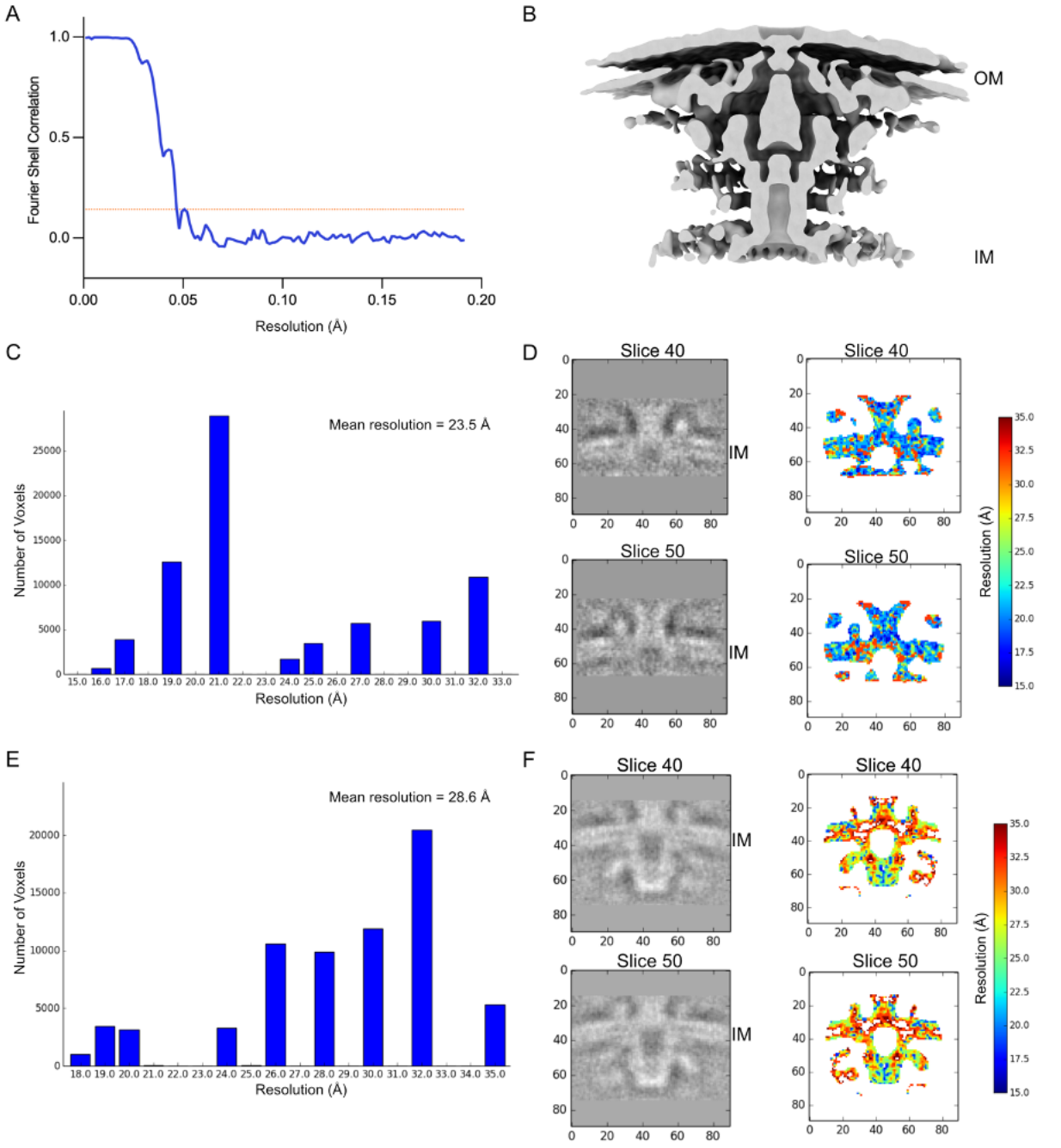


Figure 5-S3. Resolution estimations. (**A**-**B**) Resolution estimation of the EMAN2 subtomogram average of the periplasmic region with C13 symmetry applied. (**A**) Fourier Shell Correlation (FSC) curve with an orange horizontal line indicating a 0.143 threshold. $FSC_{0.143}$ = 21 Å. (**B**) Side view of the corresponding density map. (**C**-**F**) Because gold-standard processing is not implemented in I3, the resolution estimation for maps from I3 was analyzed using ResMap. Resolution was estimated using unfiltered, symmetrized maps. (**C**-**D**) Resolution estimation for the map with local refinement on the collar/IM connection. (**E**-**F**) Resolution estimation for the IMC.

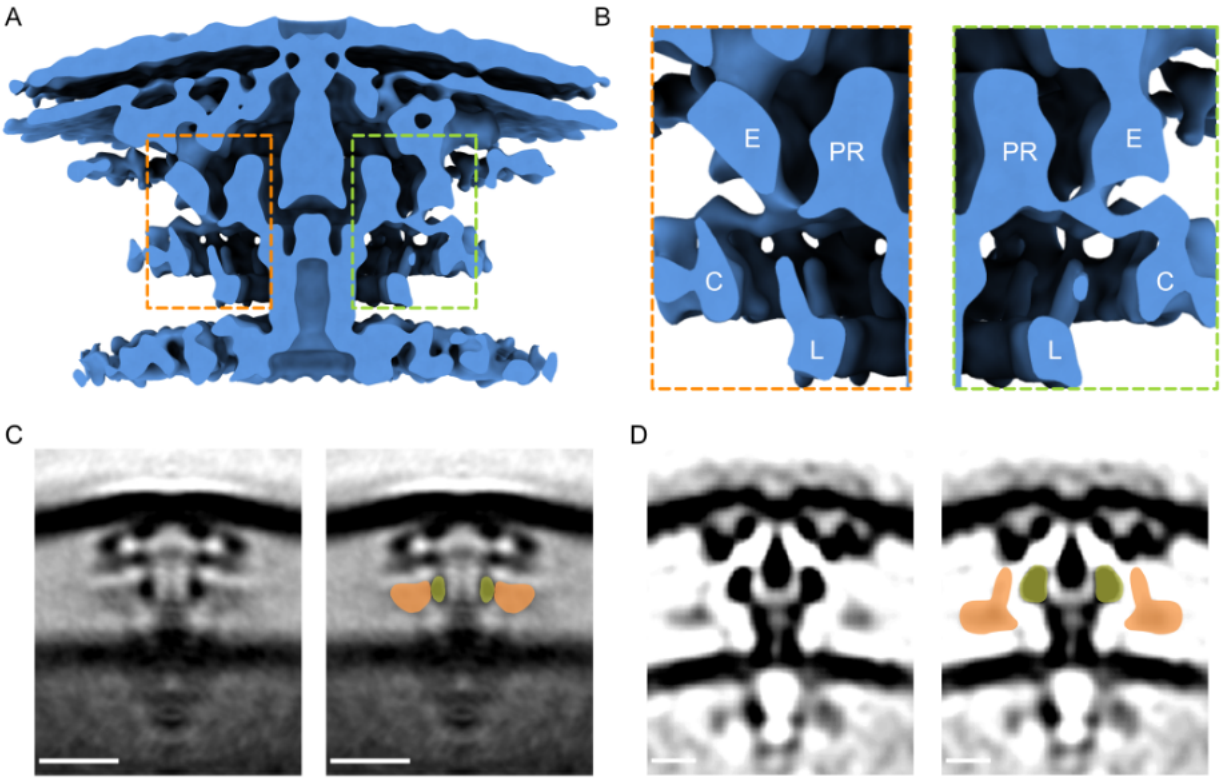


Figure 5-S4. Connectivity of the collar density. (**A**) Side view of the periplasmic complex. (**B**) Enlarged sections as indicated in (**A**) by dashed outlines. E – elbow; PR – periplasmic ring; C – collar; L – linker. (**C**) Central slices through sub-tomogram averages of *H. pylori* cag T4SS (EMD-7474) (Chang et al. 2018). Scale bars, 20 nm. (**D**) Central slices through sub-tomogram averages of *L. pneumophila* Dot/Icm T4SS reproduced from Figure 5-2E in Park et al. (Park et al. 2020). Scale bars, 10 nm. (**C**,**D**) In panels on the right, beige and green marked densities indicate collar and PR, respectively.

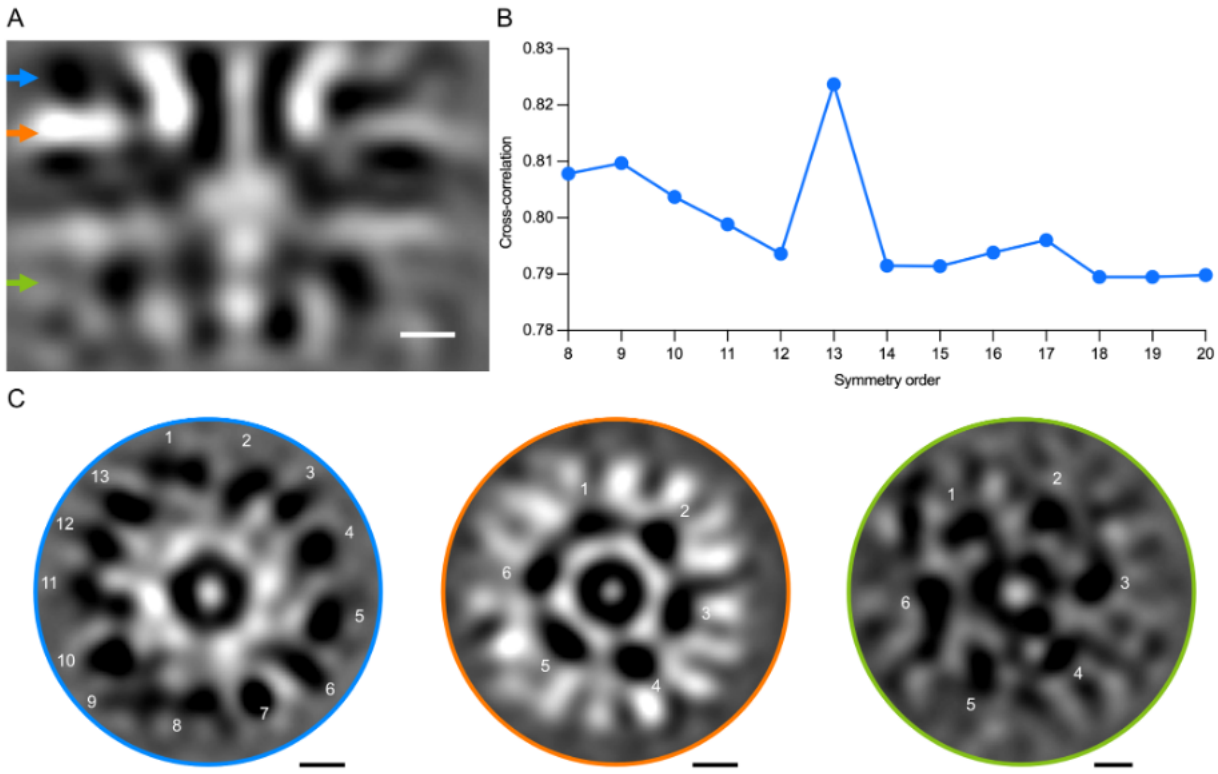


Figure 5-S5. Symmetries of the collar and IMC. (**A**) Central tomographic slice from the focused refinement of the collar/IM region. (**B**) 3D rotational cross-correlation for the collar region marked by a blue arrow in (**A**). (**C**) Cross-sectional views at the positions indicated by color-coded arrows in (**A**). Scale bars, 10 nm.

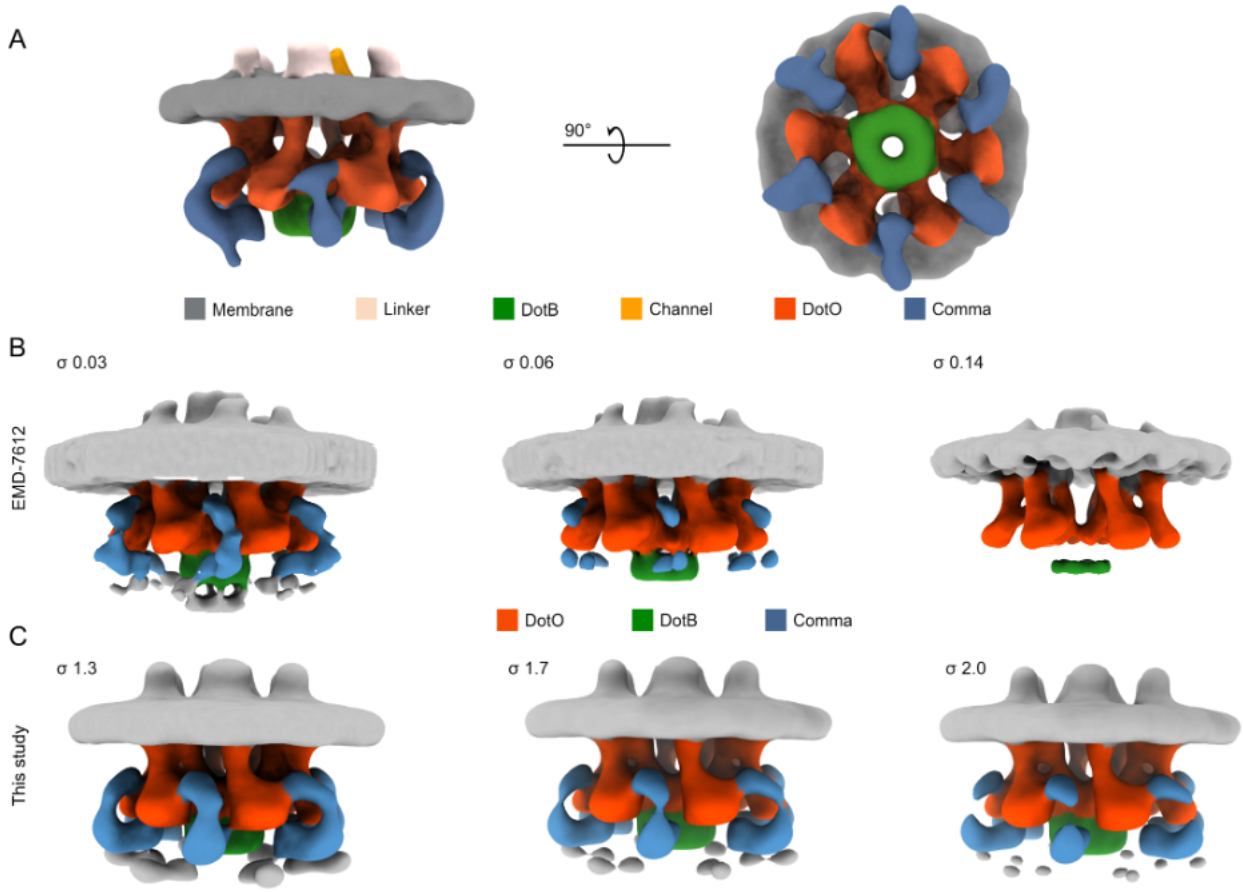


Figure 5-S6. Previously unreported density in the IMC. (**A**) 3D rendering of the unsymmetrized density from focused refinement on the IMC from Class III (Figure 5-S2A). Densities outside the IMC are masked out. (**B**,**C**) Cryo-ET density maps of the IMC displayed at different threshold levels with colored densities corresponding to DotO, DotB, and uncharacterized comma density. (**B**) EMD-7612 (Chetrit et al. 2018). (**C**) This study.

