

Visualizing Small Proteins
with the cryoEM Platform
and
The Structure of the
Vibrio cholerae
Type IV Competence Pilus
Secretin PilQ

Thesis by
Sara Jean Weaver

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

The logo for the California Institute of Technology (Caltech), featuring the word "Caltech" in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2020
(Defended November 22, 2019)

© 2020

Sara Jean Weaver
ORCID: 0000-0001-7753-6215

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Grant Jensen, for his support. He constantly inspires me and has provided me immense freedom to develop as a scientist. Grant has cultivated a lab environment stocked with the most intelligent and the kindest people you could ask for in science. The lab is a true testament to his leadership, vision, and worldview.

I am extremely grateful to Doug Rees, Mikhail Shapiro, and Rebecca Voorhees for guidance, support, and inspiration. My committee meetings were full of excellent discussion, great advice, and patience. Outside of our yearly meetings, each member of my committee was always willing me to meet. In particular, Rebecca Voorhees has been a true role model in everything from experimental work to navigating the publishing process.

I would not be here without the innumerable hours of mentorship a variety of people provided. To adequately list them all, my dissertation would double in length. I am constantly floored by the amount of time and effort people in my life have given me. I will highlight a few.

Phil Bednarek and Jack Palmer went out of their way to encourage me in high school science classes. Colin Jefcoate welcomed me into his lab after my first year of college when I could barely pipet. Michele Larsen took me under her wing in Colin's lab. At Barnard, Jennifer Mansfield and John Magyar let me loose with my own research projects. John taught me about creativity and resilience in science.

After college, working for Jay Yang was truly invaluable. Jay is an incredible scientist and gave me freedom that most technicians could never dream of. Uyen Chu taught me to channel my ideas into concrete, careful experiments.

At Caltech, Matthew Swulius mentored me for my Jensen lab rotation project, and somehow talked Grant into taking me as a student. After setting me free, he continued to advise me on experimental and existential topics. Matt has an uncanny ability to see

problems clearly and has taken some of the most beautiful tomograms in existence. David Ortega has spent numerous hours discussing science and life with me. His energy and passion constantly revitalized me. The North Lab was a key support network in the Jensen lab. Catherine Oikonomou is too wonderful to describe in words. Her advice on all topics is a key factor in the completion of this dissertation. Lauren Ann Metskas's tenacity and brilliance are a constant inspiration. Naima Sharaf's curiosity and passion motivated me to push and carve my own path forward. Her clever experiments and enthusiasm for all aspects of science gave me hope and new tools to go after my ideas. I want to thank everyone on Broad 3 for support, advice, and entertainment.

This alphabetical list highlights some of the many people I would like to thank: Christopher Barnes, Annet Blom, Catie Blunt, Emily Blythe, Ariane Briegel, Marisa Buzzeo, Patricia B. Campbell, Stephen Carter, Yi-Wei Chang, Sally Chapman, Songye Chen, Georges Chreifi, Jane Ding, Erin-Kate Escobar, Chengcheng Fan, Welison Floriano, Debnath Ghosal, Danielle Grotjahn, Samuel Ho, Kathryn Huey-Tubman, Claudia Jette, Andrew Jewett, Mohammed Kaplan, Ki Woo Kim, Rachel Krueger, Olympia Jebejian, Julia Jensen, Mark Ladinsky, Jeffery Lai, Allen Lee, Yuxi Liu, Wendy Lopata, Siobhan MacArdle, Shrawan Mageswaran, Ailiena Maggiolo, Karin Mallard, Andrey Malyutin, Alasdair McDowall, Marta Murphy, Lam Nguyen, Phong Nguyen, William Nicolas, Ariana Peck, Phoebe Ray, Lee Rettberg, Alison Ross, Leah Sabbath, Matthew Sazinsky, Andrew Schacht, Carrie Shaffer, Rebekah Silva, Eileen Smith, Shannon Stone, Poorna Subramanian, Meaghan Sullivan, Annelise Thompson, Elitza Tocheva, Agnes Tong, Steven Wang, Belinda Wenke, Wei Yang, Qing Yao, Byungkuk Yoo, Lujia Zhang, Xianjun Zhang, and Wei Zhao.

I would be nowhere without Travis Alvarez, Lauren Benton, Jose Delgado, Frances Diep, Briana Doering, Kevin Gould, Jaclyn Henderson, Evelyn Jagoda, Julia Kennedy, Daniel Miranda, and Hayley Peterson. G(TB)².

Finally, I cannot begin to express my thanks to my family, who are my biggest champions and supporters. Jonathan Jager, Teresa Weaver, Michael Weaver, Teresa Grooms, Amy

Rae Weaver, Ruby Weaver, Niles Orville Weaver, Steven Jager, Janice Jager, Robbie Jager, Barbara Jager, Otis Jager, Nina Jager, Barb and Bob Weaver, Jim and Lois Statz, Magdalena Jager Z'L, Thomas Weaver, Rachel Drysdale, Mary Weaver-Klees, Jeff Klees, and everyone else.

Thank you.

ABSTRACT

Solving protein structures by single-particle cryoelectron microscopy (cryo-EM) has become a crucial tool in structural biology. While exciting progress is being made toward the visualization of small macromolecules, the median protein size in both eukaryotes and bacteria is still beyond the reach of cryo-EM. To overcome this problem, we implemented a platform strategy in which a small protein target was rigidly attached to a large, symmetric base via a selectable adapter. Of our seven designs, the best construct used a designed ankyrin repeat protein (DARPin) rigidly fused to tetrameric rabbit muscle aldolase through a helical linker. The DARPins retained their ability to bind their target: GFP. We solved the structure of this complex to 3.0 Å resolution overall, with 5-8 Å resolution in the GFP region. As flexibility in the DARPin position limited the overall resolution of the target, we describe strategies to rigidify this element.

Natural competence is the process by which bacteria take up genetic material from their environment and integrate it into their genome using homologous recombination. In *Vibrio cholerae*, the Type IV competence pilus is thought to mediate DNA uptake by binding DNA and retracting back toward the cell. How the DNA enters the periplasm is unclear. One hypothesis suggests that the DNA-bound Type IV competence pilus retracts completely so that the DNA would pass through the outer membrane secretin pore (PilQ). PilQ is a 870 kDa outer membrane pore with C14 symmetry. Here, we purify the *V. cholerae* PilQ secretin from *V. cholerae* cells in amphipols for single particle cryogenic electron microscopy (cryoEM). We solve the structure to 3.0 Å and provide insight on the channel DNA may traverse through during uptake.

PUBLISHED CONTENT AND CONTRIBUTIONS

Yao, Q.*, Weaver, S.J.*, Mock, J.Y., Jensen, G.J., Fusion of DARPin to Aldolase Enables Visualization of Small Protein by Cryo-EM, *Structure*, 27, 1-8 (2019).

- bioRxiv Preprint: [https://doi.org/10.1101/455063\(2018\)](https://doi.org/10.1101/455063(2018))
- Conceptualization, Q.Y., S.J.W., G.J.J.; Methodology, Q.Y., S.J.W., J.Y.M; Investigation, Q.Y., S.J.W., J.Y.M; Data Curation, S.J.W.; Writing – Original Draft, S.J.W.; Writing – Review & Editing, S.J.W., Q.Y., G.J.J.; Visualization, S.J.W., Q.Y.; Supervision, G.J.J., Funding Acquisition, G.J.J.

Sara J. Weaver¹, Matthew Sazinsky², Triana Dalia³, Ankur Dalia³, Grant J. Jensen⁴. CryoEM Structure of The *Vibrio cholerae* Type IV competence pilus Secretin PilQ. *In preparation*.

- S.J.W. conceptualized the project, expressed and purified the protein, prepared samples for cryoEM, collected cryoEM data, processed cryoEM data, interpreted results, designed figures, and wrote the paper.
- M.S. purified protein, assisted with cryoEM sample prep and data collection, built the atomic model, and provided feedback on the paper.
- T.D. and A.D. conceptualized the project, obtained funding, engineered the *V. cholerae* constructs, performed transformation assays, and provided feedback on the paper.
- G.J.J. conceptualized the project, obtained funding, and provided feedback on the paper.