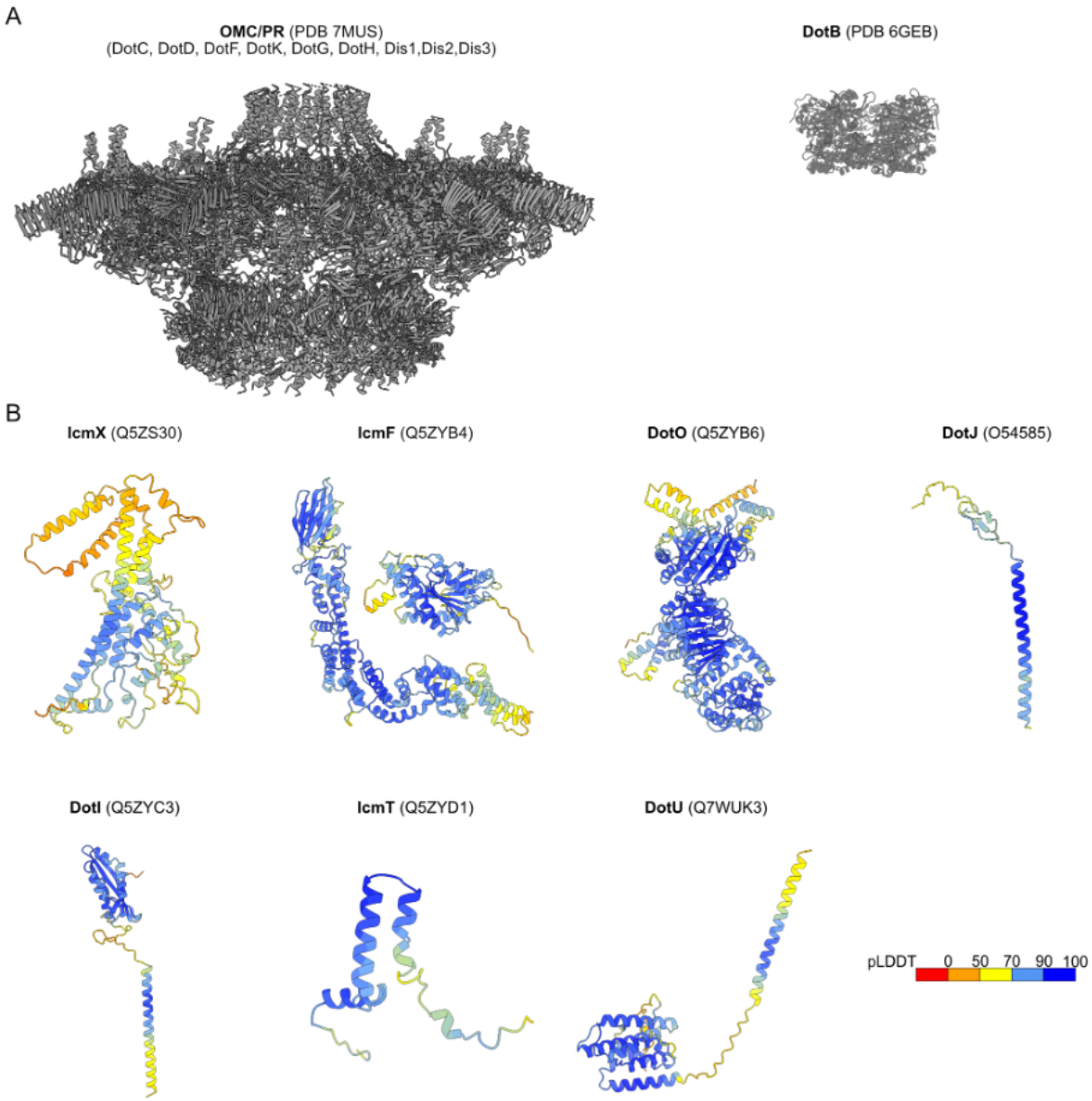


Figure 5-S7. Dot/Icm T4SS components used to build the integrative model. (**A**) Experimentally determined structures. (Left) Single-particle cryo-EM structure (PDB 7MUS) of the periplasmic core complex (Sheedlo et al. 2021) and (right) crystal structure (PDB 6GEB) of DotB (Prevost and Waksman 2018). (**B**) Monomeric AlphaFold structure predictions for Dot/Icm T4SS components. The models are colored by local model confidence estimated with a predicted local distance difference test (pLDDT) (Jumper et al. 2021). UniProtKB accession numbers are indicated in the parenthesis.

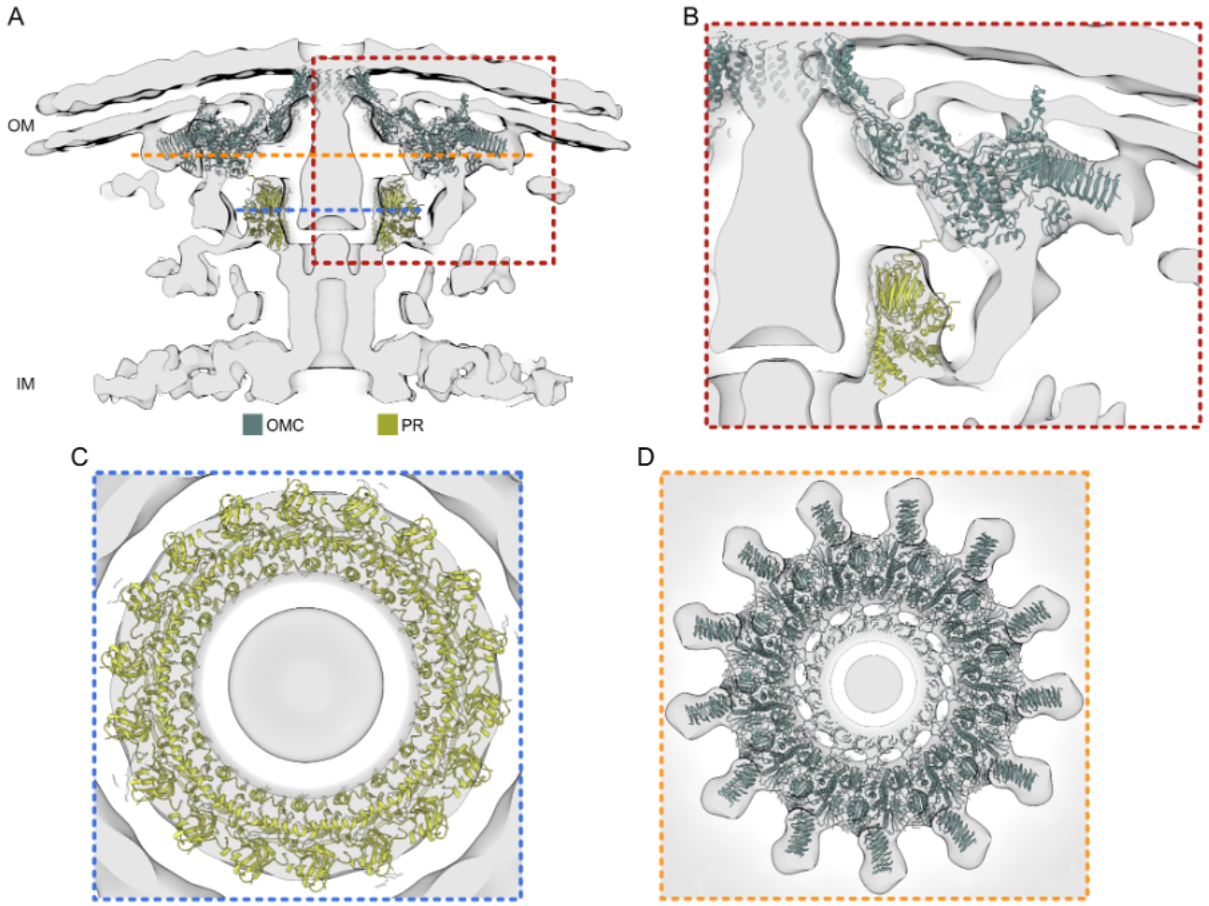


Figure 5-S8. Fitting the atomic models of the OMC and PR into the cryo-ET density map. (**A**) 13-fold symmetrized cryo-ET map (gray) of the T4SS periplasmic region with the atomic models of OMC and PR (PDB 7MUS) (Sheedlo et al. 2021). (**B**) Enlarged section from (**A**) as indicated by the red dashed box. (**C**,**D**) Cross-sections at the positions indicated by blue (**C**) and orange (**D**) dashed lines in panel (**A**).

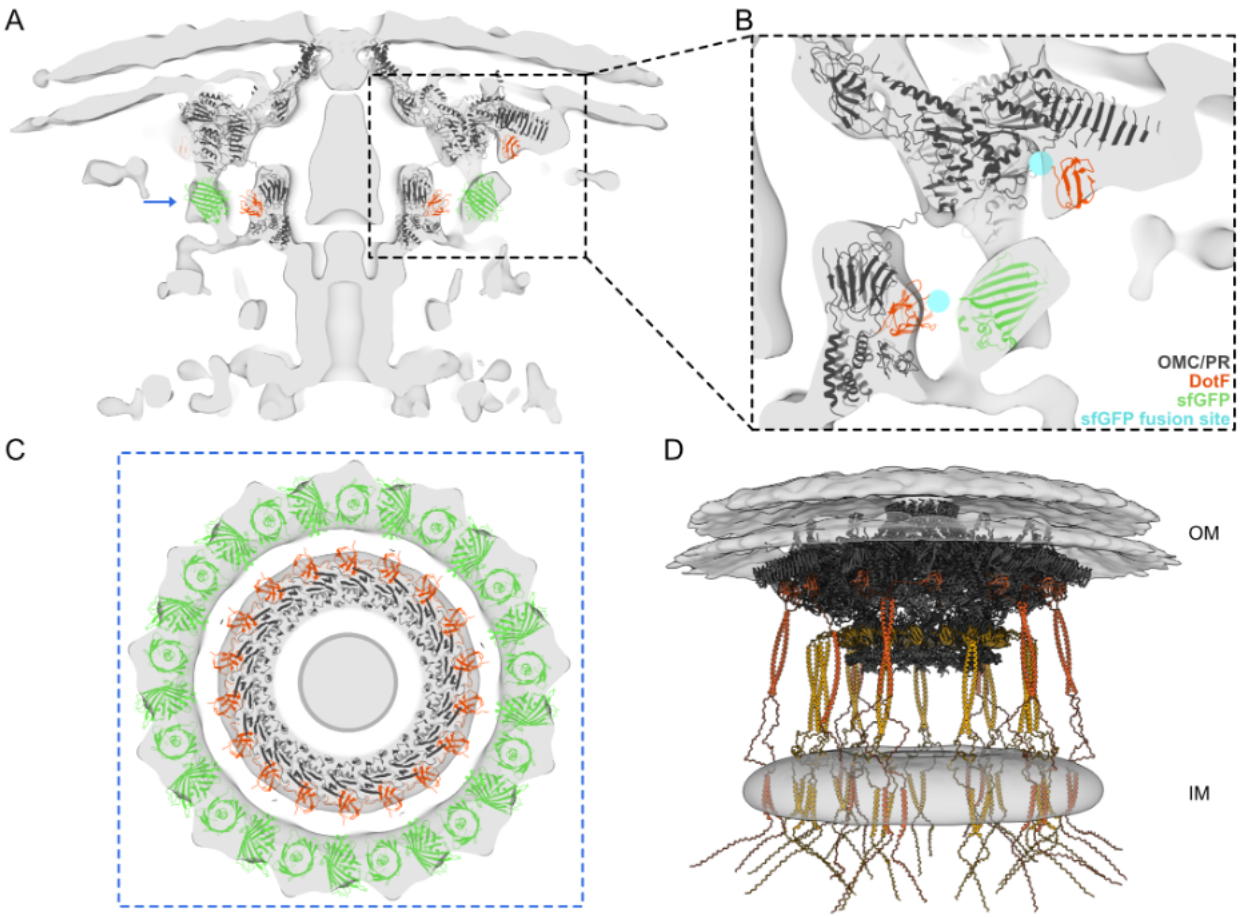


Figure 5-S9. Elbow density attributed to sfGFP. (**A**) Atomic models of the OMC and PR (PDB 7MUS) (Sheedlo et al. 2021) fit into the cryo-ET map. DotF subunits are colored in orange. The atomic model of sfGFP (green) dimer (PDB 5B61) (Choi et al. 2017) fit into the elbow density. (**B**) Inset shows enlarged section of the elbow density. The position of DotF-GFP fusion in our construct is indicated by blue spheres. (**C**) Cross-section the elbow density ring at the positions indicated by a blue arrow in panel **A**, with 26-copies of sfGFP. (D) Predicted localization of the coiled coil and transmembrane N-terminal domains of DotF. Subunits of the DotF portions localized in the OMC and PR are colored in orange and yellow, respectively.

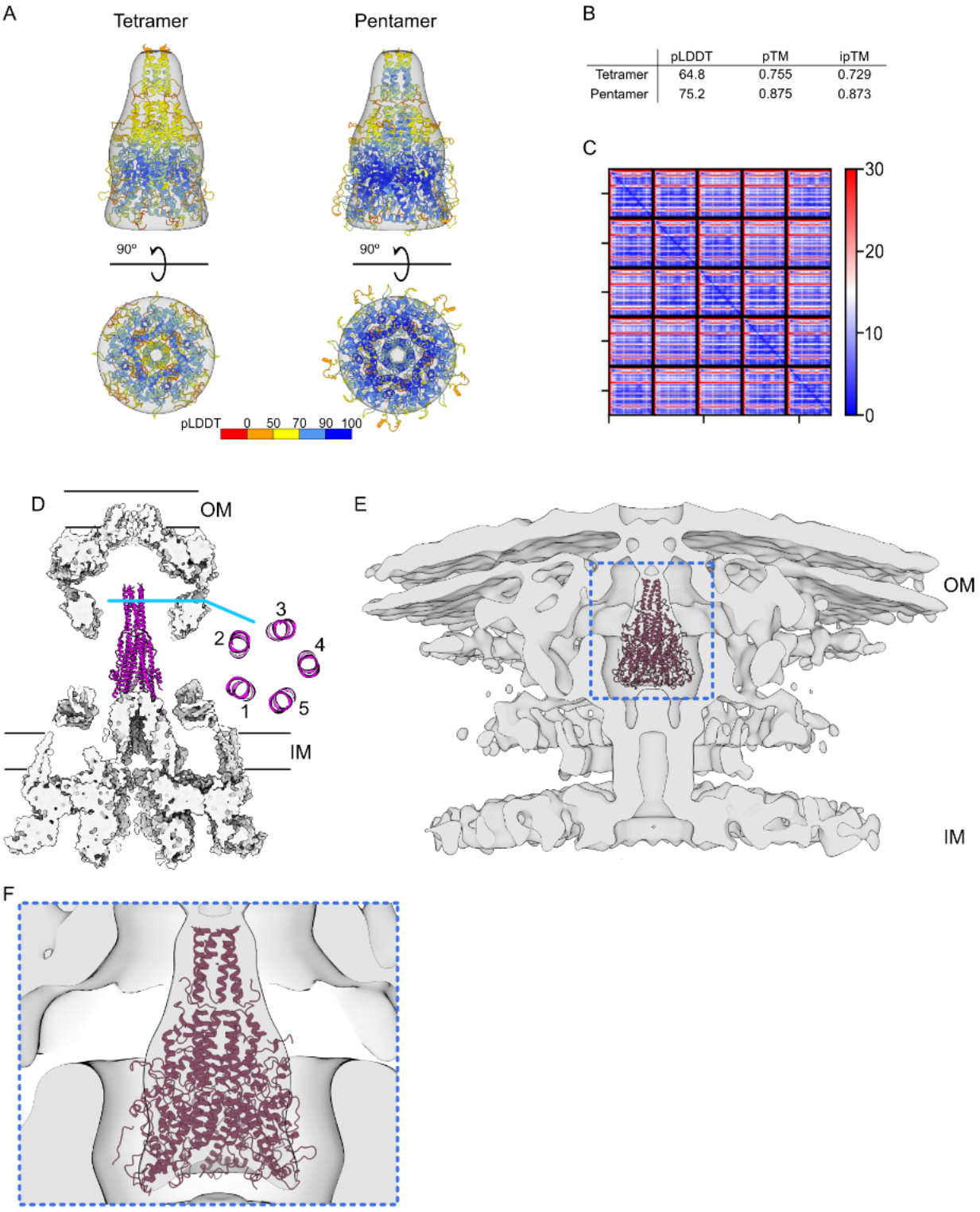


Figure 5-S10. Modeling IcmX into the plug density. (**A**) AlphaFold/ColabFold predictions for the tetramer and pentamer of IcmX colored by pLDDT fit into the segmented plug density. (**B**) Modeling scores as returned by ColabFold for the highest-ranked models. pLDDT – predicted

local distance difference test, representing local accuracy; pTM – predicted TM-score, corresponding to overall topological accuracy; ipTM – interface pTM score. (**C**) Heatmap represents the inter-chain accuracy of the model for the highest-rank pentameric model. Each pixel in the heatmap indicates alignment error in Angstroms for a pair of two residues. The metric indicates the positional uncertainty at residue x if the predicted and actual structures are aligned on residue y (Varadi et al. 2022). Blue represents low error. (**D**) Atomic model of the conjugative T4SS (PBD 7O3J, 7O3T, 7O3V, 7O41) (Macé et al. 2022) with the VirB5 pentamer forming a stalk with structural similarity to IcmX. (**E**) Rigid-body fitting of the pentameric model of IcmX into the cryo-ET map. (**F**) Enlarged view as indicated by the blue dashed outline in panel **E**.