TABLE OF CONTENTS

Table of Contents

ACKNOWLEDGEMENTS	III
ABSTRACT	VI
PUBLISHED CONTENT AND CONTRIBUTIONS	VII
TABLE OF CONTENTS	VIII
LIST OF ILLUSTRATIONS AND/OR TABLES.....	X
CHAPTER 1:.....	1
INTRODUCTION.....	1
<i>Context.....</i>	<i>1</i>
<i>Impact of structural biology.....</i>	<i>1</i>
<i>Why cryoEM?</i>	<i>2</i>
SINGLE PARTICLE ANALYSIS.....	4
<i>Quality of single particle cryoEM maps</i>	<i>6</i>
LIMITATIONS OF SINGLE PARTICLE CRYOEM	7
CONTEXT AND SIGNIFICANCE.....	9
<i>A platform to simplify small protein structure determination by cryoEM.....</i>	<i>10</i>
<i>CryoEM of a bacterial outer membrane protein</i>	<i>11</i>
CHAPTER 2:	12
FUSION OF DARPIN TO ALDOLASE ENABLES VISUALIZATION OF SMALL PROTEIN BY CRYOEM.....	12
SUMMARY	13
<i>Introduction</i>	<i>14</i>
RESULTS	16
<i>Platform strategy and design</i>	<i>16</i>
<i>Screening base candidates</i>	<i>20</i>
<i>CryoEM analysis of the GFP:DARPin-aldolase complex</i>	<i>23</i>
<i>Potential sources of heterogeneity.....</i>	<i>27</i>
<i>DARPin position caused conformational heterogeneity</i>	<i>27</i>
DISCUSSION.....	30
ACKNOWLEDGEMENTS	31
AUTHOR CONTRIBUTIONS	31
DECLARATION OF INTERESTS	32
STAR METHODS.....	32
<i>Contact for Reagent and Resource Sharing</i>	<i>32</i>
<i>Experimental Model and Subject Details</i>	<i>32</i>

<i>Method Details</i>	32
<i>Multimedia files</i>	37
CHAPTER 3:	39
CRYOEM STRUCTURE OF THE VIBRIO CHOLERAE TYPE IV COMPETENCE PILUS SECRETIN PILQ	39
ABSTRACT	39
INTRODUCTION.....	39
RESULTS AND DISCUSSION.....	42
<i>Purification of V. cholerae PilQ in amphipol</i>	42
<i>Single particle cryoEM of the Type IV competence pilus secretin PilQ</i>	44
<i>The structure of VcPilQ demonstrates similarities to other secretins</i>	49
<i>The N0 domain in VcPilQ agrees with previous structures</i>	52
<i>VcPilQ uses a novel helical coil to transition into the N3 domain</i>	54
<i>The putative outer membrane region of VcPilQ is thicker than T2SS secretins</i>	58
<i>Why is the VcPilQ putative outer membrane region taller than in T2SS secretins?</i>	60
<i>Why don't any of the putative outer membrane regions span the entire 4 nm outer membrane?</i>	61
<i>Will PilQ accommodate the Type IV competence pilus?</i>	64
CONCLUSION.....	68
<i>Cloning of the His-tagged secretin constructs</i>	80
<i>Transformation assay</i>	80
<i>Expression</i>	80
<i>Purification</i>	80
<i>Electron microscopy</i>	81
<i>Image Processing</i>	81
<i>Model Building and Refinement</i>	82
<i>Mass Spectrometry</i>	82
<i>Protein sequence alignment</i>	83
ACKNOWLEDGEMENTS	84
BIBLIOGRAPHY	85

LIST OF ILLUSTRATIONS AND/OR TABLES

<i>Number</i>	<i>Page</i>
Figure 1 The structural biology continuum.....	2
Figure 2 Single Particle cryoEM Workflow.....	3
Figure 3 Table of protein features observed at different resolutions.....	7
Figure 4 Conformational heterogeneity averages out.....	9
Figure 5 Graph of the number of PDB entries released per year using electron microscopy.....	9
Figure 6 Graphical Abstract For Chapter 2.....	13
Figure 7 The Design of the platform.....	17
Figure 8 Attempted platform designs and outcomes, related to Figure 7.....	18
Figure 9 cryoEM data for the GFP:DARPin-aldolase complex, related to Figure 8....	22
Figure 10 Major steps of the last cycle of cryoEM data processing, related to STAR Methods.....	25
Figure 11 The cryoEM structure of the DARPin-aldolase platform in complex with the target GFP.....	26
Figure 12 Symmetry expanded 3D classification of the GFP:DARPin region of the density.....	29
Figure 13 Protein domain organization in the Type IV Competence Pilus (T4P), the Type II Secretion system (T2SS) and the Type III Secretion System (T3SS)...	40
Figure 14 Structures of Type IV Pilus Secretins.....	41
Figure 15 His-tagged PilQ <i>V. cholerae</i> transformation assay.....	43
Figure 16 Protein electrophoresis of VcPilQ in amphipol.....	44
Figure 17 Example micrograph.....	45
Figure 18 VcPilQ 2D classes.....	46
Figure 19 VcPilQ 2D class demonstrates C14 symmetry.....	46
Figure 20 CryoEM density of VcPilQ.....	47
Figure 21 2D classification demonstrates hazy AMIN domain region.....	48
Figure 22 Atomic model of VcPilQ.....	50
Figure 23 RMSD of VcPilQ versus published secretin structures.....	51
Figure 24 Alignment of VcPilQ N0 domain compared to others.....	53
Figure 25 Secretin monomer atomic model.....	55
Figure 26 Conservation analysis of <i>V. cholerae</i> PilQ focusing on different protein domains.....	56
Figure 27 Electrostatic potential surface of VcPilQ.....	58
Figure 28 VcPilQ micelle density with putative outer membrane diameter labeled in Å.....	60
Figure 29 Micelle density of VcPilQ with ConSurf protein structure.....	61

Figure 30 VcPilQ fit into sub-tomogram averaging of <i>M. xanthus</i> Type IVa pilus machinery.	63
Figure 31 Approximate VcPilQ inner channel diameter distances.	64
Figure 32 Type IVa pili cryoEM structures fit into VcPilQ.	66
Figure 33 VcPilQ putative gate hinge distance.	67
Figure 34 <i>V. cholerae</i> PilQ sequence with mass spectroscopy results.	69
Figure 35 3D Fourier Shell Correlation Results.	70
Figure 36 CryoEM density around VcPilQ beta strands.	70
Figure 37 Local resolution of the VcPilQ structure.	71
Figure 38 ConSurf residue conservation analysis on VcPilQ residues 1 to 252.	72
Figure 39 ConSurf residue conservation analysis on VcPilQ residues 253-571.	73
Figure 40 cryoEM image processing summary flow chart.	74
Figure 41 Ribbon diagram of VcPilQ beta sheets.	76
Figure 42 VcPilQ is a double beta barrel.	77
Figure 43 Electrostatic potential of secretins.	78
Figure 44 Homology modeling of <i>V. cholerae</i> PilQ.	79

Chapter 1:

INTRODUCTION

Context

In the midst of my doctoral research, the resolution revolution took cryoEM from medium to high resolution. My research focused on two main projects that resulted in papers. Each project focused on single particle cryoEM and developing methods to use it to solve a variety of types of protein structures. In the second chapter, I addressed the problem of solving small protein structures by cryoEM using a platform system. In the third chapter, I investigated the structure of a multimeric outer membrane channel in *Vibrio cholerae* that is involved in DNA uptake and antibiotic resistance.

Impact of structural biology

Understanding protein structure is fundamental to understanding and manipulating protein function. Additionally, from a basic science perspective, studying the interactions of biomolecules within cells is key to understanding physiological phenomena. Structural biology aims to solve the structures of biologically relevant materials, from small biomolecules to the ultrastructure of cells and tissue, in order to understand how these macromolecules function (**Figure 1**). This data is integral to the future of cell biology, drug design, and protein engineering.

Structural biology targets range from ångstrom-scale details in a glycan on a protein to micron-scale details in the organization of cells within tissue. X-ray crystallography of protein structures can yield sub-ångstrom-level information about a protein of interest. All-atom molecular dynamics simulations and Nuclear Magnetic Resonance (NMR) spectroscopy can complement that crystal structure to lend insight into how that protein might move and flex inside an aqueous cell. Cryogenic electron microscopy (cryoEM) single particle analysis (SPA) can capture snapshots of large proteins in a near native state to fill out possible conformations. To obtain a snapshot of proteins in a cellular context,

cryogenic electron tomography (cryoET) can be used to provide nanometer-scale resolution 3D reconstructions of cells or tissues. To study the dynamics of a given protein, light and fluorescence microscopy can track labeled proteins within a living cell over time. To fully appreciate the complexity of biology, these techniques are combined to complement the strengths and weaknesses of each technique.

Proteins				Whole Cells		
Polypeptides	Small proteins and domains	Large proteins and complexes	Multi-protein reactions	Cell sections	Whole cells	Tissues
Molecular dynamics simulations	X-ray crystallography & NMR spectroscopy	cryoEM single particle analysis & X-ray crystallography	Cryoelectron tomography	Cryoelectron tomography & light microscopy	Flourescence light microscopy	Structurally and spatially explicit cell modeling

Figure 1 The structural biology continuum.