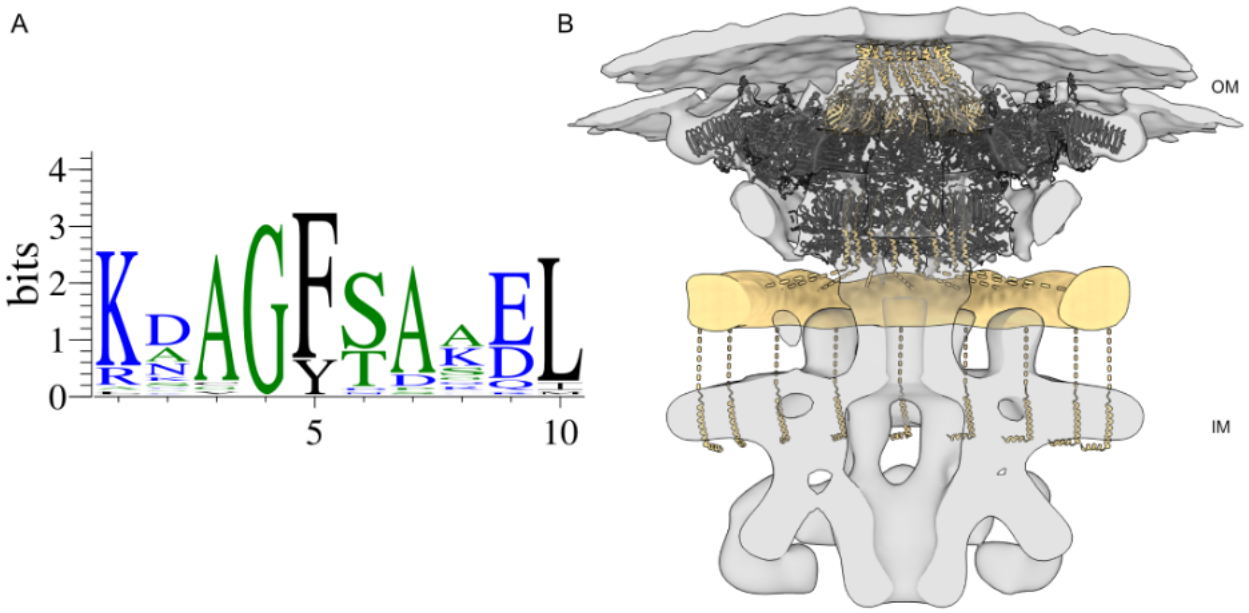


Figure 5-S11. DotG as potential collar forming protein. (**A**) Weblogo plot showing consensus sequence of the 10 amino-acid repeating sequence in DotG. (**B**) Schematic representation showing how DotG could continue from the PR to form the collar and eventually anchor in the IM. Atomic model of the OMC and PR (PDB 7MUS) (Sheedlo et al. 2021) in black. DotG highlighted in beige.

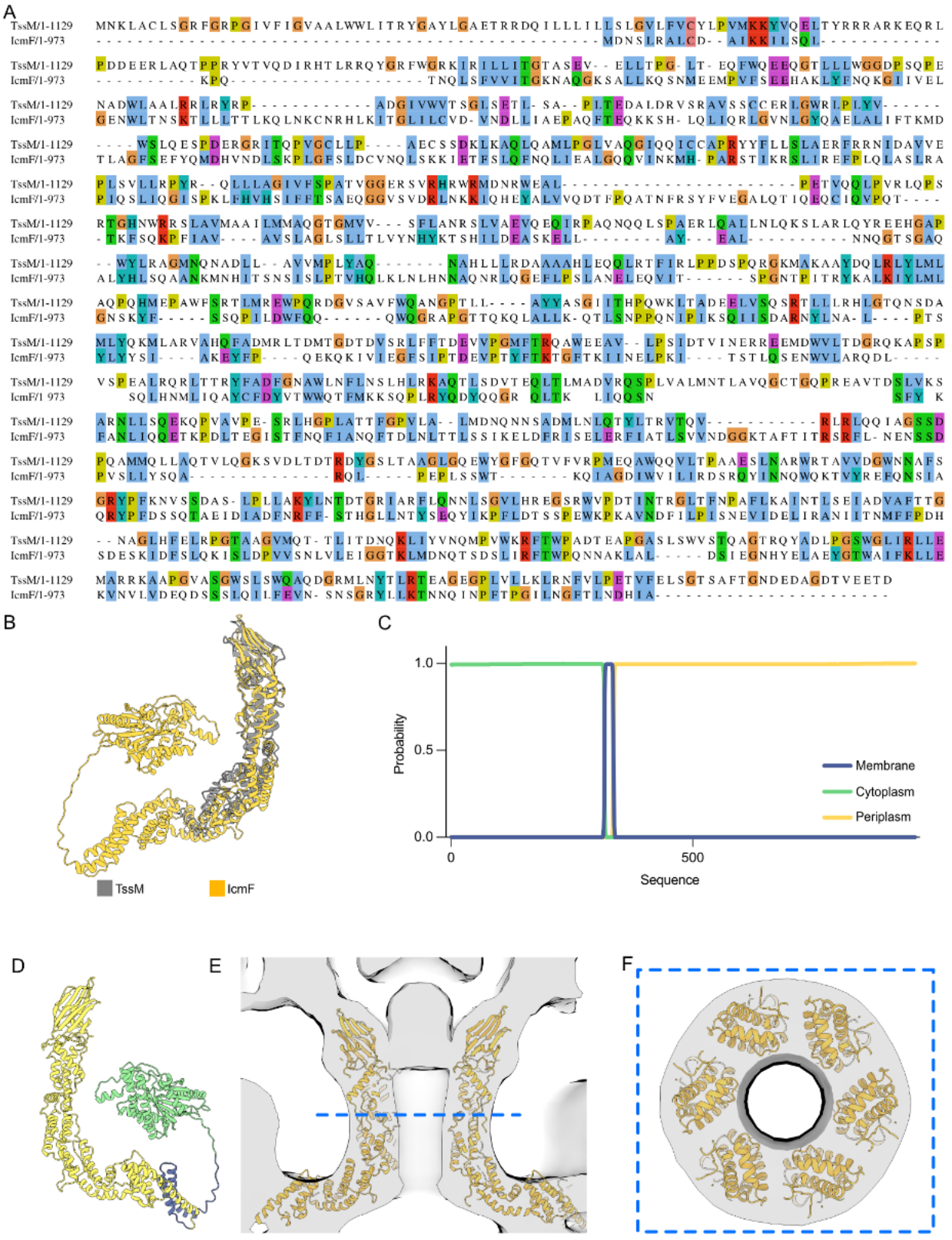


Figure 5-S12. Modeling IcmF into the cylinder density. (**A**) Protein sequence alignment between TssM (T6SS) and IcmF. (**B**) Structure alignment between the structure of one of the TssM

monomers that form the T6SS core complex channel (PDB 6IXH)(Yin, Yan, and Li 2019) and a monomer of IcmF as predicted by AlphaFold. (**C**) IcmF topology prediction by DeepTMHMM (Hallgren et al. 2022). (**D**) IcmF model colored according to the predicted topology. (**E**-**F**) Rigid-body fitting of the IcmF hexamer into the cylinder density of the cryo-ET map. (**F**) Cross-section at the position indicated by the blue dashed line in panel **E**.

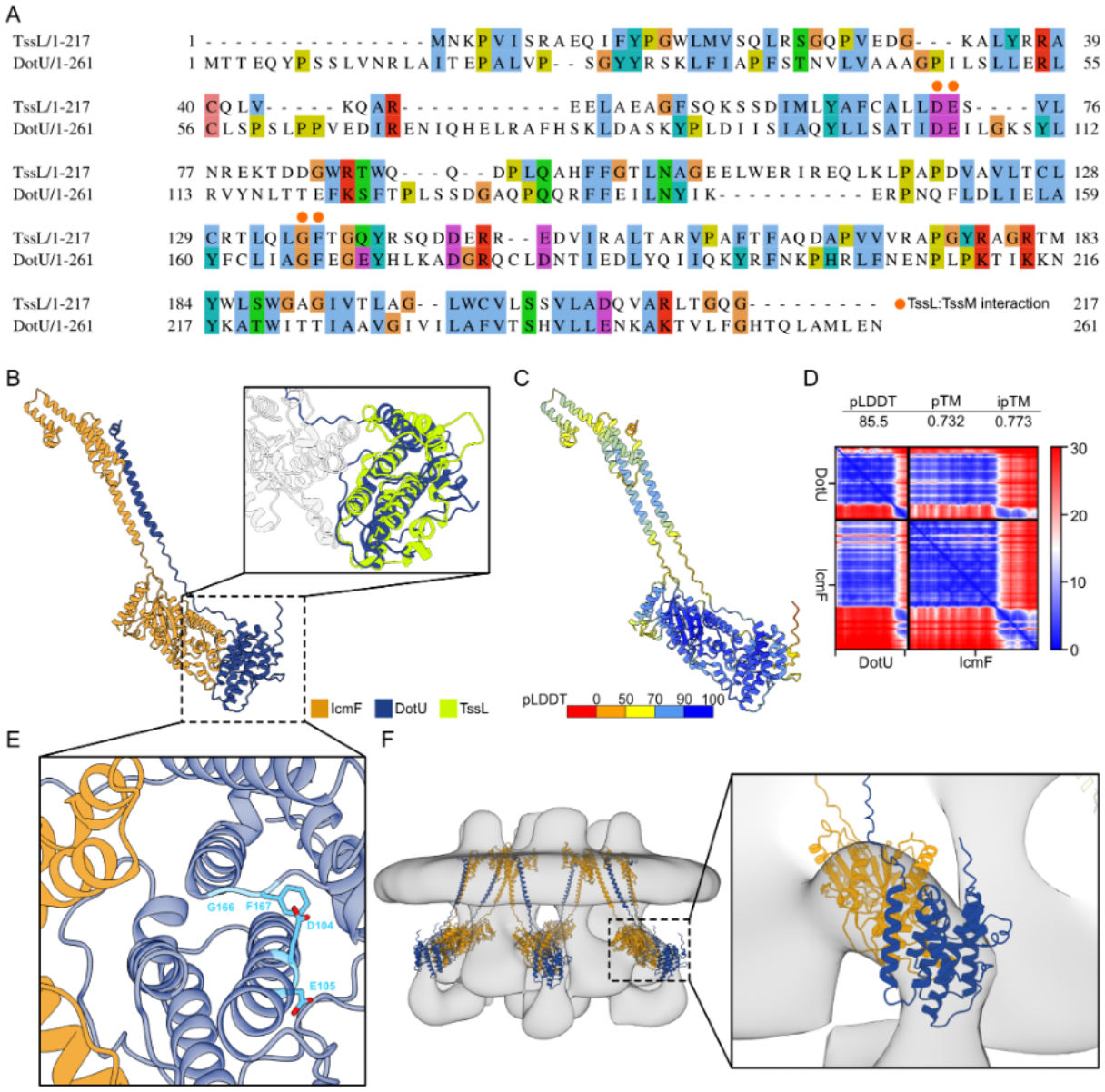


Figure 5-S13. Modeling the IcmF$_{cyto}$:DotU subcomplex. (A) Protein sequence alignment between TssL (T6SS) and DotU. Orange dots indicate the location of the residues important for TssM:TssL interaction in T6SS (Zoued et al. 2016). These residues are conserved in DotU. (**B**) AlphaFold/ColabFold prediction of the IcmF$_{cyto}$:DotU heterodimer. The inset shows the structural similarity between the DotU model and the crystal structure of TssL (PDB 3U66) (Durand et al. 2012). (**C**) IcmF$_{cyto}$:DotU model colored by pLDDT. (**D**) Modeling scores as returned by ColabFold for the highest-ranked model. For a full description see Figure 5-S10. (**E**) Location of conserved residues in DotU identified previously as important for the TssM:TssL interaction (Zoued et al. 2016). Residues are seen here to be located close to the IcmF$_{cyto}$:DotU interaction interface. Our interpretation is that point mutations disrupt tertiary structure and therefore the IcmF$_{cyto}$:DotU interaction. (**F**) Docking of the IcmF$_{cyto}$:DotU subcomplexes into the top of the comma density in the cryo-ET map.

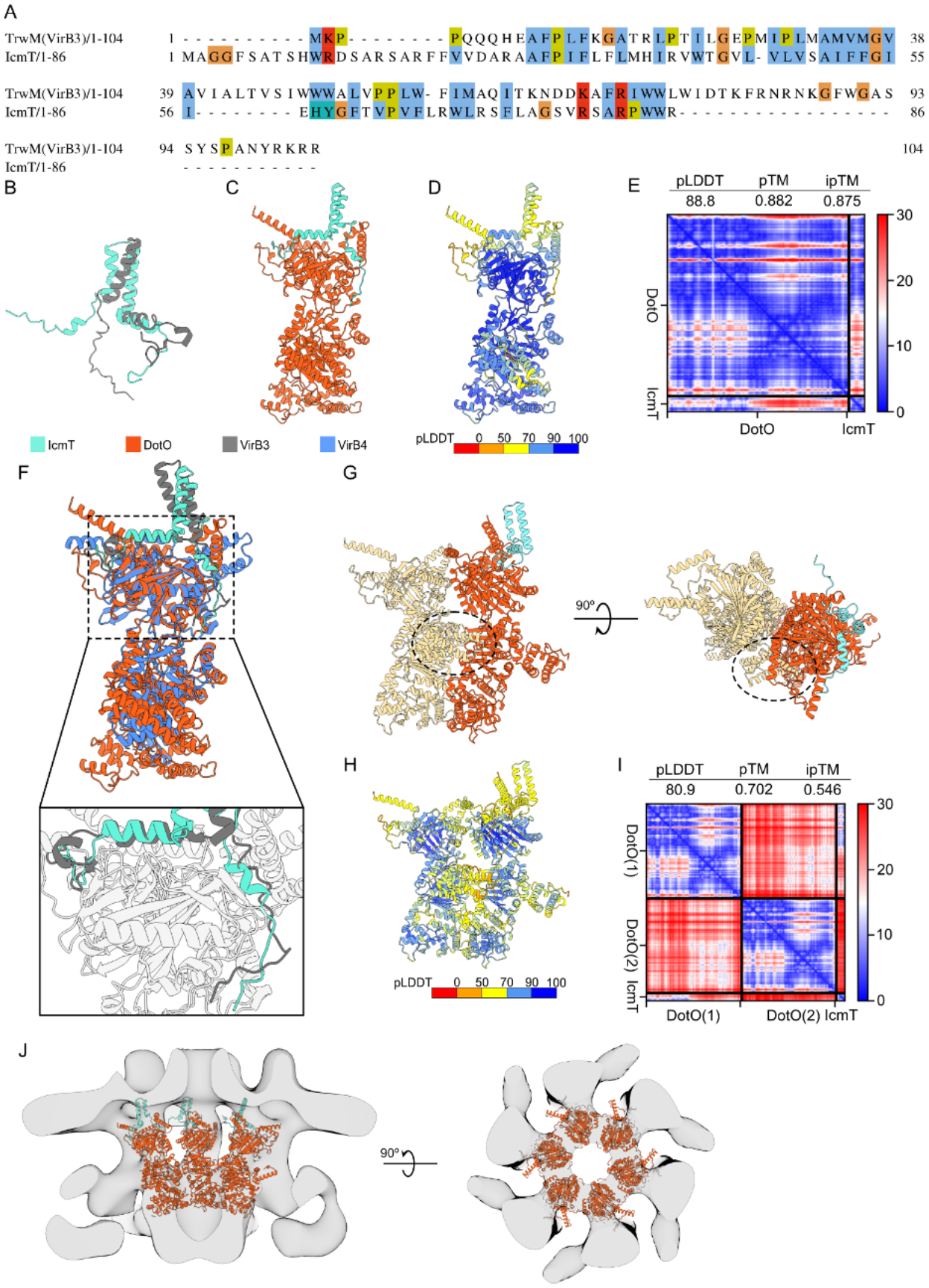

A
TrwM(VirB3)/1-104 1 - - - - - - - - - - MKP- - - - - - - - PQQQHEAFPLFKGATRLPTILGEPMIPLMAMVMGV 38
IcmT/1-86 1 MAGGFSATSHWRDSARSARFFVVDARAAFPIFLFLMHIRVWTGVL- VLVSAIFFGI 55
TrwM(VirB3)/1-104 39 AVIALTVSIWWWALVPPLW- FIMAQITKNDDKAFRIWWLWIDTKFRNRNKGFWGAS 93
IcmT/1-86 56 I- - - - - - - - EHYGFTVPVFLRWLRSFLAGSVRSARPWWR- - - - - - - - - - - - - - - - 86
TrwM(VirB3)/1-104 94 SYSPANYRKRR 104
IcmT/1-86 - - - - - - - - - - -
B
C
D
E
pLDDT pTM ipTM
88.8 0.882 0.875
DotO
IcmT
30
20
10
0
IcmT DotO VirB3 VirB4
pLDDT 0 50 70 90 100
F
G
90°
H
pLDDT 0 50 70 90 100
I
pLDDT pTM ipTM
80.9 0.702 0.546
DotO(1)
DotO(2) IcmT
J
90°

Figure 5-S14. Modeling the IcmT:DotO ATPase subcomplexes. (**A**) Protein sequence alignment between VirB3 and IcmT. (**B**) Structure alignment between the experimentally determined structure of VirB3 (PDB 7O41) (Macé et al. 2022) and AlphaFold prediction for IcmT. (**C**-**D**) Predicted structure of IcmT:DotO subcomplex from AlphaFold/ColabFold. (**D**) IcmT:DotO model colored by pLDDT. (**E**) Modeling scores as returned by ColabFold for the highest-ranked model. For a full description see Figure 5-S10. (**F**) Overlay of the IcmT:DotO predicted model with the experimental structure of homologous VirB3/VirB4 (PDB 7O41) (Macé et al. 2022). The inset shows the conserved binding site of IcmT and VirB3. (**G-I**) AlphaFold/ColabFold model for DotO dimer with IcmT. (**G**) The C-terminal portion responsible for inter-subunit interaction between DotO is indicated by a dashed circle. (**H**) Predicted model colored by pLDDT. (**I**) Modeling scores as returned by ColabFold for the highest-ranked model. For a full description see Figure 5-S10. (**J**) IcmT:DotO inner hexamer fit into the cryo-ET map (gray).

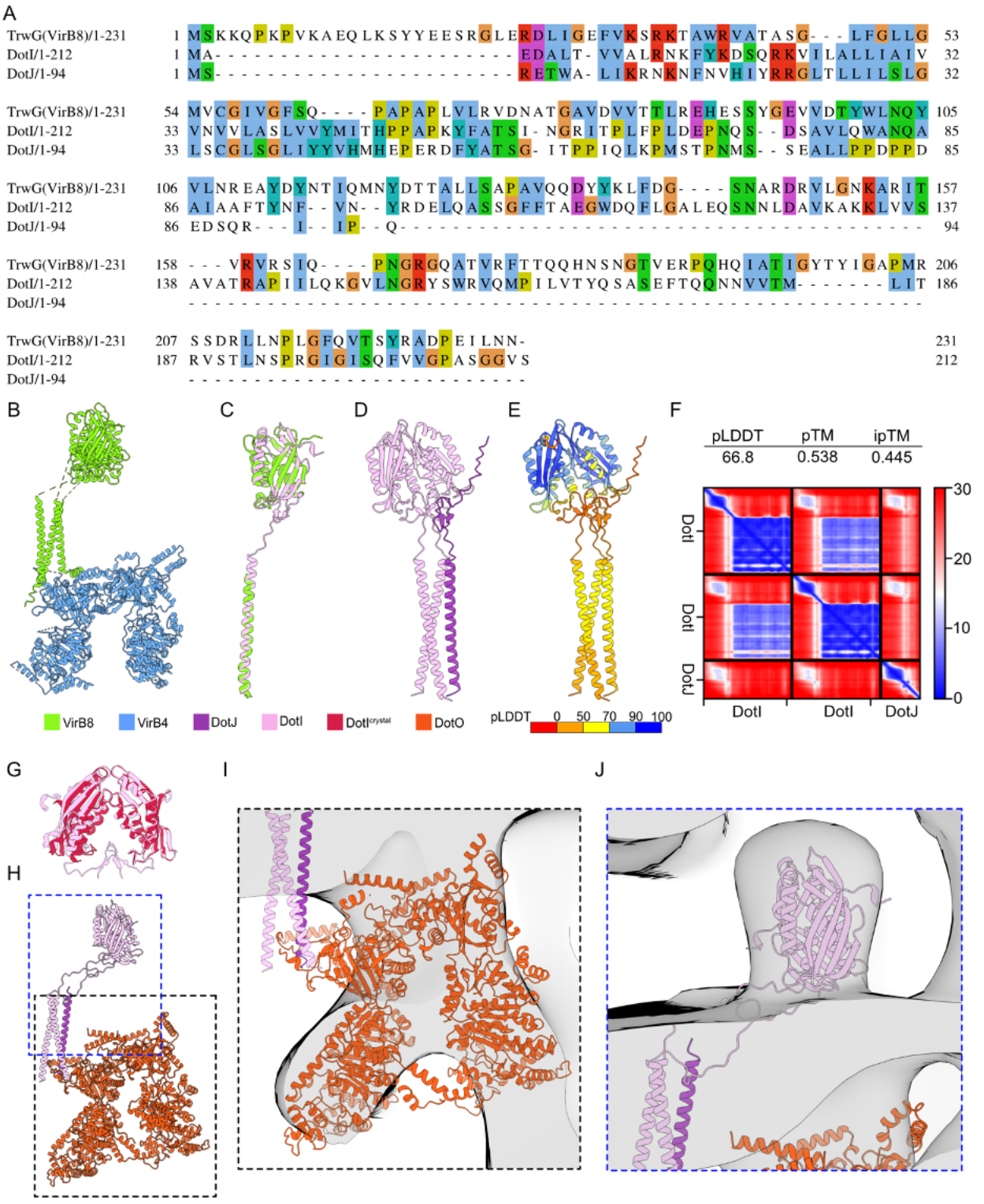


Figure 5-S15. Modeling the DotI:DotJ:DotO subcomplex. (**A**) Protein sequence alignment between VirB8, DotI, and DotJ. (**B**) Structure of the VirB8:VirB4 subcomplex (PDB 7O41) (Macé et al. 2022). (**C**) Structure alignment between the experimentally determined structure of VirB8 (PDB 7O41) (Macé et al. 2022) and AlphaFold prediction for DotI. (**D**) 2:1 DotI:DotJ subcomplex

structure predicted by ColabFold/AlphaFold. (**E**) DotI:DotJ model colored by pLDDT. (**F**) Modeling scores as returned by ColabFold for the highest-ranked model. (**G**) Alignment of the periplasmic portion of the DotI dimers comparing the AlphaFold prediction and a crystal structure (PDB 3WZ5) (Kuroda et al. 2015). (**H**) Subcomplex of DotO dimer with DotI:DotJ trimer modeled based on the VirB8:VirB4 structure. (**I**) The cytosolic portion of the DotI:DotJ:DotO subcomplex fit into the cryo-ET map. (**J**) DotI periplasmic portion fit into the cryo-ET map.

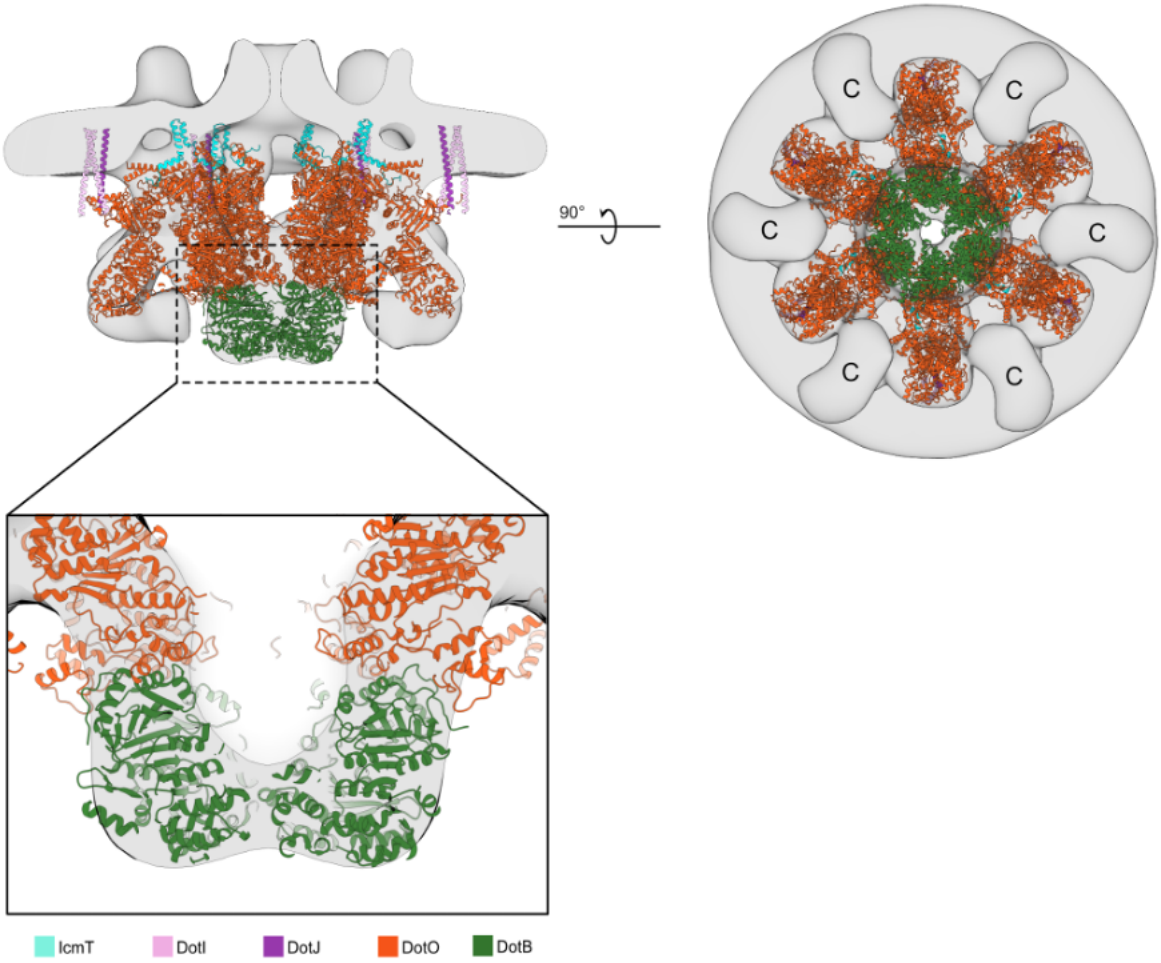


Figure 5-S16. Placement of DotB. DotB crystal structure (PDB 6GEB) (Prevost and Waksman 2018) docked into the density below the inner DotO hexamer in the cryo-ET map. The fully assembled ATPases (DotO:DotB) are anchored to the IM through IcmT (inner DotO) and DotI:DotJ (outer DotO). C – comma density.

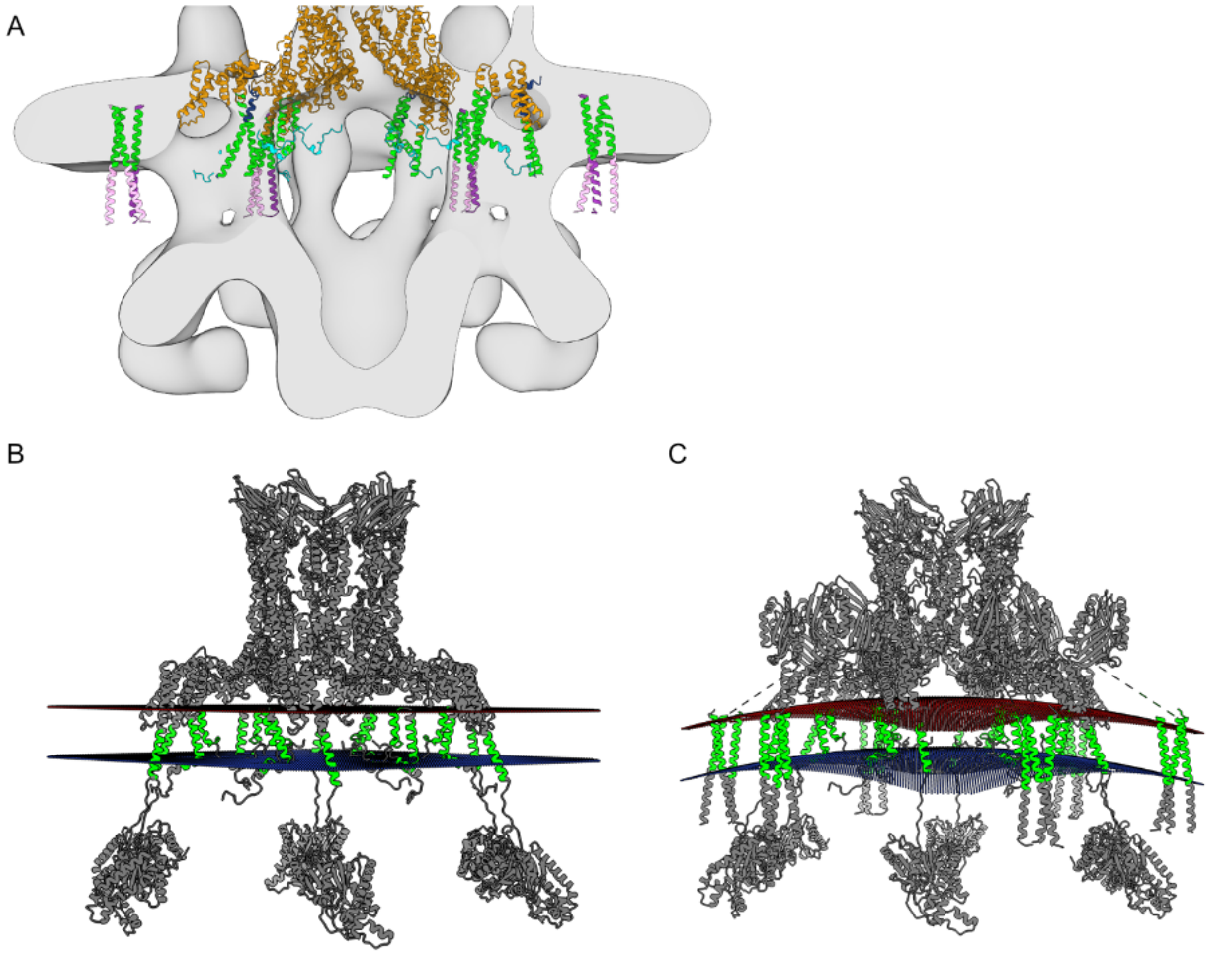


Figure 5-S17. Localization of the transmembrane domains in the IM. (**A**) Model of the IMC fit into the cryo-ET map. Cytosolic portions of the proteins are not shown. (**B**,**C**) Simulated membrane topology using PP3 package for the model with IcmT:IcmF complex (**B**) and IcmF:IcmT:DotI:DotJ complex (**C**). Predicted transmembrane regions are highlighted in green.

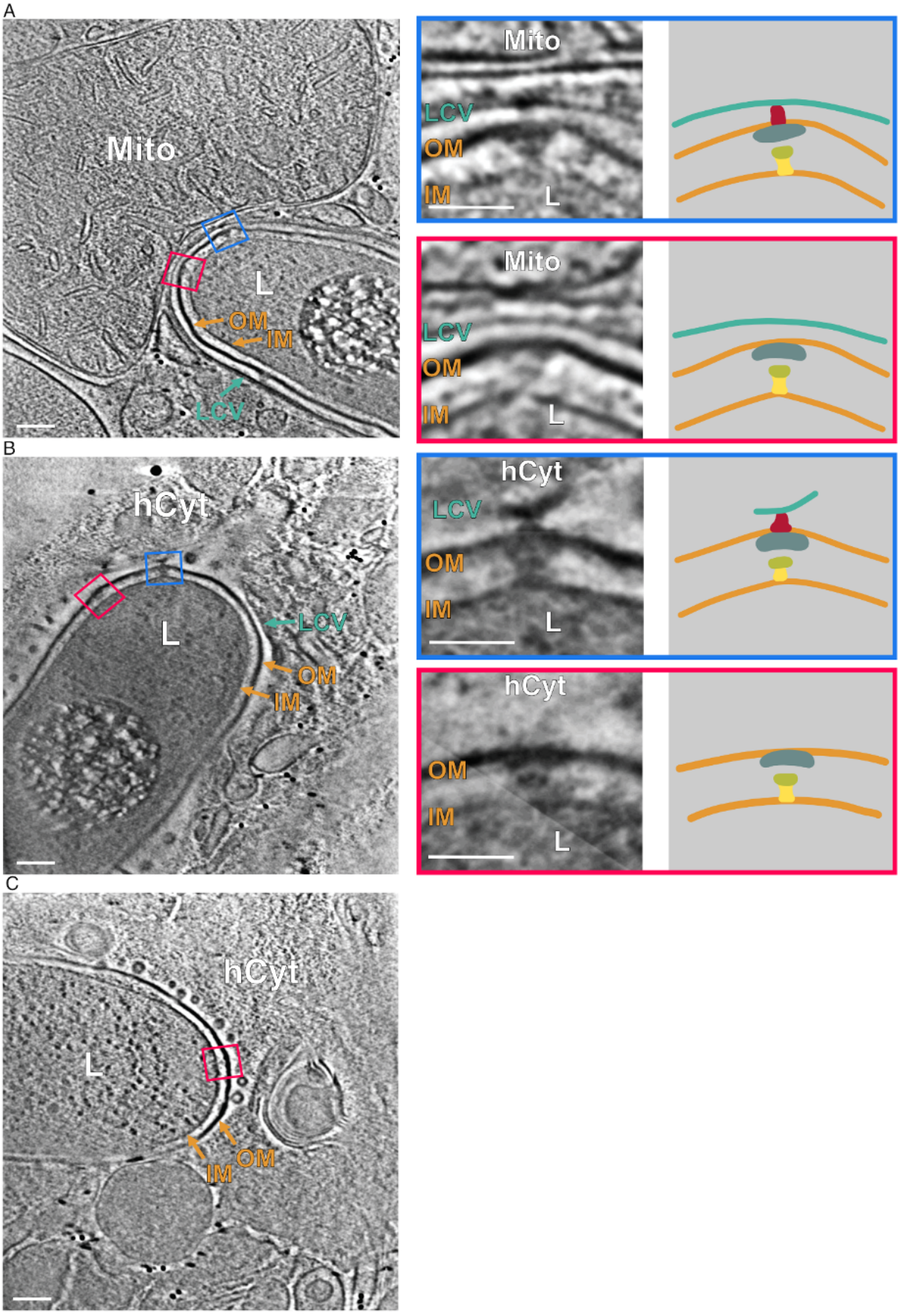

A
Mito
L
OM
IM
LCV
B
hCyt
L
LCV
OM
IM
C
hCyt
L
IM
OM
Mito
LCV
OM
IM
L
Mito
LCV
OM
IM
L
hCyt
LCV
OM
IM
L
hCyt
OM
IM
L

Figure 5-S18. Additional examples of *L. pneumophila* inside host cells during early infection. **(A-C)** Representative tomographic slices of intracellular Legionella imaged at 1 h post-infection. OM (orange) – *L. pneumophila* outer membrane; IM (orange) – *L. pneumophila* inner membrane; LCV (cyan) – *L. pneumophila* containing vacuoles. L (orange) – *L. pneumophila* cells; hER (brown) – endoplasmic reticulum of U937 cells. Particles tethered to the host membrane are enclosed by blue frames; particles without a tethering site are indicated by red frames. The panels on the right depict individual particles making the OMC, PR, and cylinder schematically. Scale bars, 100 nm for main panels and 50 nm for the insets.