Structural biology uses different techniques to interrogate targets of different sizes (top row). Small (left) to large (right) systems. Adapted from Grant Jensen.

Why cryoEM?

Cryogenic electron microscopy (cryoEM) is a microscopy technique in which thin (<500 nm thick) frozen samples are imaged in an electron microscope. Purified protein can be analyzed by single particle cryoEM to solve protein structures. Cells can be analyzed using cryogenic electron tomography (cryoET).

In both single particle cryoEM and cryoET, the sample is rapidly plunged into cooled liquid ethane, which allows amorphous, vitreous ice to form around and within the sample, rather than crystalline ice. This allows samples to be visualized in a frozen-hydrated, near-native state within the microscope without the damage the ice crystallization process produces (Y. Cheng et al. 2015). This frozen-hydrated, near-native state is an advantage of cryoEM. Additionally, in the moments before plunge freezing a cryoEM sample, the proteins are tumbling around in solution. As the cryoEM sample is frozen rapidly, proteins do not have much time to rearrange as the sample solidifies. Thus, the protein of interest is captured on a cryoEM grid in a variety of conformations and orientations. These can be analyzed by single particle cryoEM to show a variety of possible conformations in the

native protein. Achieving a similar result with x-ray crystallography would require the protein to crystallize in several different conformations, and each conformation would require an additional dataset.

Ten years ago, single particle cryoEM was consistently referred to as “blob-ology” because the structures solved using it were on the nanometer scale. Today, single particle cryoEM can now attain resolutions sufficient to solve a protein structure *de novo*. With recent technological advances, cryoEM has consistently demonstrated its ability to solve protein structures to near-atomic resolution and is now a strong competitor to x-ray crystallography for large proteins. For smaller proteins, cryoEM is possible, but not trivial. While cryoEM SPA is a developing technique and aspects like sample preparation, data analysis, and model building are still being optimized, cryoEM has the potential to become the most high-throughput structural biology technique.

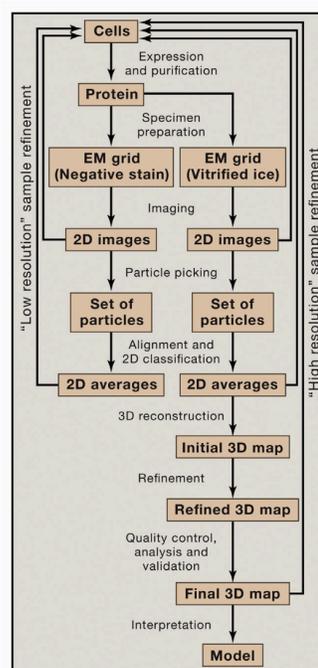


Figure 2 Single Particle cryoEM Workflow.

The workflow of a cryoEM single particle analysis project. Reproduced from (Y. Cheng et al. 2015)

Single particle analysis

A high-resolution single particle cryoEM project begins by analyzing sample quality using negatively stained samples (**Figure 2**). For a negative stain experiment, purified protein is applied to an electron microscopy grid and stained with a heavy metal stain, such as uranyl acetate or osmium tetroxide(Orlova and Saibil 2011). Two-dimensional (2D) projection images of the sample are taken at room temperature in a transmission electron microscope (TEM). Negative stain TEM allows the outline of particles to be seen, but no information about the internal structure of the particle can be obtained. It is used to assess the homogeneity of the sample, its distribution across a grid, and the ideal concentration. Depending on the resolution required to answer a given biological question, a 3D reconstruction of the sample may be prepared using negative stain data.

Whether or not a sample behaves well for negative stain TEM, cryoEM can be performed. A purified protein solution is plunge-frozen in a thin sheet of ice over a thin holey carbon substrate, supported by a small metal grid, which can be inserted into the vacuum of the microscope. For cryoEM 2D projection images are recorded as movies with a fast frame rate. When the electron beam irradiates the sample, beam-induced specimen movement is observed as translations and rotations of the sample(Y. Cheng et al. 2015). Recent studies using the movie mode of direct detectors have demonstrated the degree to which particles translate and rotate during a typical SPA electron dose exposure(Campbell et al. 2012). A typical single particle cryoEM experiment might use a total electron fluence of 60 electrons/Å² and subdivide that exposure into movie frames receiving 1-2 electron/Å² per frame so that these movements can be tracked and corrected during data processing.

In addition to correction for this motion, images are also corrected for the Contrast Transfer Function (CTF) of the microscope. In a real electron microscope, scattered electrons with different wavelengths don't contribute equally to the final image. This is particularly true at high resolution, where contrast is reduced due to this loss of information(Rosenthal and Henderson 2003). The CTF of a given microscope describes how the microscope's aberrations attenuate the resulting image.

To solve a structure by EM, first particles are identified in each 2D projection image and extracted. These extracted particles are aligned and sorted into classes. Ideally each class represents a 2D projection image of the particle taken from a different angle. If there is conformational or stoichiometric heterogeneity in the sample, some classes may represent it. Additionally, contamination is often sorted into its own class. Next, a class average is produced from the aligned particles in a given class.

Once class averages are obtained, the process of 3D reconstruction can begin. The projection theorem states that the Fourier transform of a 2D projection image of an object represents a central slice in the 3D Fourier transform of the object. Thus, the 2D Fourier transforms of the class averages of the particle represent different central slices of the 3D Fourier transform of the particle. The 2D Fourier transform of each class average is compared with the other class averages to find the common line they share. This allows one to align the 2D Fourier transforms in 3D reciprocal space. The space not populated by a class average is interpolated and the entire 3D volume is treated with an inverse Fourier transform to yield an initial 3D reconstruction. This reconstruction is projected to all angles to create 2D projection images representing class averages. The particle sub-volumes are reclassified according to these new class averages and the process is iterated to produce a high quality reconstruction of the particle.

After obtaining a 3D reconstruction of the specimen, 3D classification can be used to sort out heterogeneity within the data. In 3D classification (sometimes called heterogeneous refinement or 3D multi-reference alignment), K initial 3D references are selected and each particle is compared to each reference to find the best match. The initial references do not need to be identical. K -means classification algorithms are typically used for 3D classification (Y. Cheng et al. 2015). This step removes damaged particles, but can also identify different conformations of the specimen. Variations on 3D classification use masks to focus on different regions of the density.

Quality of single particle cryoEM maps

CryoEM reconstructions of proteins are judged by a variety of criteria, including resolution and map to model agreement. In single particle electron microscopy (cryoEM or negative stain experiments), resolution is calculated based on the Fourier Shell Correlation (FSC). During 3D reconstruction, the dataset is divided into two and each half is refined independently. The two maps are compared in Fourier space to determine to which spatial frequency ($1/\text{\AA}$) the two half sets agree. When the normalized correlation between the two half sets drops below a certain criterion (0.5 and 0.143 are used in the field, with 0.143 being more common), that cutoff point is referred to as the global resolution. However, most cryoEM maps have anisotropic resolution, so one number is insufficient to describe the overall quality of the map. The local resolution of each voxel in the map can be calculated to demonstrate which regions of the density are higher resolution (Kucukelbir, Sigworth, and Tagare 2014b). Additionally, the 3D FSC can be used to examine how preferred orientation may play into the anisotropic resolution (Tan et al. 2017).

CryoEM images are inherently noisy, due to relatively low contrast of proteins in ice, and the total dose limitations dictated by radiation damage. This is challenging at high resolution, where the CTF already dictates contrast will be lower. To interpret cryoEM maps, scientists perform post processing (including sharpening and masking) after a 3D reconstruction to distinguish signal from noise in the cryoEM density map. Post processing can mitigate some of this contrast loss by using temperature-factor (B-factor) sharpening (Rosenthal and Henderson 2003). In an unfiltered cryoEM density map, the ratio of low and high resolution structure factors is biased toward low resolution structure factors as high resolution structure factors have reduced contrast. Sharpening the map restores the balance of low resolution and high resolution structure factors using scaling (Rosenthal and Henderson 2003).

Feature	Required resolution
Proteins	
α-helix	9Å
β-helix	4Å
'random' main chain	3.7Å
aromatic side-chains	3.5Å
shaped bulbs of density for small side chains	3.2Å
interpretable conformations for side chains	2.9Å
density for main-chain carbonyl groups, identifying plane of peptide bond	2.7Å
ordered water molecules	2.7Å
dimple at center of aromatic ring	2.4Å
correct stereochemistry at C ^β of isoleucine	2.2Å
puckering of proline	2.0Å
resolving individual atoms	1.5Å

Figure 3 Table of protein features observed at different resolutions.

Based on a table in (Blow 2002).