Table 5-T1. Primers used for *L. pneumophila* strains preparation.

Primer name	Sequence (5'-3')
JVP2973	CGCGGATCCTGAAGGAGGAATTCGCCAATGATGGCAGAGCACGATCAAAATAATG
JVP2990	GGAGCGGCCGCTCATCATTTGTACAGTTCATCCATACC
JVP2992	GGAGCGGCCGCGAAGATAGTTGAGGCGTATATGGCTG
JVP2993	CCCGAGCTCCTCTCTATTAATTAGTCTTATATATTGTTCC

CONCLUSIONS

This thesis explores the structure and its functional implications for two large bacterial macromolecular complexes, GVs and T4SS. Despite being vastly different, they both constitute a significant challenge for any structural biology method. GVs are perhaps the largest macromolecular assemblies known, and in some cases, can reach several micrometers in length. In addition to being large, GVs' dimensions are highly irregular, adding an additional layer of difficulty for any method that requires averaging. Another consequence of GV's size and mechanical properties is their tendency to flatten in the ice, which breaks helical symmetry. Finally, GVs are highly sensitive to radiation damage; they rapidly shrink when exposed to an electron beam, reducing the maximum electron dose that can be used for imaging. This leads to noisy images which are more difficult to process and reconstruct to obtain high-quality tomograms.

L. pneumophila T4SS, on the other hand, presents a very different set of challenges. There is no reported successful purification of the intact Dot/Icm T4SS, requiring it to be investigated in intact cells. This results in a thick sample (250-300 nm) with only a few particles per tomogram (~3-5), calling for the collection of a large dataset to obtain a sufficient number of particles that can be averaged and classified. Moreover, T4SS is a highly dynamic system that can exist in several different configurations and exhibits symmetry mismatches between separate components, further complicating data processing.

These challenges can be overcome using integrative modeling guided by cryo-ET maps. The integrative model revealed that GvpA subunits polymerize in an amyloid-like fashion to form the

GV shell, which is further reinforced by the GvpC rod that creates a helical spiral around the GV cylinder. This model provided answers to fundamental questions on how GVs are formed explaining their mechanical resilience, and how gas diffuses in and out of the GVs. Additionally, this work elucidated the evolutionary relationships between different types of GVs, and how they might regulate their mechanical properties between different species. Importantly, cryo-ET provided insight into the biogenesis of GVs, revealing that they might assemble from a single point at the center of the cylinder by adding individual subunits of GvpA in opposite directions. These findings provide valuable information for further development of GVs as acoustic reporter genes. However, this research focused mainly on fully assembled GVs. Future work should focus on understanding how the assembly process is initiated and how GVs grow as bicones and eventually transition to cylinders. GV gene clusters encode proteins other than GvpA and GvpC with unknown functions, usually referred to as "accessory proteins." It is crucial to understand the role of these accessory proteins to manipulate them and develop new variants of GVs with desired functions for their biotechnological applications. Finally, to fully embrace GVs' potential, it will be necessary to study their interactions with other cellular components and the consequences of their heterologous overexpression on cellular machinery and fitness.

Similarly, a nearly complete model of the Dot/Icm T4SS is an important milestone in the field that builds on several decades of functional, biochemical, and structural work dedicated to understanding this molecular machine. For the first time, we have a clear picture of the localization of the key protein components of the system and their interactions. This work provides important clues on how the system might work as a complex network, communicating signals from outside in and inside out to secrete over 300 different proteins and hijack the host's biology. The integrative

model of the T4SS will serve as a framework to build and test hypotheses on how substrates are translocated to the host. For example in the cytoplasmic portion of the system, there are several enzymes with NTPase activity that must be responsible for substrate delivery, unfolding, and pushing it through the system, but the exact function of each of these proteins is unclear. Another mystery is the size of the complex. If the T4SS is translocating unfolded cargo, why is it so complex and large? Finally, the integrative model presented here is based on an inactive system, but to fully understand its mechanism, it will be necessary to solve the structure of the system in an actively secreting form.

Both GVs and T4SS are examples of large and complex biological assemblies that are very difficult to study using traditional structural biology methods. Integrative modeling has proven to be successful in both cases and has helped to resolve decades-long mysteries about both systems. However, this methodology is still young and will continue to grow, especially with current advances in protein structure prediction and the rising interest in cryo-ET. This necessitates the development of standardized methods for validating integrative models to further advance structural biology and move away from studying isolated proteins without their physiological context.

BIBLIOGRAPHY

Achenbach, Jan. 2012. *Wave Propagation in Elastic Solids*. Elsevier. https://play.google.com/store/books/details?id=XyQ6ibRvst4C.

Adams, Paul D., Pavel V. Afonine, Gábor Bunkóczi, Vincent B. Chen, Ian W. Davis, Nathaniel Echols, Jeffrey J. Headd, et al. 2010. "PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution." *Acta Crystallographica. Section D, Biological Crystallography* 66 (Pt 2): 213–21. https://doi.org/10.1107/S0907444909052925.

Agulleiro, Jose-Ignacio, and Jose-Jesus Fernandez. 2015. "Tomo3D 2.0--Exploitation of Advanced Vector Extensions (AVX) for 3D Reconstruction." *Journal of Structural Biology* 189 (2): 147–52. https://doi.org/10.1016/j.jsb.2014.11.009.

Alvarez-Martinez, Cristina E., and Peter J. Christie. 2009. "Biological Diversity of Prokaryotic Type IV Secretion Systems." *Microbiology and Molecular Biology Reviews: MMBR* 73 (4): 775–808. https://doi.org/10.1128/MMBR.00023-09.

Ammar, Ehab M., Xiaoyi Wang, and Christopher V. Rao. 2018. "Regulation of Metabolism in Escherichia Coli during Growth on Mixtures of the Non-Glucose Sugars: Arabinose, Lactose, and Xylose." *Scientific Reports* 8 (1): 609. https://doi.org/10.1038/s41598-017-18704-0.

Amyot, Whitney M., Dennise deJesus, and Ralph R. Isberg. 2013. "Poison Domains Block Transit of Translocated Substrates via the Legionella Pneumophila Icm/Dot System." *Infection and Immunity* 81 (9): 3239–52. https://doi.org/10.1128/IAI.00552-13.

Archer, David B., and Norman R. King. 1984. "Isolation of Gas Vesicles from Methanosarcina Barken." *Microbiology* 130 (1): 167–72. https://www.microbiologyresearch.org/content/journal/micro/10.1099/00221287-130-1-167.

Ashkenazy, Haim, Shiran Abadi, Eric Martz, Ofer Chay, Itay Mayrose, Tal Pupko, and Nir Ben-Tal. 2016. "ConSurf 2016: An Improved Methodology to Estimate and Visualize Evolutionary Conservation in Macromolecules." *Nucleic Acids Research* 44 (W1): W344–50. https://doi.org/10.1093/nar/gkw408.

Backert, Steffen, and Elisabeth Grohmann. 2018. *Type IV Secretion in Gram-Negative and Gram-Positive Bacteria*. Springer. https://play.google.com/store/books/details?id=CjtRDwAAQBAJ.

Bar-Zion, Avinoam, Atousa Nourmahnad, David R. Mittelstein, Shirin Shivaei, Sangjin Yoo, Marjorie T. Buss, Robert C. Hurt, et al. 2021. "Acoustically Triggered Mechanotherapy Using Genetically Encoded Gas Vesicles." *Nature Nanotechnology* 16 (12): 1403–12. https://doi.org/10.1038/s41565-021-00971-8.

Bar-Zion, Avinoam, Atousa Nourmahnad, David R. Mittelstein, Sangjin Yoo, Dina Malounda, Mohamad Abedi, Audrey Lee-Gosselin, David Maresca, and Mikhail G. Shapiro. 2019. "Acoustically Detonated Biomolecules for Genetically Encodable Inertial Cavitation." *Cold Spring Harbor Laboratory*. https://doi.org/10.1101/620567.

Beard, Steven J., P. A. Davis, D. Iglesias-Rodrı´guez, O. M. Skulberg, and Anthony E. Walsby. 2000. "Gas Vesicle Genes in Planktothrix Spp. from Nordic Lakes: Strains with Weak Gas Vesicles Possess a Longer Variant of gvpCThe GenBank Accession Numbers for the New Sequences in This Paper Are AJ253125–253133." *Microbiology* 146 (8): 2009–18. https://doi.org/10.1099/00221287-146-8-2009.

Beard, S. J., B. A. Handley, P. K. Hayes, and Anthony E. Walsby. 1999. "The Diversity of Gas Vesicle Genes in Planktothrix Rubescens from Lake Zürich." *Microbiology* 145 (Pt 10) (October): 2757–68. https://doi.org/10.1099/00221287-145-10-2757.

Belenky, M., R. Meyers, and J. Herzfeld. 2004. "Subunit Structure of Gas Vesicles: A MALDI-TOF Mass Spectrometry Study." *Biophysical Journal*. https://www.sciencedirect.com/science/article/pii/S0006349504741284.

Berhanu, Workalemahu M., Erik J. Alred, Nathan A. Bernhardt, and Ulrich H. E. Hansmann. 2015. "All-Atom Simulation of Amyloid Aggregates." *Physics Procedia* 68 (January): 61–68. https://doi.org/10.1016/j.phpro.2015.07.110.

Bisagni, C. 2000. "Numerical Analysis and Experimental Correlation of Composite Shell Buckling and Post-Buckling." *Composites Part B Engineering* 31 (8): 655–67. https://doi.org/10.1016/S1359-8368(00)00031-7.

Böck, Désirée, Dario Hüsler, Bernhard Steiner, João M. Medeiros, Amanda Welin, Katarzyna A. Radomska, Wolf-Dietrich Hardt, Martin Pilhofer, and Hubert Hilbi. 2021. "The Polar Legionella Icm/Dot T4SS Establishes Distinct Contact Sites with the Pathogen Vacuole Membrane." *mBio* 12 (5): e0218021. https://doi.org/10.1128/mBio.02180-21.

Bourdeau, Raymond W., Audrey Lee-Gosselin, Anupama Lakshmanan, Arash Farhadi, Sripriya Ravindra Kumar, Suchita P. Nety, and Mikhail G. Shapiro. 2018. "Acoustic Reporter Genes for Noninvasive Imaging of Microorganisms in Mammalian Hosts." *Nature* 553 (7686): 86–90. https://doi.org/10.1038/nature25021.

Buchholz, B. E., P. K. Hayes, and A. E. Walsby. 1993. "The Distribution of the Outer Gas Vesicle Protein, GvpC, on the Anabaena Gas Vesicle, and Its Ratio to GvpA." *Journal of General Microbiology* 139 (10): 2353–63. https://doi.org/10.1099/00221287-139-10-2353.

Burt, Alister, Lorenzo Gaifas, Tom Dendooven, and Irina Gutsche. 2021. "A Flexible Framework for Multi-Particle Refinement in Cryo-Electron Tomography." *PLoS Biology* 19 (8): e3001319. https://doi.org/10.1371/journal.pbio.3001319.

Cai, Kun, Bo-Ying Xu, Yong-Liang Jiang, Ying Wang, Yuxing Chen, Cong-Zhao Zhou, and Qiong Li. 2020. "The Model Cyanobacteria Anabaena Sp. PCC 7120 Possess an Intact but Partially Degenerated Gene Cluster Encoding Gas Vesicles." *BMC Microbiology* 20 (1): 110. https://doi.org/10.1186/s12866-020-01805-8.

Callaway, Ewen. 2020. "Revolutionary Cryo-EM Is Taking over Structural Biology." *Nature* 578 (7794): 201. https://doi.org/10.1038/d41586-020-00341-9.

Cascales, Eric, and Peter J. Christie. 2003. "The Versatile Bacterial Type IV Secretion Systems." *Nature Reviews. Microbiology* 1 (2): 137–49. https://doi.org/10.1038/nrmicro753.

Castaño-Díez, Daniel, Mikhail Kudryashev, Marcel Arheit, and Henning Stahlberg. 2012. "Dynamo: A Flexible, User-Friendly Development Tool for Subtomogram Averaging of Cryo-EM Data in High-Performance Computing Environments." *Journal of Structural Biology* 178 (2): 139–51. https://doi.org/10.1016/j.jsb.2011.12.017.

Castaño-Díez, Daniel, Mikhail Kudryashev, and Henning Stahlberg. 2017. "Dynamo Catalogue: Geometrical Tools and Data Management for Particle Picking in Subtomogram Averaging of Cryo-Electron Tomograms." *Journal of Structural Biology* 197 (2): 135–44. https://doi.org/10.1016/j.jsb.2016.06.005.

Chang, Jeong Ho, and Yeon-Gil Kim. 2015. "Crystal Structure of the Bacterial Type VI Secretion System Component TssL from Vibrio Cholerae." *Journal of Microbiology* 53 (1): 32–37. https://doi.org/10.1007/s12275-015-4539-0.

Chang, Yi-Wei, Lee A. Rettberg, Davi R. Ortega, and Grant J. Jensen. 2017. "In Vivo Structures of an Intact Type VI Secretion System Revealed by Electron Cryotomography." *EMBO Reports* 18 (7): 1090–99. https://doi.org/10.15252/embr.201744072.

Chang, Yi-Wei, Carrie L. Shaffer, Lee A. Rettberg, Debnath Ghosal, and Grant J. Jensen. 2018. "In Vivo Structures of the Helicobacter Pylori Cag Type IV Secretion System." *Cell Reports* 23 (3): 673–81. https://doi.org/10.1016/j.celrep.2018.03.085.

Chen, Muyuan, James M. Bell, Xiaodong Shi, Stella Y. Sun, Zhao Wang, and Steven J. Ludtke. 2019. "A Complete Data Processing Workflow for Cryo-ET and Subtomogram Averaging." *Nature Methods* 16 (11): 1161–68. https://doi.org/10.1038/s41592-019-0591-8.

Chen, Zhen, Momoko Shiozaki, Kelsey M. Haas, Shumei Zhao, Caiying Guo, Benjamin J. Polacco, Zhiheng Yu, et al. 2022. "De Novo Protein Identification in Mammalian Sperm Using High-Resolution in Situ Cryo-Electron Tomography." *bioRxiv*. https://doi.org/10.1101/2022.09.28.510016.

Cherin, Emmanuel, Johan M. Melis, Raymond W. Bourdeau, Melissa Yin, Dennis M. Kochmann, F. Stuart Foster, and Mikhail G. Shapiro. 2017. "Acoustic Behavior of Halobacterium Salinarum Gas Vesicles in the High-Frequency Range: Experiments and Modeling." *Ultrasound in Medicine & Biology* 43 (5): 1016–30. https://doi.org/10.1016/j.ultrasmedbio.2016.12.020.

Chetrit, David, Bo Hu, Peter J. Christie, Craig R. Roy, and Jun Liu. 2018. "A Unique Cytoplasmic ATPase Complex Defines the Legionella Pneumophila Type IV Secretion Channel." *Nature Microbiology*. https://doi.org/10.1038/s41564-018-0165-z.

Choi, Jae Young, Tae-Ho Jang, and Hyun Ho Park. 2017. "The Mechanism of Folding Robustness Revealed by the Crystal Structure of Extra-Superfolder GFP." *FEBS Letters* 591 (2): 442–47. https://doi.org/10.1002/1873-3468.12534.

Chreifi, Georges, Songye Chen, and Grant J. Jensen. 2021. "Rapid Tilt-Series Method for Cryo-Electron Tomography: Characterizing Stage Behavior during FISE Acquisition." *Journal of Structural Biology* 213 (2): 107716. https://doi.org/10.1016/j.jsb.2021.107716.

Chreifi, Georges, Songye Chen, Lauren Ann Metskas, Mohammed Kaplan, and Grant J. Jensen. 2019. "Rapid Tilt-Series Acquisition for Electron Cryotomography." *Journal of Structural Biology* 205 (2): 163–69. https://doi.org/10.1016/j.jsb.2018.12.008.

Christie, Peter J. 2016. "The Mosaic Type IV Secretion Systems." *EcoSal Plus* 7 (1). https://doi.org/10.1128/ecosalplus.ESP-0020-2015.

Christie, Peter J., Laura Gomez Valero, and Carmen Buchrieser. 2017. "Biological Diversity and Evolution of Type IV Secretion Systems." *Current Topics in Microbiology and Immunology* 413: 1–30. https://doi.org/10.1007/978-3-319-75241-9_1.

Chung, Jeong Min, Michael J. Sheedlo, Anne M. Campbell, Neha Sawhney, Arwen E. Frick-Cheng, Dana Borden Lacy, Timothy L. Cover, and Melanie D. Ohi. 2019. "Structure of the Helicobacter Pylori Cag Type IV Secretion System." *eLife* 8 (June). https://doi.org/10.7554/eLife.47644.