Since global resolution can be misleading, some scientists prefer to judge the quality of a map based on examining features in the cryoEM density. For example, at certain resolution thresholds secondary structure, like alpha helices (9 Å) and beta sheets (4 Å) can be distinguished (**Figure 3**)(Blow 2002). For many projects, a major goal involves obtaining better than 3 Å resolution. At 3 Å resolution, side chains can be distinguished in the cryoEM density, so the protein chain can be traced *de novo* (without fitting in a previously solved or predicted atomic model). For proteins with novel structures, this threshold is essential, since a similar protein from the PDB can't be docked into the structure to interpret it.

Limitations of single particle cryoEM

Contrast is inherently low in cryoEM experiments because the density of the protein is similar to that of the ice surrounding it(Richard Henderson 1992). Higher electron fluence can increase contrast, but must be balanced with the resulting radiation damage (e.g. breaking covalent bonds, releasing gas)(R Henderson 1995). During an exposure,

interactions between electrons and the sample lead to beam-induced specimen movement. This blurring reduces our ability to precisely align each particle. Large, densely folded proteins are easier to solve by single particle cryoEM because they produce higher contrast per unit area than small or extended conformation proteins, so it is easier to align the particles.

Another challenge in single particle cryoEM is the non-specific interaction of particles with each other (aggregation) and with the air-water interface(Taylor and Glaeser 2008). It is advantageous to image proteins in the thinnest layer of ice possible that still provides enough orientations for reconstruction. Sometimes thin ice can exclude proteins, or will favor some conformations over others. This can lead to a preferred orientation problem that makes it difficult to solve the protein structure. However, simulations have demonstrated that proteins hit the air-water interface on the millisecond timescale when a thin film of liquid is formed, like during plunge-freezing(Taylor and Glaeser 2008). Forces at the air-water interface likely damage or completely unfold the protein. In some cases, the problem can be circumvented by applying a substrate, like a thin film of amorphous carbon, graphene, or graphene oxide, to the grid(Taylor and Glaeser 2008; D’Imprima et al. 2019; Y. Cheng et al. 2015). Some proteins preferentially associate with the substrate rather than the air-water interface. This interaction could also attenuate the protein, but in some cases it is the only way to solve the structure. Regardless, many cryoEM structures have been solved without addressing the air-water interface.

Strategies to address conformational heterogeneity in single particle cryoEM are needed. Particles are tumbling randomly in a physiologically relevant solution at the time of plunge-freezing, so conformational heterogeneity is expected. Conformational heterogeneity is more common in proteins with flexible regions. If a protein is rigid, it adopts fewer specific conformations on the grid and is easier to align and reconstruct. If the protein is flexible and adopts a continuous gradient of conformations, it will be a great challenge to tease apart the nuances of the movement (**Figure 3**). If the flexible regions are not well aligned they will either cancel out or appear in the final image as a blur.

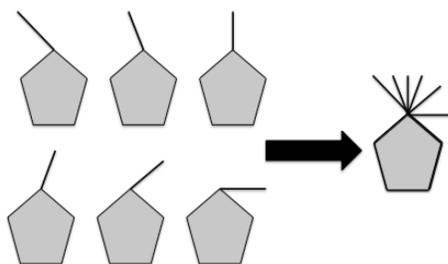


Figure 4 Conformational heterogeneity averages out

The illustration highlights the difficulty of aligning particles where one part exhibits conformational heterogeneity. In the averaging process, the flexible black tail of the pentagon will be canceled out due to the heterogeneity.

Context and Significance

Over the course of my graduate studies (2014-2019), the reach of cryogenic electron microscopy (cryoEM) has exploded across structural biology. In 2017, Jacques Dubochet, Joachim Frank and Richard Henderson won the Chemistry Nobel Prize "for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution" ("The Nobel Prize in Chemistry 2017" 2017). A survey of the number of PDB entries released each year that use the technique "electron microscopy" reveals a strong upward trend (**Figure 5**) ("RCSB PDB: Homepage" n.d.; Berman et al. 2000).

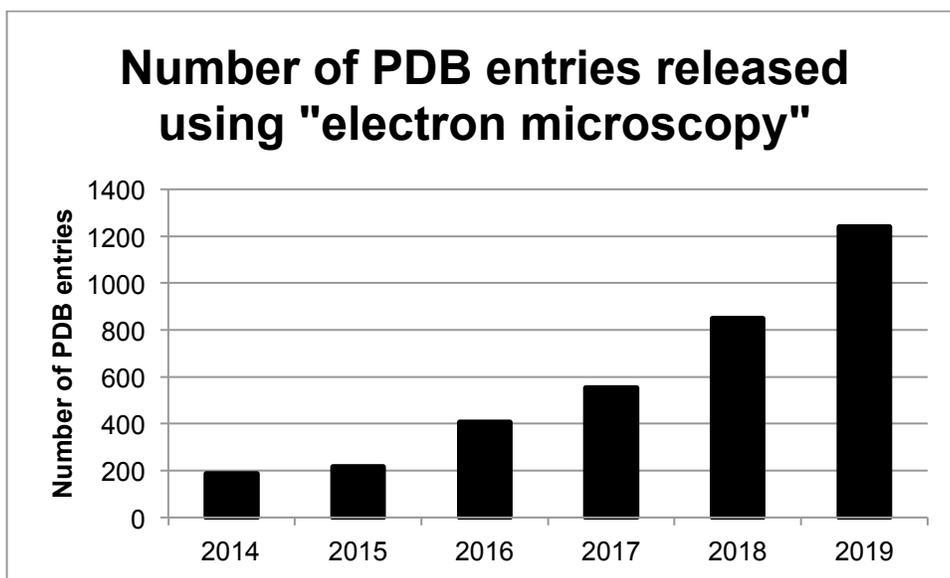


Figure 5 Graph of the number of PDB entries released per year using electron microscopy. As of 2019/11/07, 1,240 PDB entries using electron microscopy have been released.

A platform to simplify small protein structure determination by cryoEM

When the platform project (described in Chapter 2) was initiated in 2015, the highest resolution cryoEM structure solved was the 2.2 Å structure of beta-galactosidase, a 465 kDa tetramer, in complex with an inhibitor (Bartesaghi et al. 2015), and the smallest protein structure solved to better than 3.5 Å resolution (the threshold where a protein chain can be mapped *de novo* into cryoEM density) was the 170 kDa gamma-secretase complex (Bai et al. 2015). The average protein size in eukaryotes is ~50 kDa, so this left a huge swath of biology that cryoEM could not access (Brocchieri and Karlin 2005). I aimed to develop a high-throughput system where small proteins could be reversibly attached to a larger platform protein to facilitate solving the structure of the small protein. Our hope was that the platform would reduce the complexity and cost of solving structures, and make cryoEM more accessible at different universities. After many tests, our most successful platform (DARPin-aldolase) was able to resolve the structure of our target protein (Green Fluorescent Protein, 27 kDa) to medium resolution (5 to 8 Å).

In the years since the initiation of my platform project, cryoEM sample preparation and algorithms have improved. At this time, the smallest reported protein to be solved to better than 3.5 Å is streptavidin (52 kDa, 3.2 Å) (Fan et al. 2019), while the highest resolution structure is of mouse heavy-chain apoferritin (1.62 Å) (Danev, Yanagisawa, and Kikkawa 2019). The theoretical size limit for single particle cryoEM was predicted to be around 38 kDa (R Henderson 1995). While scientists are pushing closer to that goal (for example, the ~6 Å resolution structure of the 43 kDa isolable kinase domain of protein kinase A) (Herzik, Wu, and Lander 2019), solving structures of proteins below 200 kDa remains difficult. Some small proteins can now be solved by cryoEM experiments with incredibly thin ice using the most sophisticated instruments and software. However, these experiments still require a great deal of expertise and expense.

Since the submission of my work, other platform strategies have been reported. Liu *et. al* reported the structure of GFP to 3.8 Å by a DARPin-based platform system (Liu, Huynh, and Yeates 2019). Uchański *et. al* reported a modified nanobody (the ‘megabody’) that

could improve the preferred orientation of a membrane protein complex and facilitate its solution to high resolution(Uchański et al. 2019). Currently, some platform systems work for certain proteins, but not others. In the future, it's likely that an entire toolbox of platforms will be developed and one will simply try everything on the list until something works.

CryoEM of a bacterial outer membrane protein

After completing the platform project, I moved my attention to solving a membrane protein structure from protein expressed in *Vibrio cholerae*. The secretin PilQ acts as a channel for the Type IV competence pilus to exit *V. cholerae* cells, pick up environmental DNA, and bring it into the cell for use. I chose to purify PilQ directly from *V. cholerae*, rather than via recombinant expression in *E. coli*, so that I could retain native protein folding. I solved the first high resolution structure of a Type IV competence pilus secretin and provided insight into the mechanism of DNA uptake during natural competence.

Chapter 2:

FUSION OF DARPIN TO ALDOLASE ENABLES VISUALIZATION OF SMALL PROTEIN BY CRYOEM

Qing Yao^{1,4}, Sara J. Weaver^{2,4}, Jee-Young Mock² and Grant J. Jensen^{1,3,5*}

¹Division of Biology and Biological Engineering, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125

²Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125

³Howard Hughes Medical Institute, 1200 E. California Blvd., Pasadena, CA 91125

⁴These authors contributed equally to this work

⁵Lead Contact

*Corresponding author

Email: jensen@caltech.edu (GJJ)

Yao, Q.*, Weaver, S.J.*, Mock, J.Y., Jensen, G.J., Fusion of DARPIn to Aldolase Enables

Visualization of Small Protein by Cryo-EM, *Structure*, 27, 1-8 (2019).

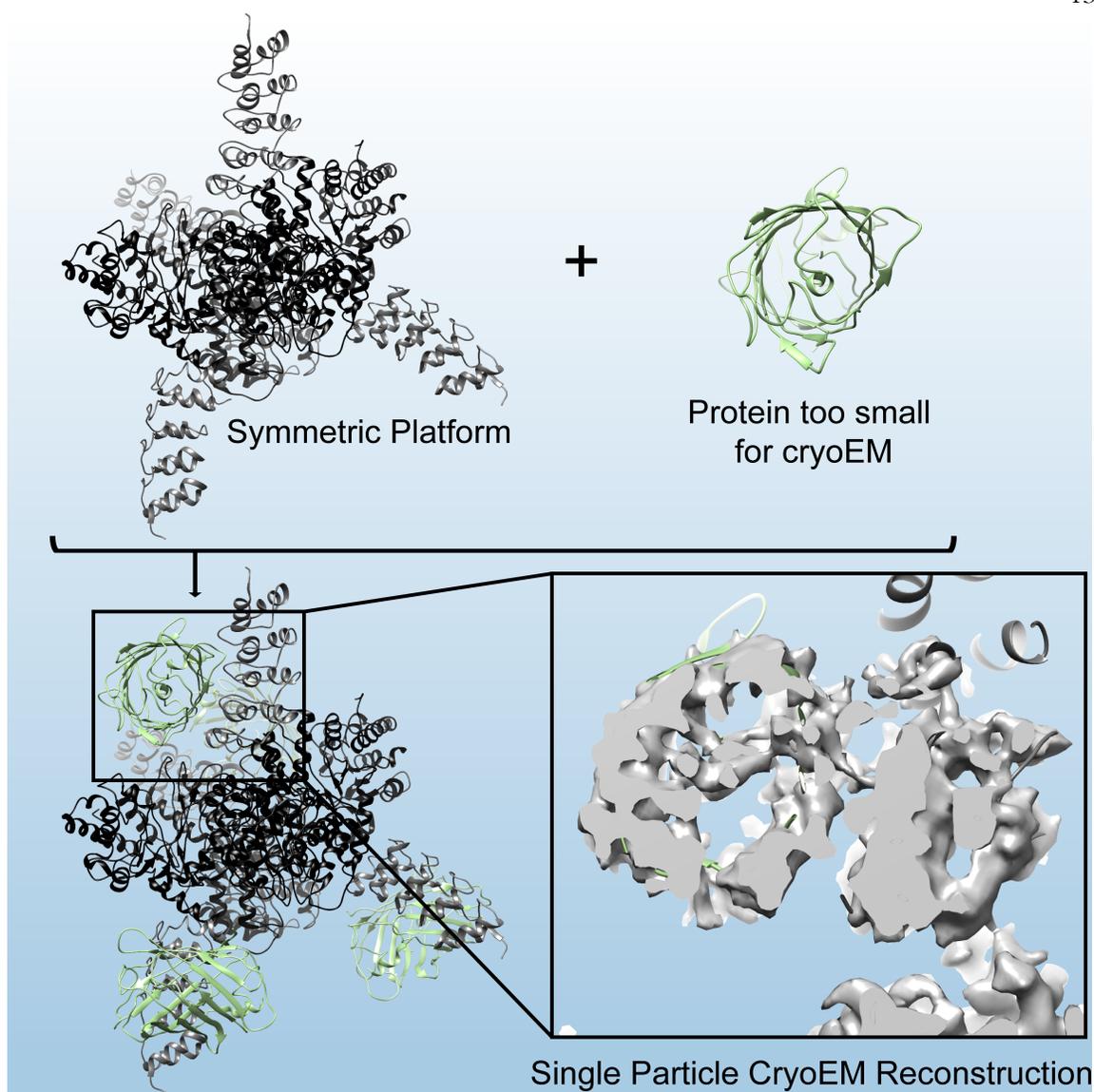


Figure 6 Graphical Abstract For Chapter 2

Summary

Solving protein structures by single particle cryo-electron microscopy (cryoEM) has become a crucial tool in structural biology. While exciting progress is being made towards the visualization of small macromolecules, the median protein size in both eukaryotes and bacteria is still beyond the reach of cryoEM. To overcome this problem, we implemented a platform strategy where a small protein target was rigidly attached to a large, symmetric base via a selectable adapter. Of our seven designs, the best construct used designed

ankyrin repeat protein (DARPin) rigidly fused to tetrameric rabbit muscle aldolase through a helical linker. The DARPins retained its ability to bind its target: green fluorescent protein (GFP). We solved the structure of this complex to 3.0 Å resolution overall, with 5 to 8 Å resolution in the GFP region. As flexibility in the DARPin position limited the overall resolution of the target, we describe strategies to rigidify this element.

Introduction

Single particle cryoEM can reveal the structures of large macromolecular complexes to near atomic resolution. To solve a protein structure by single particle cryoEM, purified proteins are rapidly frozen in a thin layer of vitreous ice. A transmission electron microscope is used to collect projection images of the protein. Individual proteins are identified in the ice and their orientations are computationally determined. The projection images are then combined to calculate a 3D reconstruction of the protein.

A fundamental challenge in single particle cryoEM is that small proteins do not produce enough contrast in noisy projection images to precisely determine their orientation. Richard Henderson estimated that with ideal images, a 3 Å structure could be reconstructed for a 40 kDa protein (R Henderson 1995). Unfortunately, real electron micrographs are imperfect so this theoretical minimum of macromolecular size has never been reached. The smallest protein to be solved to near atomic resolution so far by cryoEM is hemoglobin (64 kDa) (Khoshouei et al. 2017), but the median protein lengths in both bacteria (27 kDa) and eukaryotes (36 kDa) are about two times smaller (Brocchieri and Karlin 2005). Consequently, many proteins in biology are beyond the reach of high-resolution structure determination by single particle cryoEM.

Over the years, several strategies to overcome the size limit problem in single particle cryoEM have been suggested. Two major themes have emerged to increase the target mass and improve its orientation determination. First, the target can be decorated with antibody fragments (Jensen and Kornberg 1998; Wu et al. 2012). Second, the target can be rigidly attached to a large platform protein. The platform is typically composed of a base protein and an adapter. The purpose of the base protein is to increase the molecular weight, which

facilitates accurate particle picking and precise particle orientation determination. The adapter can be customized (a covalent fusion between the target and the base) or general (a selectable adapter that facilitates non-covalent binding of the target to the platform base). Covalent approaches have utilized direct fusions between the target protein and the base either via a flexible linker adapter (Kratz, Böttcher, and Nassal 1999) or a helical junction adapter (Coscia et al. 2016; Liu et al. 2018). For the platform to be successful, the adapter must be rigidly attached to the base. The flexible linker adapter was therefore insufficient to determine the structure of the target (Kratz, Böttcher, and Nassal 1999), but the use of a helix-forming peptide linker (Liu et al. 2018; Padilla, Colovos, and Yeates 2001) or direct concatenation of two helices (Coscia et al. 2016; Jeong et al. 2016) has shown promise. Most recently, Liu et al. demonstrated that a rigid, continuous α -helix could be formed by linking the terminal α -helices of a designed ankyrin repeat protein (DARPin) and a nanocage subunit through a helix-forming peptide linker (Padilla, Colovos, and Yeates 2001). Notably, Liu et al. were able to show the structure of the 17 kDa DARPin to 3.5 to 5 Å local resolution (Liu et al. 2018). Unfortunately, these strategies are limited to target proteins with a terminal α -helix, and their implementation requires that the length of the helical junction adapter must be customized for each new target. Utilizing a non-covalent platform strategy with a selectable adapter (like an antibody or a DARPin) has the potential to be generally applicable, regardless of the structure of the target, since the selectable adapter could be raised against any target using phage display, while the invariant nature of the adapter framework region would allow the one-time optimization of a rigid attachment point between it and the base. Along these lines, Liu et al. suggested that their DARPin-nanocage could display a small protein for structure determination by cryoEM (Liu et al. 2018). A follow up to Liu et al., 2018 was posted to the bioRxiv while this manuscript was under review (Liu, Huynh, and Yeates 2018). In the follow up paper, Liu and coworkers reported the fusion of their previously reported nanocage with a DARPin against GFP. They engineered the DARPin to have additional stabilizing mutations designs (Kramer et al. 2010) and achieved a near atomic resolution structure of the target GFP.