Combe, Colin W., Lutz Fischer, and Juri Rappsilber. 2015. "xiNET: Cross-Link Network Maps with Residue Resolution." *Molecular & Cellular Proteomics: MCP* 14 (4): 1137–47. https://doi.org/10.1074/mcp.O114.042259.

C.P.van Zundert, Gydo, Bijvoet Center for Biomolecular Research, Faculty of Science-Chemistry, Utrecht University, Utrecht, the Netherlands, and Alexandre M.J.J. Bonvin. 2015. "Fast and Sensitive Rigid-Body Fitting into Cryo-EM Density Maps with PowerFit." *AIMS Biophysics* 2 (2): 73–87. https://doi.org/10.3934/biophy.2015.2.73.

Crooks, Gavin E., Gary Hon, John-Marc Chandonia, and Steven E. Brenner. 2004. "WebLogo: A Sequence Logo Generator." *Genome Research* 14 (6): 1188–90. https://doi.org/10.1101/gr.849004.

Deutsch, Eric W., Nuno Bandeira, Vagisha Sharma., et al. 2020. "The ProteomeXchange Consortium in 2020: Enabling 'Big Data'approaches in Proteomics." *Nucleic Acids and Molecular Biology*. https://academic.oup.com/nar/article-abstract/48/D1/D1145/5612546.

Dooling, Kathleen L., Karrie-Ann Toews, Lauri A. Hicks, Laurel E. Garrison, Brian Bachaus, Shelley Zansky, L. Rand Carpenter, et al. 2015. "Active Bacterial Core Surveillance for Legionellosis - United States, 2011-2013." *MMWR. Morbidity and Mortality Weekly Report* 64 (42): 1190–93. https://doi.org/10.15585/mmwr.mm6442a2.

Dubnau, David, and Melanie Blokesch. 2019. "Mechanisms of DNA Uptake by Naturally Competent Bacteria." *Annual Review of Genetics* 53 (December): 217–37. https://doi.org/10.1146/annurev-genet-112618-043641.

Dunton, Peter G., and Anthony E. Walsby. 2005. "The Diameter and Critical Collapse Pressure of Gas Vesicles in Microcystis Are Correlated with GvpCs of Different Length." *FEMS Microbiology Letters* 247 (1): 37–43. https://doi.org/10.1016/j.femsle.2005.04.026.

Durand, Eric, Abdelrahim Zoued, Silvia Spinelli, Paul J. H. Watson, Marie-Stéphanie Aschtgen, Laure Journet, Christian Cambillau, and Eric Cascales. 2012. "Structural Characterization and Oligomerization of the TssL Protein, a Component Shared by Bacterial Type VI and Type IVb Secretion Systems." *The Journal of Biological Chemistry* 287 (17): 14157–68. https://doi.org/10.1074/jbc.M111.338731.

Durie, Clarissa L., Michael J. Sheedlo, Jeong Min Chung, Brenda G. Byrne, Min Su, Thomas Knight, Michele Swanson, D. Borden Lacy, and Melanie D. Ohi. 2020. "Structural Analysis of the Legionella Pneumophila Dot/Icm Type IV Secretion System Core Complex." *eLife* 9 (September). https://doi.org/10.7554/eLife.59530.

Dutka, Przemysław, Dina Malounda, Lauren Ann Metskas, Songye Chen, Robert C. Hurt, George J. Lu, Grant J. Jensen, and Mikhail G. Shapiro. 2021. "Measuring Gas Vesicle Dimensions by Electron Microscopy." *Protein Science: A Publication of the Protein Society* 30 (5): 1081–86. https://doi.org/10.1002/pro.4056.

Dutka, Przemysław, Lauren Ann Metskas, Robert C. Hurt, Hossein Salahshoor, Ting-Yu Wang, Dina Malounda, George Lu, Tsui-Fen Chou, Mikhail G. Shapiro, and Grant J. Jensen. 2022. "Structure of Anabaena Flos-Aquae Gas Vesicles Revealed by Cryo-ET." *bioRxiv*. https://doi.org/10.1101/2022.06.21.496981.

Edgar, Robert C. 2004a. "MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput." *Nucleic Acids Research* 32 (5): 1792–97. https://doi.org/10.1093/nar/gkh340.

Robert C Edgar. 2004b. "MUSCLE: A Multiple Sequence Alignment Method with Reduced Time and Space Complexity." *BMC Bioinformatics* 5 (August): 113. https://doi.org/10.1186/1471-2105-5-113.

Egelman, Edward H. 2015. "Three-Dimensional Reconstruction of Helical Polymers." *Archives of Biochemistry and Biophysics* 581 (September): 54–58. https://doi.org/10.1016/j.abb.2015.04.004.

Emsley, P., B. Lohkamp, W. G. Scott, and K. Cowtan. 2010. "Features and Development of Coot." *Acta Crystallographica. Section D, Biological Crystallography* 66 (Pt 4): 486–501. https://doi.org/10.1107/S0907444910007493.

Englert, Christoph, Mary Horne, and Felicitas Pfeifer. 1990. “Expression of the Major Gas Vesicle Protein Gene in the Halophilic archaebacteriumHaloferax Mediterranei Is Modulated by Salt.” *Molecular & General Genetics: MGG* 222 (2): 225–32. https://doi.org/10.1007/BF00633822.

Escoll, Pedro, Ok-Ryul Song, Flávia Viana, Bernhard Steiner, Thibault Lagache, Jean-Christophe Olivo-Marin, Francis Impens, Priscille Brodin, Hubert Hilbi, and Carmen Buchrieser. 2017. “Legionella Pneumophila Modulates Mitochondrial Dynamics to Trigger Metabolic Repurposing of Infected Macrophages.” *Cell Host & Microbe* 22 (3): 302–16.e7. https://doi.org/10.1016/j.chom.2017.07.020.

Evans, Richard, Michael O’Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, et al. 2022. “Protein Complex Prediction with AlphaFold-Multimer.” *bioRxiv*. https://doi.org/10.1101/2021.10.04.463034.

Farhadi, Arash, Manuel Bedrossian, Justin Lee, Gabrielle H. Ho, Mikhail G. Shapiro, and Jay L. Nadeau. 2020. “Genetically Encoded Phase Contrast Agents for Digital Holographic Microscopy.” *Nano Letters* 20 (11): 8127–34. https://doi.org/10.1021/acs.nanolett.0c03159.

Farhadi, Arash, Gabrielle H. Ho, Daniel P. Sawyer, Raymond W. Bourdeau, and Mikhail G. Shapiro. 2019. “Ultrasound Imaging of Gene Expression in Mammalian Cells.” *Science* 365 (6460): 1469–75. https://doi.org/10.1126/science.aax4804.

Farhadi, Arash, Gabrielle Ho, Martin Kunth, Bill Ling, Anupama Lakshmanan, George Lu, Raymond W. Bourdeau, Leif Schröder, and Mikhail G. Shapiro. 2018. “Recombinantly Expressed Gas Vesicles as Nanoscale Contrast Agents for Ultrasound and Hyperpolarized MRI.” *AIChE Journal. American Institute of Chemical Engineers* 64 (8): 2927–33. https://doi.org/10.1002/aic.16138.

Frank, Joachim. 2006. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*. Oxford University Press. https://play.google.com/store/books/details?id=kjGKy2LeWnUC.

Fromm, Simon A., Kate M. O’Connor, Michael Purdy, Pramod R. Bhatt, Gary Loughran, John F. Atkins, Ahmad Jomaa, and Simone Mattei. 2023. “The Translating Bacterial Ribosome at 1.55 Å Resolution Generated by Cryo-EM Imaging Services.” *Nature Communications* 14 (1): 1095. https://doi.org/10.1038/s41467-023-36742-3.

Gemmer, Max, Marten L. Chaillet, Joyce van Loenhout, Rodrigo Cuevas Arenas, Dimitrios Vismpas, Mariska Gröllers-Mulderij, Fujiet A. Koh, et al. 2023. “Visualization of Translation and Protein Biogenesis at the ER Membrane.” *Nature* 614 (7946): 160–67. https://doi.org/10.1038/s41586-022-05638-5.

Ghosal, Debnath, Yi-wei Chang, Kwangcheol C. Jeong, Joseph P. Vogel, and Grant J. Jensen. 2017. “*In Situ* Structure of the *Legionella* Dot/Icm Type IV Secretion System by Electron Cryotomography.” *EMBO Reports*. https://doi.org/10.15252/embr.201643598.

Ghosal, Debnath, Kwangcheol C. Jeong, Yi-Wei Chang, Jacob Gyore, Lin Teng, Adam Gardner, Joseph P. Vogel, and Grant J. Jensen. 2019. “Molecular Architecture, Polar Targeting and Biogenesis of the Legionella Dot/Icm T4SS.” *Nature Microbiology*. https://doi.org/10.1038/s41564-019-0427-4.

Goddard, Thomas D., Conrad C. Huang, Elaine C. Meng, Eric F. Pettersen, Gregory S. Couch, John H. Morris, and Thomas E. Ferrin. 2018. “UCSF ChimeraX: Meeting Modern Challenges in Visualization and Analysis.” *Protein Science: A Publication of the Protein Society* 27 (1): 14–25. https://doi.org/10.1002/pro.3235.

Gomez-Valero, Laura, Alvaro Chiner-Oms, Iñaki Comas, and Carmen Buchrieser. 2019. "Evolutionary Dissection of the Dot/Icm System Based on Comparative Genomics of 58 Legionella Species." *Genome Biology and Evolution* 11 (9): 2619–32. https://doi.org/10.1093/gbe/evz186.

Gosline, John, Margo Lillie, Emily Carrington, Paul Guerette, Christine Ortlepp, and Ken Savage. 2002. "Elastic Proteins: Biological Roles and Mechanical Properties." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 357 (1418): 121–32. https://doi.org/10.1098/rstb.2001.1022.

Griffiths, A. E., A. E. Walsby, and P. K. Hayes. 1992. "The Homologies of Gas Vesicle Proteins." *Journal of General Microbiology* 138 (6): 1243–50. https://doi.org/10.1099/00221287-138-6-1243.

Hallgren, Jeppe, Konstantinos D. Tsirigos, Mads Damgaard Pedersen, José Juan Almagro Armenteros, Paolo Marcatili, Henrik Nielsen, Anders Krogh, and Ole Winther. 2022. "DeepTMHMM Predicts Alpha and Beta Transmembrane Proteins Using Deep Neural Networks." *bioRxiv*. https://doi.org/10.1101/2022.04.08.487609.

Hare, Stephen, Richard Bayliss, Christian Baron, and Gabriel Waksman. 2006. "A Large Domain Swap in the VirB11 ATPase of Brucella Suis Leaves the Hexameric Assembly Intact." *Journal of Molecular Biology* 360 (1): 56–66. https://doi.org/10.1016/j.jmb.2006.04.060.

Hayes, P. K., B. Buchholz, and A. E. Walsby. 1992. "Gas Vesicles Are Strengthened by the Outer-Surface Protein, GvpC." *Archives of Microbiology* 157 (3): 229–34. https://doi.org/10.1007/BF00245155.

Hayes, P. K., and A. E. Walsby. 1986. "The Inverse Correlation between Width and Strength of Gas Vesicles in Cyanobacteria." *British Phycological Journal* 21 (2): 191–97. https://doi.org/10.1080/00071618600650221.

He, Shaoda, and Sjors H. W. Scheres. 2017. "Helical Reconstruction in RELION." *Journal of Structural Biology* 198 (3): 163–76. https://doi.org/10.1016/j.jsb.2017.02.003.

Huang, Yen-Hua, and Cheng-Yang Huang. 2018. "The Glycine-Rich Flexible Region in SSB Is Crucial for PriA Stimulation." *RSC Advances* 8 (61): 35280–88. https://doi.org/10.1039/c8ra07306f.

Hubber, Andree, and Craig R. Roy. 2010. "Modulation of Host Cell Function by Legionella Pneumophila Type IV Effectors." *Annual Review of Cell and Developmental Biology* 26: 261–83. https://doi.org/10.1146/annurev-cellbio-100109-104034.

Huber, Stefan T., Dion Terwiel, Wiel H. Evers, David Maresca, and Arjen J. Jakobi. 2022. "Cryo-EM Structure of Gas Vesicles for Buoyancy-Controlled Motility." *bioRxiv*. https://doi.org/10.1101/2022.05.08.489936.

Hu, Bo, Pratick Khara, and Peter J. Christie. 2019. "Structural Bases for F Plasmid Conjugation and F Pilus Biogenesis in *Escherichia Coli*." *Proceedings of the National Academy of Sciences of the United States of America* 116 (28): 14222–27. https://doi.org/10.1073/pnas.1904428116.

Hurt, Robert C., Marjorie T. Buss, Mengtong Duan, Katie Wong, Mei Yi You, Daniel P. Sawyer, Margaret B. Swift, et al. 2023. "Genomically Mined Acoustic Reporter Genes for Real-Time in Vivo Monitoring of Tumors and Tumor-Homing Bacteria." *Nature Biotechnology*, January. https://doi.org/10.1038/s41587-022-01581-y.

Hylton, Ryan K., and Matthew T. Swulius. 2021. “Challenges and Triumphs in Cryo-Electron Tomography.” *iScience* 24 (9): 102959. https://doi.org/10.1016/j.isci.2021.102959.

Jain, Rinku, William J. Rice, Radhika Malik, Robert E. Johnson, Louise Prakash, Satya Prakash, Iban Ubarretxena-Belandia, and Aneel K. Aggarwal. 2019. “Cryo-EM Structure and Dynamics of Eukaryotic DNA Polymerase δ Holoenzyme.” *Nature Structural & Molecular Biology* 26 (10): 955–62. https://doi.org/10.1038/s41594-019-0305-z.

Jaskolski, Mariusz, Zbigniew Dauter, and Alexander Wlodawer. 2014. “A Brief History of Macromolecular Crystallography, Illustrated by a Family Tree and Its Nobel Fruits.” *The FEBS Journal* 281 (18): 3985–4009. https://doi.org/10.1111/febs.12796.

Jost, Alisa, and Felicitas Pfeifer. 2022. “Interaction of the Gas Vesicle Proteins GvpA, GvpC, GvpN, and GvpO of Halobacterium Salinarum.” *Frontiers in Microbiology* 13 (July): 971917. https://doi.org/10.3389/fmicb.2022.971917.

Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, et al. 2021. “Highly Accurate Protein Structure Prediction with AlphaFold.” *Nature* 596 (7873): 583–89. https://doi.org/10.1038/s41586-021-03819-2.

Kanehisa, M. 1997. “Linking Databases and Organisms: GenomeNet Resources in Japan.” *Trends in Biochemical Sciences* 22 (11): 442–44. https://doi.org/10.1016/s0968-0004(97)01130-4.

Kaplan, Mohammed, Doulin C. Shepherd, Naveen Vankadari, Ki Woo Kim, Charles L. Larson, Przemysław Dutka, Paul A. Beare, et al. 2022. “Structural Remodeling of Coxiella Burnetii during Its Biphasic Developmental Cycle Revealed by Cryo-Electron Tomography.” *bioRxiv*. https://doi.org/10.1101/2022.08.23.505044.

Khara, Pratick, Liqiang Song, Peter J. Christie, and Bo Hu. 2021. “In Situ Visualization of the pKM101-Encoded Type IV Secretion System Reveals a Highly Symmetric ATPase Energy Center.” *mBio* 12 (5): e0246521. https://doi.org/10.1128/mBio.02465-21.

Knitsch, Regine, Marie Schneefeld, Kerstin Weitzel, and Felicitas Pfeifer. 2017. “Mutations in the Major Gas Vesicle Protein GvpA and Impacts on Gas Vesicle Formation in Haloferax Volcanii.” *Molecular Microbiology* 106 (4): 530–42. https://doi.org/10.1111/mmi.13833.

Komeili, Arash, Zhuo Li, Dianne K. Newman, and Grant J. Jensen. 2006. “Magnetosomes Are Cell Membrane Invaginations Organized by the Actin-like Protein MamK.” *Science* 311 (5758): 242–45. https://doi.org/10.1126/science.1123231.

Koukos, P. I., and A. M. J. J. Bonvin. 2020. “Integrative Modelling of Biomolecular Complexes.” *Journal of Molecular Biology* 432 (9): 2861–81. https://doi.org/10.1016/j.jmb.2019.11.009.

Krause, Duncan C., Songye Chen, Jian Shi, Ashley J. Jensen, Edward S. Sheppard, and Grant J. Jensen. 2018. “Electron Cryotomography of Mycoplasma Pneumoniae Mutants Correlates Terminal Organelle Architectural Features and Function.” *Molecular Microbiology* 108 (3): 306–18. https://doi.org/10.1111/mmi.13937.

Kremer, J. R., D. N. Mastronarde, and J. R. McIntosh. 1996. “Computer Visualization of Three-Dimensional Image Data Using IMOD.” *Journal of Structural Biology* 116 (1): 71–76. https://doi.org/10.1006/jsbi.1996.0013.

Kubori, Tomoko, and Hiroki Nagai. 2016. “The Type IVB Secretion System: An Enigmatic Chimera.” *Current Opinion in Microbiology* 29 (February): 22–29. https://doi.org/10.1016/j.mib.2015.10.001.