Here we report the outcomes of a variety of new designs.

Results

Platform strategy and design

The goal of our study was to design a generally applicable platform to solve small protein structures by single particle cryoEM. We explored several candidate base proteins and selectable adapters (**Figure 7A**, **Figure 8A**). We favored bases that were easy to purify and that had already been solved to high-resolution by single particle cryoEM. We reasoned that oligomeric and symmetric (as a globular protein, or as a helical tube) bases would be best.

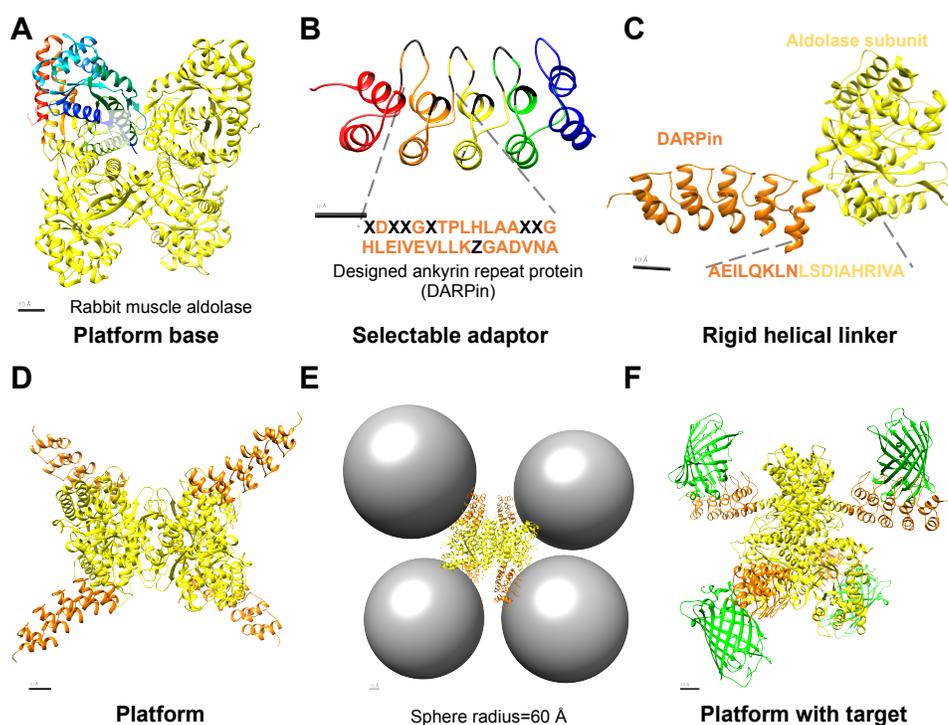


Figure 7 The Design of the platform

(A) The platform base was homotetrameric rabbit muscle aldolase (PDB ID code 5VY5). One subunit was depicted with rainbow coloring and N and C labels to indicate the orientation of the monomer chain. The other three identical subunits are shown in yellow. (B) The selectable adapter was a Designed Ankyrin Repeat Protein (DARPin) (PDB 5MA6). Shown below is a close-up view of the repetitive motif of DARPin with its amino acid sequence (orange). Using a phage display library, DARPins can be generated against a protein target. The selectable residues are depicted in black as X (any amino acid except cysteine or proline) or Z (amino acids asparagine, histidine or tyrosine) (Brauchle et al. 2014). (C) The final helix of the C-terminal cap of the DARPin (orange) was directly fused to the first alpha helix of aldolase (yellow) to form the platform subunit. (D) The D2 symmetry of the DARPin-aldolase fusion demonstrates ample space for target binding. (E) Spheres (radius=60 Å) were drawn in the position where each DARPin binds its target. A globular protein of up to 740 kDa could be accommodated on the DARPin-aldolase platform without steric clash. (F) The model of the DARPin-aldolase platform in complex with GFP (green) is shown. See also **Figure 8** and Movie S1.

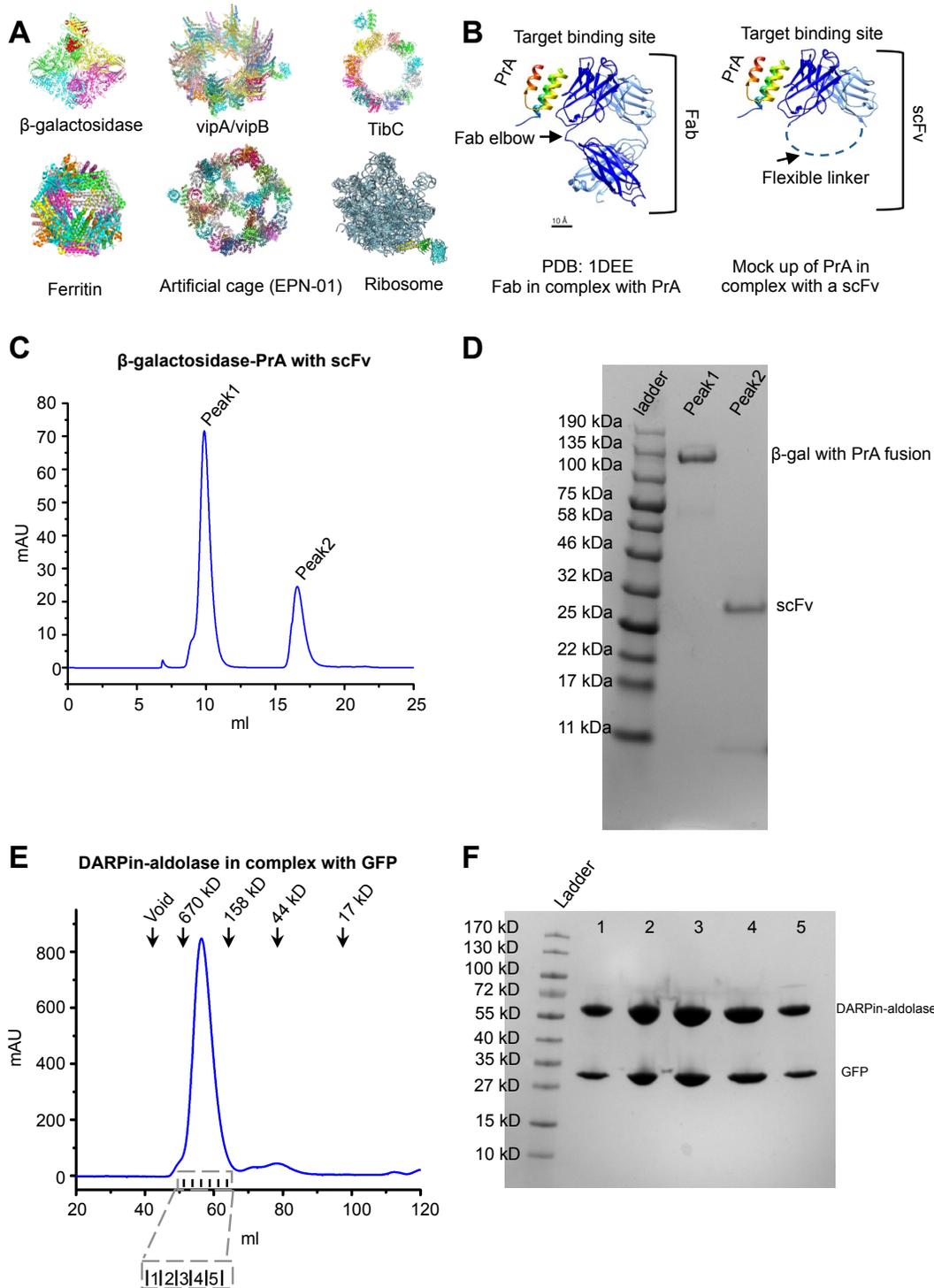


Figure 8 Attempted platform designs and outcomes, related to Figure 7.