Kucukelbir, Alp, Fred J. Sigworth, and Hemant D. Tagare. 2014. “Quantifying the Local Resolution of Cryo-EM Density Maps.” *Nature Methods* 11 (1): 63–65. https://doi.org/10.1038/nmeth.2727.

Kühlbrandt, Werner. 2014. "Biochemistry. The Resolution Revolution." *Science*. https://doi.org/10.1126/science.1251652.

Kunth, Martin, George J. Lu, Christopher Witte, Mikhail G. Shapiro, and Leif Schröder. 2018. "Protein Nanostructures Produce Self-Adjusting Hyperpolarized Magnetic Resonance Imaging Contrast through Physical Gas Partitioning." *ACS Nano* 12 (11): 10939–48. https://doi.org/10.1021/acsnano.8b04222.

Kuroda, Takuya, Tomoko Kubori, Xuan Thanh Bui, Akihiro Hyakutake, Yumiko Uchida, Katsumi Imada, and Hiroki Nagai. 2015. "Molecular and Structural Analysis of Legionella DotI Gives Insights into an Inner Membrane Complex Essential for Type IV Secretion." *Scientific Reports* 5 (June): 10912. https://doi.org/10.1038/srep10912.

Lakshmanan, Anupama, Arash Farhadi, Suchita P. Nety, Audrey Lee-Gosselin, Raymond W. Bourdeau, David Maresca, and Mikhail G. Shapiro. 2016. "Molecular Engineering of Acoustic Protein Nanostructures." *ACS Nano* 10 (8): 7314–22. https://doi.org/10.1021/acsnano.6b03364.

Lakshmanan, Anupama, Zhiyang Jin, Suchita P. Nety, Daniel P. Sawyer, Audrey Lee-Gosselin, Dina Malounda, Mararet B. Swift, David Maresca, and Mikhail G. Shapiro. 2020. "Acoustic Biosensors for Ultrasound Imaging of Enzyme Activity." *Nature Chemical Biology* 16 (9): 988–96. https://doi.org/10.1038/s41589-020-0591-0.

Lakshmanan, Anupama, George J. Lu, Arash Farhadi, Suchita P. Nety, Martin Kunth, Audrey Lee-Gosselin, David Maresca, et al. 2017. "Preparation of Biogenic Gas Vesicle Nanostructures for Use as Contrast Agents for Ultrasound and MRI." *Nature Protocols* 12 (10): 2050–80. https://doi.org/10.1038/nprot.2017.081.

Lam, Vinson, and Elizabeth Villa. 2021. "Practical Approaches for Cryo-FIB Milling and Applications for Cellular Cryo-Electron Tomography." *Methods in Molecular Biology* 2215: 49–82. https://doi.org/10.1007/978-1-0716-0966-8_3.

Lancey, Claudia, Muhammad Tehseen, Souvika Bakshi, Matthew Percival, Masateru Takahashi, Mohamed A. Sobhy, Vlad S. Raducanu, et al. 2021. "Cryo-EM Structure of Human Pol ϰ Bound to DNA and Mono-Ubiquitylated PCNA." *Nature Communications* 12 (1): 6095. https://doi.org/10.1038/s41467-021-26251-6.

Lanzi, Luca. 2004. "A Numerical and Experimental Investigation on Composite Stiffened Panels into Post-Buckling." *Thin-Walled Structures* 42 (12): 1645–64. https://doi.org/10.1016/j.tws.2004.06.001.

Liberta, Falk, Sarah Loerch, Matthies Rennegarbe, Angelika Schierhorn, Per Westermark, Gunilla T. Westermark, Bouke P. C. Hazenberg, Nikolaus Grigorieff, Marcus Fändrich, and Matthias Schmidt. 2019. "Cryo-EM Fibril Structures from Systemic AA Amyloidosis Reveal the Species Complementarity of Pathological Amyloids." *Nature Communications* 10 (1): 1104. https://doi.org/10.1038/s41467-019-09033-z.

Li, N., and M. C. Cannon. 1998. "Gas Vesicle Genes Identified in Bacillus Megaterium and Functional Expression in Escherichia Coli." *Journal of Bacteriology* 180 (9): 2450–58. https://doi.org/10.1128/JB.180.9.2450-2458.1998.

Liu, Yun-Tao, Heng Zhang, Hui Wang, Chang-Lu Tao, Guo-Qiang Bi, and Z. Hong Zhou. 2022. "Isotropic Reconstruction for Electron Tomography with Deep Learning." *Nature Communications* 13 (1): 6482. https://doi.org/10.1038/s41467-022-33957-8.

Li, Yang Grace, Bo Hu, and Peter J. Christie. 2019. "Biological and Structural Diversity of Type IV Secretion Systems." *Microbiology Spectrum* 7 (2). https://doi.org/10.1128/microbiolspec.PSIB-0012-2018.

Lockwood, Daniel C., Himani Amin, Tiago R. D. Costa, and Gunnar N. Schroeder. 2022. "The Legionella Pneumophila Dot/Icm Type IV Secretion System and Its Effectors." *Microbiology* 168 (5). https://doi.org/10.1099/mic.0.001187.

Logger, Laureen, Marie-Stéphanie Aschtgen, Marie Guérin, Eric Cascales, and Eric Durand. 2016. "Molecular Dissection of the Interface between the Type VI Secretion TssM Cytoplasmic Domain and the TssG Baseplate Component." *Journal of Molecular Biology* 428 (22): 4424–37. https://doi.org/10.1016/j.jmb.2016.08.032.

Lomize, Andrei L., Spencer C. Todd, and Irina D. Pogozheva. 2022. "Spatial Arrangement of Proteins in Planar and Curved Membranes by PPM 3.0." *Protein Science: A Publication of the Protein Society* 31 (1): 209–20. https://doi.org/10.1002/pro.4219.

Lu, George J., Li-Dek Chou, Dina Malounda, Amit K. Patel, Derek S. Welsbie, Daniel L. Chao, Tirunelveli Ramalingam, and Mikhail G. Shapiro. 2020. "Genetically Encodable Contrast Agents for Optical Coherence Tomography." *ACS Nano* 14 (7): 7823–31. https://doi.org/10.1021/acsnano.9b08432.

Lu, George J., Arash Farhadi, Jerzy O. Szablowski, Audrey Lee-Gosselin, Samuel R. Barnes, Anupama Lakshmanan, Raymond W. Bourdeau, and Mikhail G. Shapiro. 2018. "Acoustically Modulated Magnetic Resonance Imaging of Gas-Filled Protein Nanostructures." *Nature Materials* 17 (5): 456–63. https://doi.org/10.1038/s41563-018-0023-7.

Macé, Kévin, Abhinav K. Vadakkepat, Adam Redzej, Natalya Lukoyanova, Clasien Oomen, Nathalie Braun, Marta Ukleja, et al. 2022. "Cryo-EM Structure of a Type IV Secretion System." *Nature* 607 (7917): 191–96. https://doi.org/10.1038/s41586-022-04859-y.

Ma, Lay-Sun, Franz Narberhaus, and Erh-Min Lai. 2012. "IcmF Family Protein TssM Exhibits ATPase Activity and Energizes Type VI Secretion." *The Journal of Biological Chemistry* 287 (19): 15610–21. https://doi.org/10.1074/jbc.M111.301630.

Maley, Adam M., George J. Lu, Mikhail G. Shapiro, and Robert M. Corn. 2017. "Characterizing Single Polymeric and Protein Nanoparticles with Surface Plasmon Resonance Imaging Measurements." *ACS Nano* 11 (7): 7447–56. https://doi.org/10.1021/acsnano.7b03859.

Maresca, David, Anupama Lakshmanan, Mohamad Abedi, Avinoam Bar-Zion, Arash Farhadi, George J. Lu, Jerzy O. Szablowski, Di Wu, Sangjin Yoo, and Mikhail G. Shapiro. 2018. "Biomolecular Ultrasound and Sonogenetics." *Annual Review of Chemical and Biomolecular Engineering* 9 (June): 229–52. https://doi.org/10.1146/annurev-chembioeng-060817-084034.

Maresca, David, Anupama Lakshmanan, Audrey Lee-Gosselin, Johan M. Melis, Yu-Li Ni, Raymond W. Bourdeau, Dennis M. Kochmann, and Mikhail G. Shapiro. 2017. "Nonlinear Ultrasound Imaging of Nanoscale Acoustic Biomolecules." *Applied Physics Letters*. https://doi.org/10.1063/1.4976105.

Maresca, David, Daniel P. Sawyer, Guillaume Renaud, Audrey Lee-Gosselin, and Mikhail G. Shapiro. 2018. "Nonlinear X-Wave Ultrasound Imaging of Acoustic Biomolecules." *Physical Review. X* 8 (4). https://doi.org/10.1103/physrevx.8.041002.

Mastronarde, David N. 2005. "Automated Electron Microscope Tomography Using Robust Prediction of Specimen Movements." *Journal of Structural Biology* 152 (1): 36–51. https://doi.org/10.1016/j.jsb.2005.07.007.

Mastronarde, David N., and Susannah R. Held. 2017. "Automated Tilt Series Alignment and Tomographic Reconstruction in IMOD." *Journal of Structural Biology* 197 (2): 102–13. https://doi.org/10.1016/j.jsb.2016.07.011.

Matthews, M., and C. R. Roy. 2000. "Identification and Subcellular Localization of the Legionella Pneumophila IcmX Protein: A Factor Essential for Establishment of a Replicative Organelle in Eukaryotic Host Cells." *Infection and Immunity* 68 (7): 3971–82. https://doi.org/10.1128/IAI.68.7.3971-3982.2000.

Meir, Amit, Kevin Macé, Natalya Lukoyanova, David Chetrit, Manuela K. Hospenthal, Adam Redzej, Craig Roy, and Gabriel Waksman. 2020. "Mechanism of Effector Capture and Delivery by the Type IV Secretion System from Legionella Pneumophila." *Nature Communications* 11 (1): 2864. https://doi.org/10.1038/s41467-020-16681-z.

Merkley, Eric D., Steven Rysavy, Abdullah Kahraman, Ryan P. Hafen, Valerie Daggett, and Joshua N. Adkins. 2014. "Distance Restraints from Crosslinking Mass Spectrometry: Mining a Molecular Dynamics Simulation Database to Evaluate Lysine-Lysine Distances." *Protein Science: A Publication of the Protein Society* 23 (6): 747–59. https://doi.org/10.1002/pro.2458.

Metskas, Lauren Ann, Davi Ortega, Luke M. Oltrogge, Cecilia Blikstad, Tom Laughlin, David F. Savage, and Grant J. Jensen. 2022. "Rubisco Forms a Lattice inside Alpha-Carboxysomes." *bioRxiv*. https://doi.org/10.1101/2022.01.24.477598.

Mirdita, Milot, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. 2022. "ColabFold: Making Protein Folding Accessible to All." *Nature Methods*, May. https://doi.org/10.1038/s41592-022-01488-1.

Mosalaganti, Shyamal, Agnieszka Obarska-Kosinska, Marc Siggel, Reiya Taniguchi, Beata Turoňová, Christian E. Zimmerli, Katarzyna Buczak, et al. 2022. "AI-Based Structure Prediction Empowers Integrative Structural Analysis of Human Nuclear Pores." *Science* 376 (6598): eabm9506. https://doi.org/10.1126/science.abm9506.

Nagai, H., and C. R. Roy. 2001. "The DotA Protein from Legionella Pneumophila Is Secreted by a Novel Process That Requires the Dot/Icm Transporter." *The EMBO Journal* 20 (21): 5962–70. https://doi.org/10.1093/emboj/20.21.5962.

Nguyen, Van Son, Badreddine Douzi, Eric Durand, Alain Roussel, Eric Cascales, and Christian Cambillau. 2018. "Towards a Complete Structural Deciphering of Type VI Secretion System." *Current Opinion in Structural Biology* 49 (April): 77–84. https://doi.org/10.1016/j.sbi.2018.01.007.

Offner, S., U. Ziese, G. Wanner, D. Typke, and F. Pfeifer. 1998. "Structural Characteristics of Halobacterial Gas Vesicles." *Microbiology* 144 (Pt 5) (May): 1331–42. https://doi.org/10.1099/00221287-144-5-1331.

Oikonomou, Catherine M., and Grant J. Jensen. 2017. "Cellular Electron Cryotomography: Toward Structural Biology In Situ." *Annual Review of Biochemistry* 86 (June): 873–96. https://doi.org/10.1146/annurev-biochem-061516-044741.

Park, Donghyun, David Chetrit, Bo Hu, Craig R. Roy, and Jun Liu. 2020. "Analysis of Dot/Icm Type IVB Secretion System Subassemblies by Cryoelectron Tomography Reveals Conformational Changes Induced by DotB Binding." *mBio* 11 (1). https://doi.org/10.1128/mBio.03328-19.

Paul, Thomas J., Zachary Hoffmann, Congzhou Wang, Maruda Shanmugasundaram, Jason DeJoannis, Alexander Shekhtman, Igor K. Lednev, Vamsi K. Yadavalli, and Rajeev Prabhakar. 2016. "Structural and Mechanical Properties of Amyloid Beta Fibrils: A Combined Experimental and Theoretical Approach." *Journal of Physical Chemistry Letters* 7 (14): 2758–64. https://doi.org/10.1021/acs.jpclett.6b01066.

Perez-Riverol, Yasset, Attila Csordas, Jingwen Bai, Manuel Bernal-Llinares, Suresh Hewapathirana, Deepti J. Kundu, Avinash Inuganti, et al. 2019. "The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data." *Nucleic Acids Research* 47 (D1): D442–50. https://doi.org/10.1093/nar/gky1106.

Pettersen, Eric F., Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. 2004. "UCSF Chimera--a Visualization System for Exploratory Research and Analysis." *Journal of Computational Chemistry* 25 (13): 1605–12. https://doi.org/10.1002/jcc.20084.

Peukes, Julia, Xiaoli Xiong, Simon Erlendsson, Kun Qu, William Wan, Leslie J. Calder, Oliver Schraidt, et al. 2020. "The Native Structure of the Assembled Matrix Protein 1 of Influenza A Virus." *Nature* 587 (7834): 495–98. https://doi.org/10.1038/s41586-020-2696-8.

Pfeifer, Felicitas. 2012. "Distribution, Formation and Regulation of Gas Vesicles." *Nature Reviews. Microbiology* 10 (10): 705–15. https://doi.org/10.1038/nrmicro2834.

Pintilie, Grigore, and Wah Chiu. 2012. "Comparison of Segger and Other Methods for Segmentation and Rigid-Body Docking of Molecular Components in Cryo-EM Density Maps." *Biopolymers* 97 (9): 742–60. https://doi.org/10.1002/bip.22074.

Pintilie, Grigore D., Junjie Zhang, Thomas D. Goddard, Wah Chiu, and David C. Gossard. 2010. "Quantitative Analysis of Cryo-EM Density Map Segmentation by Watershed and Scale-Space Filtering, and Fitting of Structures by Alignment to Regions." *Journal of Structural Biology* 170 (3): 427–38. https://doi.org/10.1016/j.jsb.2010.03.007.

Prevost, Marie S., and Gabriel Waksman. 2018. "X-Ray Crystal Structures of the Type IVb Secretion System DotB ATPases." *Protein Science*. https://doi.org/10.1002/pro.3439.

Punjani, Ali, John L. Rubinstein, David J. Fleet, and Marcus A. Brubaker. 2017. "cryoSPARC: Algorithms for Rapid Unsupervised Cryo-EM Structure Determination." *Nature Methods* 14 (3): 290–96. https://doi.org/10.1038/nmeth.4169.

Qiu, Jiazhang, and Zhao-Qing Luo. 2017. "Legionella and Coxiella Effectors: Strength in Diversity and Activity." *Nature Reviews. Microbiology* 15 (10): 591–605. https://doi.org/10.1038/nrmicro.2017.67.

Rabut, Claire, Di Wu, Bill Ling, Zhiyang Jin, Dina Malounda, and Mikhail G. Shapiro. 2021. "Ultrafast Amplitude Modulation for Molecular and Hemodynamic Ultrasound Imaging." *Applied Physics Letters* 118 (24): 244102. https://doi.org/10.1063/5.0050807.

Ramsay, Joshua P., Neil R. Williamson, David R. Spring, and George P. C. Salmond. 2011. "A Quorum-Sensing Molecule Acts as a Morphogen Controlling Gas Vesicle Organelle Biogenesis and Adaptive Flotation in an Enterobacterium." *Proceedings of the National Academy of Sciences of the United States of America* 108 (36): 14932–37. https://doi.org/10.1073/pnas.1109169108.

Rapisarda, Chiara, Yassine Cherrak, Romain Kooger, Victoria Schmidt, Riccardo Pellarin, Laureen Logger, Eric Cascales, Martin Pilhofer, Eric Durand, and Rémi Fronzes. 2019. "In Situ and High-Resolution Cryo-EM Structure of a Bacterial Type VI Secretion System Membrane Complex." *The EMBO Journal* 38 (10). https://doi.org/10.15252/embj.2018100886.