(A) Models of the six of the platform base proteins tested here. The seventh model (aldolase) tested is shown in **Figure 7**. The fused DARPin (in green) and target GFP (in cyan) were shown for only one subunit for clarity. (B) Our Protein A/scFv selectable adapter design was inspired by PDB

1DEE (Graille et al. 2001). On the left, PDB 1DEE shows the PrA three helix bundle (represented in rainbow from red to blue as N- to C-termini respectively) in complex with a Fab. On the right, a mock up of a scFv binding PrA (based on PDB 1DEE) is shown. A flexible linker (dashed line) connects the two beta sandwich domains of the scFv. Scale bar, 10 Å. (C) Briefly, the β -gal-PrA fusion protein was mixed with equal copies of scFv and incubated overnight. The mixture was resolved into two peaks (1 and 2) by gel filtration chromatography using a Superdex 200 10/300 GL column (GE Healthcare). (D) SDS-Page gel stained with Coomassie Blue of peaks 1 and 2 from (C) showed that the beta-galactosidase-PrA fusion protein and the scFv appeared in different peaks, suggesting that a stable complex did not form. Bands are labeled with β -gal with PrA or scFv. (E) Gel filtration chromatography of the purified DARPin-aldolase platform in complex with GFP on Superdex 200 10/300 GL column yielded one main peak. The black arrows mark the molecular weight calibration and void volume. Fractions 1 to 5 are labeled. (F) SDS-Page gel stained with Coomassie Blue of fractions 1 to 5 from (E). The bands representing the DARPin aldolase platform subunit and the GFP are labeled.

As selectable adapters, we first considered antibody fragments (Fabs and scFvs). Fabs have a flexible elbow connecting two immunoglobulin regions, whereas scFvs are made up of one immunoglobulin region (**Figure 8B**). The Fab elbow could introduce flexibility, so we preferred the smaller, more compact scFv (**Figure 8B**). However, because the beta sandwich immunoglobulin fold of a scFv could be difficult to rigidly fuse to the surface of a platform base, we identified Staphylococcus Protein A (PrA) as a linker that could bind the invariant region of a scFv (Graille et al. 2001) (PDB 1DEE, **Figure 8B**). As PrA is a three-helix bundle, we reasoned that it could be rigidly attached to a base via a helical linker. Thus in one of our designs, the C-terminal helix of PrA was fused to the N-terminal helix of the base protein. Since PrA is capable of binding the invariant scFv framework, the base-PrA:scFv interfaces would not need to be redesigned for each new target. Unfortunately, in our biochemical experiments, we observed that the PrA:scFv interaction did not remain stable through a gel filtration column (**Figure 8C-D**), indicating that the binding affinity was not strong enough for our purposes. Further mutagenesis to the PrA:scFv interface may strengthen the interaction. Regardless, a fundamental concern with this design is that two non-covalent binding interactions are required (PrA:scFv, and scFv:target), which could lead to occupancy issues. As a result, we moved to DARPins as our selectable adapter (**Figure 7B**).

In our designs, the final alpha helix of the DARPin C-terminal cap (C-cap) was directly fused to the first α -helix of the base (**Figure 7C, Methods**). All DARPin libraries use a C-cap to stabilize the protein, so we expect it will be straightforward to swap in any DARPin built on the same framework (**Figure 7B**). In the base-DARPin platform design, only one

non-covalent interaction is required (between the DARPin and the target), which results in a more predictable and stable complex. We chose a DARPin that formed a stable complex with GFP with picomolar range binding affinity as a first test case (Brauchle et al., 2014) and screened several base-DARPin candidates.

For all of our DARPin-base constructs, the fusion was designed using the molecular graphics program COOT (Emsley and Cowtan 2004). On the base protein, a N- or C-terminal alpha helix was identified as the fusion site (see **Methods**). The base alpha helix and the DARPin terminal alpha helix were aligned in COOT. We assumed that the alpha helicity and the orientation of the fusion helix would not be significantly perturbed by the fusion. For some constructs, a helix-forming peptide linker was added between the base and the DARPin to achieve the desired fusion helix length (Padilla, Colovos, and Yeates 2001). The possible steric clashes of the designed model were visually examined and the open binding space was maximized. For example, the ferritin-DARPin fusion design was abandoned because adding a DARPin to each ferritin subunit didn't leave enough space for a target to bind (PDB 1EUM, **Figure 8A**) (Stillman et al. 2001).

Screening base candidates

We performed expression trials for two base-PrA candidates and six base-DARPin candidates (**Figure 7D, Figure 8A**). These bases included the *E. coli* ribosome, (Noeske et al. 2015; Shoji et al. 2011) β -galactosidase (β -gal) (Bartesaghi et al. 2015), the vipA/vipB helical tube (Kudryashev et al. 2015), an artificial nanocage based on EPN-01 (Votteler et al. 2016), TibC (Yao et al. 2014), and aldolase (Herzik, Wu, and Lander 2017).

Our initial expression trials utilized the PrA/scFv strategy discussed in the previous section with β -gal. Concurrently, we found that the ribosomal protein L29-PrA fusion could be expressed, but we were unable to incorporate it into Δ L29 *E. coli* ribosomes (Shoji et al. 2011).

Because β -gal tetramerization requires the N- and C-termini of each subunit (Ullmann, Jacob, and Monod 1967), an internal DARPin insertion was used, flanked by a helix-

forming peptide (at the DARPin N-cap) and a flexible linker (at the DARPin C-cap)(Padilla, Colovos, and Yeates 2001). Biochemically the β -gal-DARPin platform formed a stable complex with GFP, but no cryoEM density was observed for the DARPin or GFP in our 3 Å reconstruction. This means that the helical linker was flexible relative to the β -gal base. Our design for the EPN-01 based nanocage also inserted the DARPin into the middle of the sequence. The EPN-01 DARPin fusion protein failed the *E. coli* expression test.

We therefore focused on bases with a terminal α -helix that could be rigidly fused to the DARPin. The vipA/vipB, TibC, and aldolase proteins all had long terminal α -helices to facilitate direct fusion (**Figure 7, Figure 8A**). In our experiments, the helical tube vipA-DARPin/vipB platform exhibited poor expression in *E. coli*. The purified TibC-DARPin platform formed a stable complex with GFP, but the complex demonstrated aggregation and preferred orientation on plunge frozen grids. In contrast, the DARPin-aldolase platform was well-behaved (**Figure 8E-F, Figure 9**).