Ravikumar, Arjun, Garri A. Arzumanyan, Muaeen K. A. Obadi, Alex A. Javanpour, and Chang C. Liu. 2018. "Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds." *Cell* 175 (7): 1946–57.e13. https://doi.org/10.1016/j.cell.2018.10.021.

Renaud, Jean-Paul, Ashwin Chari, Claudio Ciferri, Wen-Ti Liu, Hervé-William Rémigy, Holger Stark, and Christian Wiesmann. 2018. "Cryo-EM in Drug Discovery: Achievements, Limitations and Prospects." *Nature Reviews. Drug Discovery* 17 (7): 471–92. https://doi.org/10.1038/nrd.2018.77.

Renault, Louis, Bernard Guibert, and Jacqueline Cherfils. 2003. "Structural Snapshots of the Mechanism and Inhibition of a Guanine Nucleotide Exchange Factor." *Nature* 426 (6966): 525–30. https://doi.org/10.1038/nature02197.

Robb, Craig S., Francis E. Nano, and Alisdair B. Boraston. 2012. "The Structure of the Conserved Type Six Secretion Protein TssL (DotU) from Francisella Novicida." *Journal of Molecular Biology* 419 (5): 277–83. https://doi.org/10.1016/j.jmb.2012.04.003.

Ruiz, Federico M., Juvenal Lopez, C. Gastón Ferrara, Elena Santillana, Yanis R. Espinosa, Mario F. Feldman, and Antonio Romero. 2020. "Structural Characterization of TssL from Acinetobacter Baumannii: A Key Component of the Type VI Secretion System." *Journal of Bacteriology* 202 (17). https://doi.org/10.1128/JB.00210-20.

Salahshoor, Hossein, Yuxing Yao, Przemysław Dutka, Nivin N. Nyström, Zhiyang Jin, Ellen Min, Dina Malounda, Grant J. Jensen, Michael Ortiz, and Mikhail G. Shapiro. 2022. "Geometric Effects in Gas Vesicle Buckling under Ultrasound." *Biophysical Journal* 121 (21): 4221–28. https://doi.org/10.1016/j.bpj.2022.09.004.

Sali, A., and T. L. Blundell. 1993. "Comparative Protein Modelling by Satisfaction of Spatial Restraints." *Journal of Molecular Biology* 234 (3): 779–815. https://doi.org/10.1006/jmbi.1993.1626.

Schindelin, Johannes, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, et al. 2012. "Fiji: An Open-Source Platform for Biological-Image Analysis." *Nature Methods* 9 (7): 676–82. https://doi.org/10.1038/nmeth.2019.

Schur, Florian K. M., Martin Obr, Wim J. H. Hagen, William Wan, Arjen J. Jakobi, Joanna M. Kirkpatrick, Carsten Sachse, Hans-Georg Kräusslich, and John A. G. Briggs. 2016. "An Atomic Model of HIV-1 Capsid-SP1 Reveals Structures Regulating Assembly and Maturation." *Science* 353 (6298): 506–8. https://doi.org/10.1126/science.aaf9620.

Segal, Gil, Michal Feldman, and Tal Zusman. 2005. "The Icm/Dot Type-IV Secretion Systems of Legionella Pneumophila and Coxiella Burnetii." *FEMS Microbiology Reviews* 29 (1): 65–81. https://doi.org/10.1016/j.femsre.2004.07.001.

Sexton, Jessica A., Jennifer L. Miller, Aki Yoneda, Thomas E. Kehl-Fie, and Joseph P. Vogel. 2004. "Legionella Pneumophila DotU and IcmF Are Required for Stability of the Dot/Icm Complex." *Infection and Immunity* 72 (10): 5983–92. https://doi.org/10.1128/IAI.72.10.5983-5992.2004.

Sexton, Jessica A., Hye-Jeong Yeo, and Joseph P. Vogel. 2005. "Genetic Analysis of the Legionella Pneumophila DotB ATPase Reveals a Role in Type IV Secretion System Protein Export." *Molecular Microbiology* 57 (1): 70–84. https://doi.org/10.1111/j.1365-2958.2005.04667.x.

Shapiro, Mikhail G., Patrick W. Goodwill, Arkosnato Neogy, Melissa Yin, F. Stuart Foster, David V. Schaffer, and Steven M. Conolly. 2014. "Biogenic Gas Nanostructures as Ultrasonic Molecular Reporters." *Nature Nanotechnology* 9 (4): 311–16. https://doi.org/10.1038/nnano.2014.32.

Shapiro, Mikhail G., R. Matthew Ramirez, Lindsay J. Sperling, George Sun, Jinny Sun, Alexander Pines, David V. Schaffer, and Vikram S. Bajaj. 2014. "Genetically Encoded Reporters for Hyperpolarized Xenon Magnetic Resonance Imaging." *Nature Chemistry* 6 (7): 629–34. https://doi.org/10.1038/nchem.1934.

Sheedlo, Michael J., Jeong Min Chung, Neha Sawhney, Clarissa L. Durie, Timothy L. Cover, Melanie D. Ohi, and D. Borden Lacy. 2020. "Cryo-EM Reveals Species-Specific Components within the Helicobacter Pylori Cag Type IV Secretion System Core Complex." *eLife* 9 (September). https://doi.org/10.7554/eLife.59495.

Sheedlo, Michael J., Clarissa L. Durie, Jeong Min Chung, Louise Chang, Jacquelyn Roberts, Michele Swanson, Dana Borden Lacy, and Melanie D. Ohi. 2021. "Cryo-EM Reveals New Species-Specific Proteins and Symmetry Elements in the Legionella Pneumophila Dot/Icm T4SS." *eLife*. https://doi.org/10.7554/elife.70427.

Sheedlo, Michael J., Melanie D. Ohi, D. Borden Lacy, and Timothy L. Cover. 2022. "Molecular Architecture of Bacterial Type IV Secretion Systems." *PLoS Pathogens* 18 (8): e1010720. https://doi.org/10.1371/journal.ppat.1010720.

Shim, Jongmin, Claude Perdigou, Elizabeth R. Chen, Katia Bertoldi, and Pedro M. Reis. 2012. "Buckling-Induced Encapsulation of Structured Elastic Shells under Pressure." *Proceedings of the National Academy of Sciences of the United States of America* 109 (16): 5978–83. https://doi.org/10.1073/pnas.1115674109.

Sievers, Fabian, Andreas Wilm, David Dineen, Toby J. Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, et al. 2011. "Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega." *Molecular Systems Biology* 7 (1): 539. https://www.embopress.org/doi/abs/10.1038/Msb.2011.75.

Simon, R. D. 1981. "Morphology and Protein Composition of Gas Vesicles from Wild Type and Gas Vacuole Defective Strains of Halobacterium Salinarium Strain 5." *Microbiology* 125 (1): 103–11. https://doi.org/10.1099/00221287-125-1-103.

Stepanenko, Olesya V., Olga V. Stepanenko, Irina M. Kuznetsova, Vladimir N. Uversky, and Konstantin K. Turoverov. 2016. "Peculiarities of the Super-Folder GFP Folding in a Crowded Milieu." *International Journal of Molecular Sciences* 17 (11). https://doi.org/10.3390/ijms17111805.

Tan, Yong Zi, Philip R. Baldwin, Joseph H. Davis, James R. Williamson, Clinton S. Potter, Bridget Carragher, and Dmitry Lyumkis. 2017. "Addressing Preferred Specimen Orientation in Single-Particle Cryo-EM through Tilting." *Nature Methods* 14 (8): 793–96. https://doi.org/10.1038/nmeth.4347.

Tegunov, Dimitry, and Patrick Cramer. 2019. "Real-Time Cryo-Electron Microscopy Data Preprocessing with Warp." *Nature Methods* 16 (11): 1146–52. https://doi.org/10.1038/s41592-019-0580-y.

Tegunov, Dimitry, Liang Xue, Christian Dienemann, Patrick Cramer, and Julia Mahamid. 2021. "Multi-Particle Cryo-EM Refinement with M Visualizes Ribosome-Antibiotic Complex at 3.5 Å in Cells." *Nature Methods* 18 (2): 186–93. https://doi.org/10.1038/s41592-020-01054-7.

Thompson, J. Michael T. 2015. "Advances in Shell Buckling: Theory and Experiments." *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering* 25 (01): 1530001. https://doi.org/10.1142/S0218127415300013.

Timoshenko, Stephen, Sergius Woinowsky-Krieger, and Others. 1959. *Theory of Plates and Shells*. Vol. 2. McGraw-hill New York. https://cds.cern.ch/record/102847/files/0070858209_TOC.pdf.

Tivol, William F., Ariane Briegel, and Grant J. Jensen. 2008. "An Improved Cryogen for Plunge Freezing." *Microscopy and Microanalysis: The Official Journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada* 14 (5): 375–79. https://doi.org/10.1017/S1431927608080781.

VanRheenen, Susan M., Guillaume Duménil, and Ralph R. Isberg. 2004. "IcmF and DotU Are Required for Optimal Effector Translocation and Trafficking of the Legionella Pneumophila Vacuole." *Infection and Immunity* 72 (10): 5972–82. https://doi.org/10.1128/IAI.72.10.5972-5982.2004.

Varadi, Mihaly, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, et al. 2022. "AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models." *Nucleic Acids Research* 50 (D1): D439–44. https://doi.org/10.1093/nar/gkab1061.

Vincent, Carr D., Jonathan R. Friedman, Kwang Cheol Jeong, Emily C. Buford, Jennifer L. Miller, and Joseph P. Vogel. 2006. "Identification of the Core Transmembrane Complex of the Legionella Dot/Icm Type IV Secretion System." *Molecular Microbiology* 62 (5): 1278–91. https://doi.org/10.1111/j.1365-2958.2006.05446.x.

Waaland, J. R., and D. Branton. 1969. "Gas Vacuole Development in a Blue-Green Alga." *Science* 163 (3873): 1339–41. https://doi.org/10.1126/science.163.3873.1339.

Wadhwa, Navish, and Howard C. Berg. 2022. "Bacterial Motility: Machinery and Mechanisms." *Nature Reviews. Microbiology* 20 (3): 161–73. https://doi.org/10.1038/s41579-021-00626-4.

Walker, J. E., and A. E. Walsby. 1983. "Molecular Weight of Gas-Vesicle Protein from the Planktonic Cyanobacterium Anabaena Flos-Aquae and Implications for Structure of the Vesicle." *Biochemical Journal* 209 (3): 809–15. https://doi.org/10.1042/bj2090809.

Walsby, A. E. 1994a. "Gas Vesicles." *Microbiological Reviews* 58 (1): 94–144. https://doi.org/10.1128/mr.58.1.94-144.1994.

Walsby, A. E., and A. Bleything. 1988. "The Dimensions of Cyanobacterial Gas Vesicles in Relation to Their Efficiency in Providing Buoyancy and Withstanding Pressure." *Microbiology* 134 (10): 2635–45. https://doi.org/10.1099/00221287-134-10-2635.

Walsby, A. E., and P. K. Hayes. 1988. "The Minor Cyanobacterial Gas Vesicle Protein, GVPc, Is Attached to the Outer Surface of the Gas Vesicle." *Microbiology* 134 (10): 2647–57. https://doi.org/10.1099/00221287-134-10-2647.

Walsby, A. E., N. P. Revsbech, and D. H. Griffel. 1992. "The Gas Permeability Coefficient of the Cyanobacterial Gas Vesicle Wall." *Journal of General Microbiology* 138 (4): 837–45. https://doi.org/10.1099/00221287-138-4-837.

Walsby, Anthony Edward, and Gordon Elliott Fogg. 1971. "The Pressure Relationships of Gas Vacuoles." *Proceedings of the Royal Society of London. Series B. Biological Sciences* 178 (1052): 301–26. https://doi.org/10.1098/rspb.1971.0067.

Wang, Jing, Maj Brodmann, and Marek Basler. 2019. "Assembly and Subcellular Localization of Bacterial Type VI Secretion Systems." *Annual Review of Microbiology* 73 (September): 621–38. https://doi.org/10.1146/annurev-micro-020518-115420.

Wan, William, Mairi Clarke, Michael J. Norris, Larissa Kolesnikova, Alexander Koehler, Zachary A. Bornholdt, Stephan Becker, Erica Ollmann Saphire, and John Ag Briggs. 2020. "Ebola and Marburg Virus Matrix Layers Are Locally Ordered Assemblies of VP40 Dimers." *eLife* 9 (October). https://doi.org/10.7554/eLife.59225.

Waterhouse, Andrew M., James B. Procter, David M. A. Martin, Michèle Clamp, and Geoffrey J. Barton. 2009. "Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench." *Bioinformatics* 25 (9): 1189–91. https://doi.org/10.1093/bioinformatics/btp033.

Winkler, Hanspeter. 2007. "3D Reconstruction and Processing of Volumetric Data in Cryo-Electron Tomography." *Journal of Structural Biology* 157 (1): 126–37. https://doi.org/10.1016/j.jsb.2006.07.014.

Wittmann, Bruce J., Kadina E. Johnston, Patrick J. Almhjell, and Frances H. Arnold. 2022. "evSeq: Cost-Effective Amplicon Sequencing of Every Variant in a Protein Library." *ACS Synthetic Biology* 11 (3): 1313–24. https://doi.org/10.1021/acssynbio.1c00592.

Wu, Di, Diego Baresch, Colin Cook, Dina Malounda, David Maresca, Maria Paulene Abundo, David Reza Mittelstein, and Mikhail G. Shapiro. 2019. "Genetically Encoded Nanostructures Enable Acoustic Manipulation of Engineered Cells." https://doi.org/10.1101/691105.

Wu, Di, Diego Baresch, Colin Cook, Zhichao Ma, Mengtong Duan, Dina Malounda, David Maresca, et al. 2023. "Biomolecular Actuators for Genetically Selective Acoustic Manipulation of Cells." *Science Advances* 9 (8): eadd9186. https://doi.org/10.1126/sciadv.add9186.

Xie, Guohui, Xuan Du, Hongmiao Hu, Sisi Li, Xiaofeng Cao, Steven E. Jacobsen, and Jiamu Du. 2023. "Structure and Mechanism of the Plant RNA Polymerase V." *Science* 379 (6638): 1209–13. https://doi.org/10.1126/science.adf8231.

Xu, Bo Ying, Ya Nan Dai, Kang Zhou, Yun Tao Liu, Qianqian Sun, Yan Min Ren, Yuxing Chen, and Cong Zhao Zhou. 2014. "Structure of the Gas Vesicle Protein GvpF from the Cyanobacterium Microcystis Aeruginosa." *Acta Crystallographica. Section D, Biological Crystallography* 70 (Pt 11): 3013–22. https://doi.org/10.1107/S1399004714021312.

Xue, Liang, Swantje Lenz, Maria Zimmermann-Kogadeeva, Dimitry Tegunov, Patrick Cramer, Peer Bork, Juri Rappsilber, and Julia Mahamid. 2022. "Visualizing Translation Dynamics at Atomic Detail inside a Bacterial Cell." *Nature* 610 (7930): 205–11. https://doi.org/10.1038/s41586-022-05255-2.

Yin, Meng, Zhaofeng Yan, and Xueming Li. 2019. "Architecture of Type VI Secretion System Membrane Core Complex." *Cell Research* 29 (3): 251–53. https://doi.org/10.1038/s41422-018-0130-7.

Zhang, Shuai, An Huang, Avinoam Bar-Zion, Jiaying Wang, Oscar Vazquez Mena, Mikhail G. Shapiro, and James Friend. 2020. "The Vibration Behavior of Sub-micrometer Gas Vesicles in Response to Acoustic Excitation Determined via Laser Doppler Vibrometry." *Advanced Functional Materials* 30 (13): 2000239. https://doi.org/10.1002/adfm.202000239.

Zheng, Shawn Q., Eugene Palovcak, Jean-Paul Armache, Kliment A. Verba, Yifan Cheng, and David A. Agard. 2017. "MotionCor2: Anisotropic Correction of Beam-Induced Motion for Improved Cryo-Electron Microscopy." *Nature Methods* 14 (4): 331–32. https://doi.org/10.1038/nmeth.4193.

Zivanov, Jasenko, Takanori Nakane, Björn O. Forsberg, Dari Kimanius, Wim Jh Hagen, Erik Lindahl, and Sjors Hw Scheres. 2018. "New Tools for Automated High-Resolution Cryo-EM Structure Determination in RELION-3." *eLife* 7 (November). https://doi.org/10.7554/eLife.42166.

Zivanov, Jasenko, Joaquín Otón, Zunlong Ke, Andriko von Kügelgen, Euan Pyle, Kun Qu, Dustin Morado, et al. 2022. "A Bayesian Approach to Single-Particle Electron Cryo-Tomography in RELION-4.0." *eLife* 11 (December). https://doi.org/10.7554/eLife.83724.

Zoued, Abdelrahim, Chloé J. Cassaro, Eric Durand, Badreddine Douzi, Alexandre P. España, Christian Cambillau, Laure Journet, and Eric Cascales. 2016. "Structure-Function Analysis of the TssL Cytoplasmic Domain Reveals a New Interaction between the Type VI Secretion Baseplate and Membrane Complexes." *Journal of Molecular Biology* 428 (22): 4413–23. https://doi.org/10.1016/j.jmb.2016.08.030.