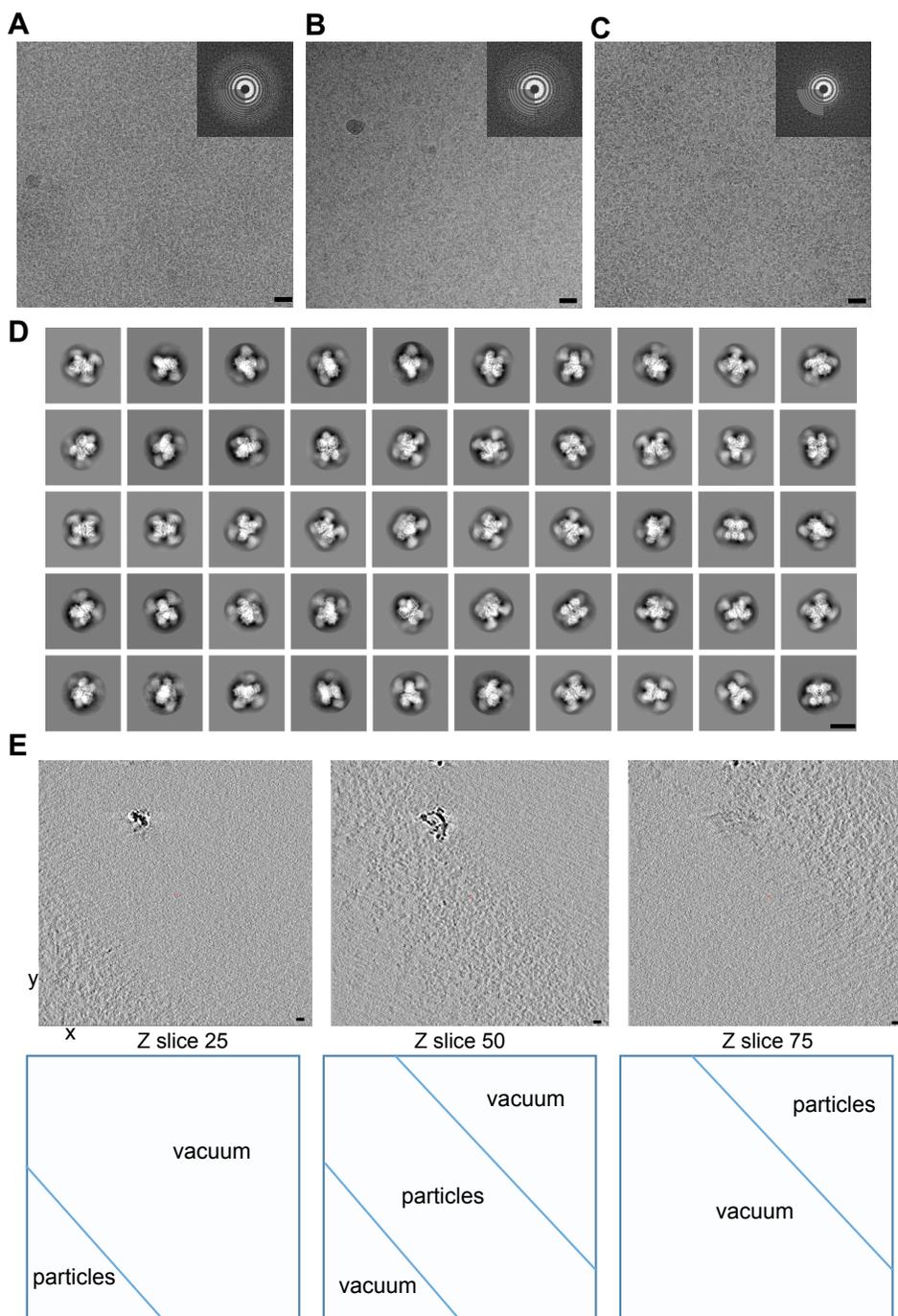


Figure 9 cryoEM data for the GFP:DARPin-aldolase complex, related to **Figure 8**
(A)-(C) Motion-corrected, dose weighted micrographs of DARPin-aldolase platform in complex with GFP in vitreous ice (left) with the Fourier transformation (inset). Micrographs were collected at 0° tilt in the first session **(A)**, at 0° tilt in the second session **(B)** and at 26° tilt in the third session **(C)**. Each micrograph has been low-pass filtered to 10 Å to enhance the contrast. Scale bar, 20 nm.
(D) Representative 2D classification results from Relion. The 2D class images are 300 Å by 300 Å. Scale bar, 100 Å.
(E) Tomography of the GFP:DARPin-aldolase grids showed a single layer of

particles suspended in thin ice. Three Z slices ($z=25$, $z=50$, $z=75$) at a z thickness of 10 slices (to increase contrast) were selected from the tomogram. Axes are X and Y. Scale bars, 10 nm. The tomogram was 100 z slices thick. The ice in the tomogram runs through the ZX plane diagonally with vacuum (empty space) on either side. Below each image is a schematic of which region represents particles in ice and which region represents vacuum above or below the layer of ice.

In our DARPIn-aldolase platform, the C-terminal α -helix of the DARPIn was directly concatenated to the N-terminal α -helix of aldolase (**Figure 7C-D**). The D2 symmetry of the DARPIn-aldolase platform provided extensive space for the target and could potentially accommodate a globular protein of up to 740 kDa without steric clash (**Figure 7E-F, S1 Movie**). The purified GFP:DARPIn-aldolase complex was stable in a gel filtration column with an apparent 1:1 stoichiometry of DARPIn-aldolase to the target (GFP) (**Figure 8E-F**).

CryoEM analysis of the GFP:DARPIn-aldolase complex

To solve the structure of GFP bound to the DARPIn-aldolase platform, we collected 1,681 micrographs on a Titan Krios (**Figure 10**). Because the thin ice forced a slight preferred orientation issue, an additional 1,180 micrographs were collected at 26° tilt (Naydenova and Russo 2017). High quality micrographs were selected after CTF determination (**Figure 9A**). The final Relion autopicking round used 2D class references generated earlier, and yielded 841,776 particles from 1,548 micrographs (**Figure 10**). The particles were converted to cryoSPARC, where 2D classes with strong secondary structure were selected for reconstruction and converted back to Relion. Representative 2D classes from Relion show clear secondary structure (**Figure 9D**). The GFP:DARPIn-aldolase complex reconstruction yielded an overall resolution of 3 Å with C1 symmetry (**Figure 10, S2 Movie**). Further classification suggested too much conformational heterogeneity to apply D2 symmetry. The aldolase core and the helical linker were resolved to near atomic resolution (**Figure 11B-D, S2 Movie**). The DARPIn and GFP exhibited a local resolution of 4 to 8 Å, with discontinuous regions of higher resolution of 3.5 Å (**Figure 11, S2 Movie**). Although the resolution in the GFP and DARPIn portion was not sufficient to build a model or assign sequence *de novo*, the static X-ray structures of GFP and the DARPIn could be reliably docked into the map (**Figure 12A, S2 Movie**). During the design phase we made an approximate mock up of the DARPIn-aldolase construct in Coot, but did not use computation methods to relax the structure. Because of this, we cannot

quantitatively compare the DARPin-aldolase model from our cryoEM structure to the design. Additionally, at the current resolution, we cannot precisely describe the side chain interactions that facilitate the GFP binding to the DARPin, but the orientation looks similar to that described by the crystal structure of a similar GFP:DARPin interaction (PDB 5MA6)(Hansen et al. 2017).

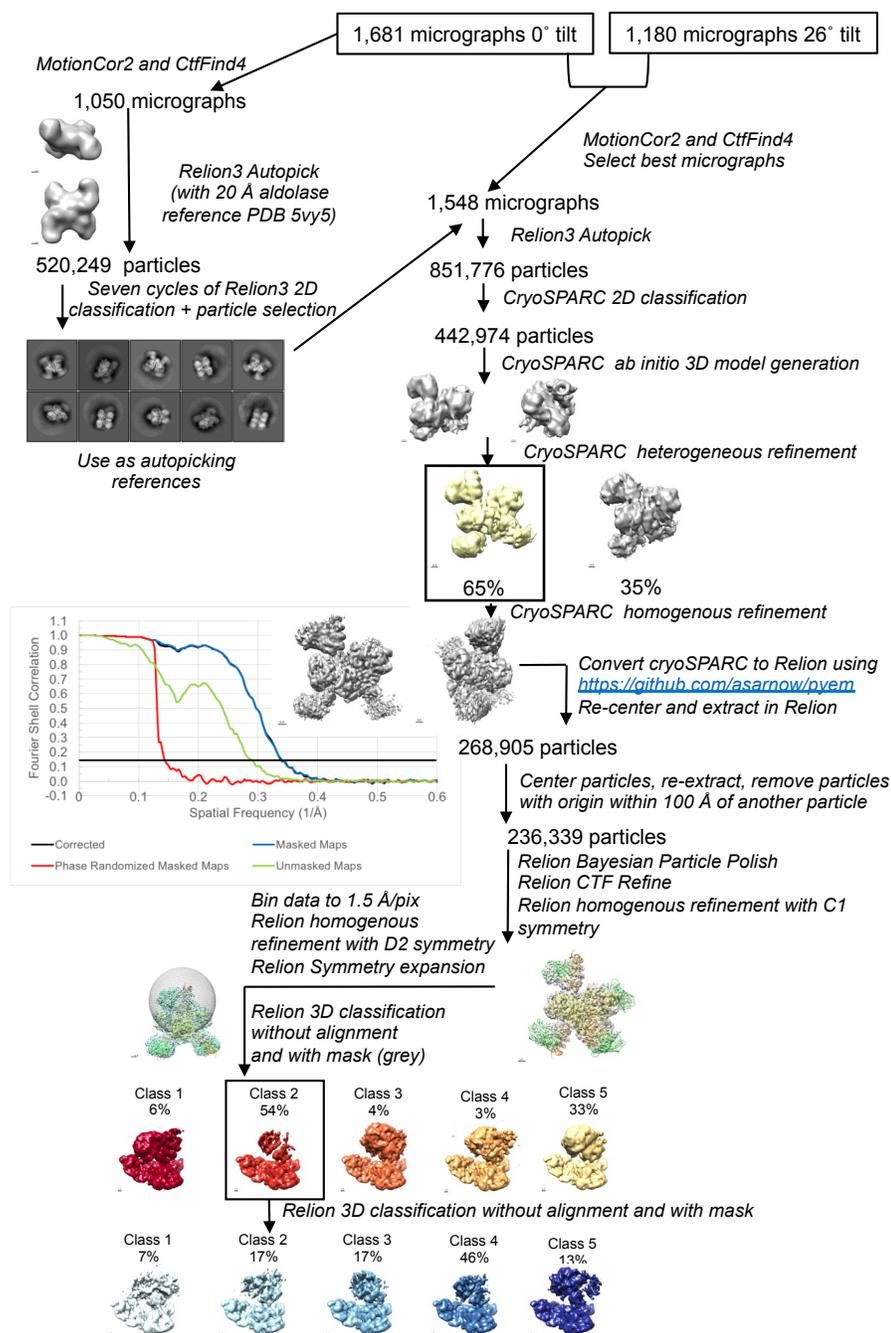


Figure 10 Major steps of the last cycle of cryoEM data processing, related to STAR Methods
Tilted data was collected several months after the initial data collection sessions, so several generations of cryoEM data processing contributed to the final data pipeline in this paper. The major steps in the final cycle of cryoEM data processing are summarized in this flowchart. Scale bars are 10 or 20 Å as indicated per image.

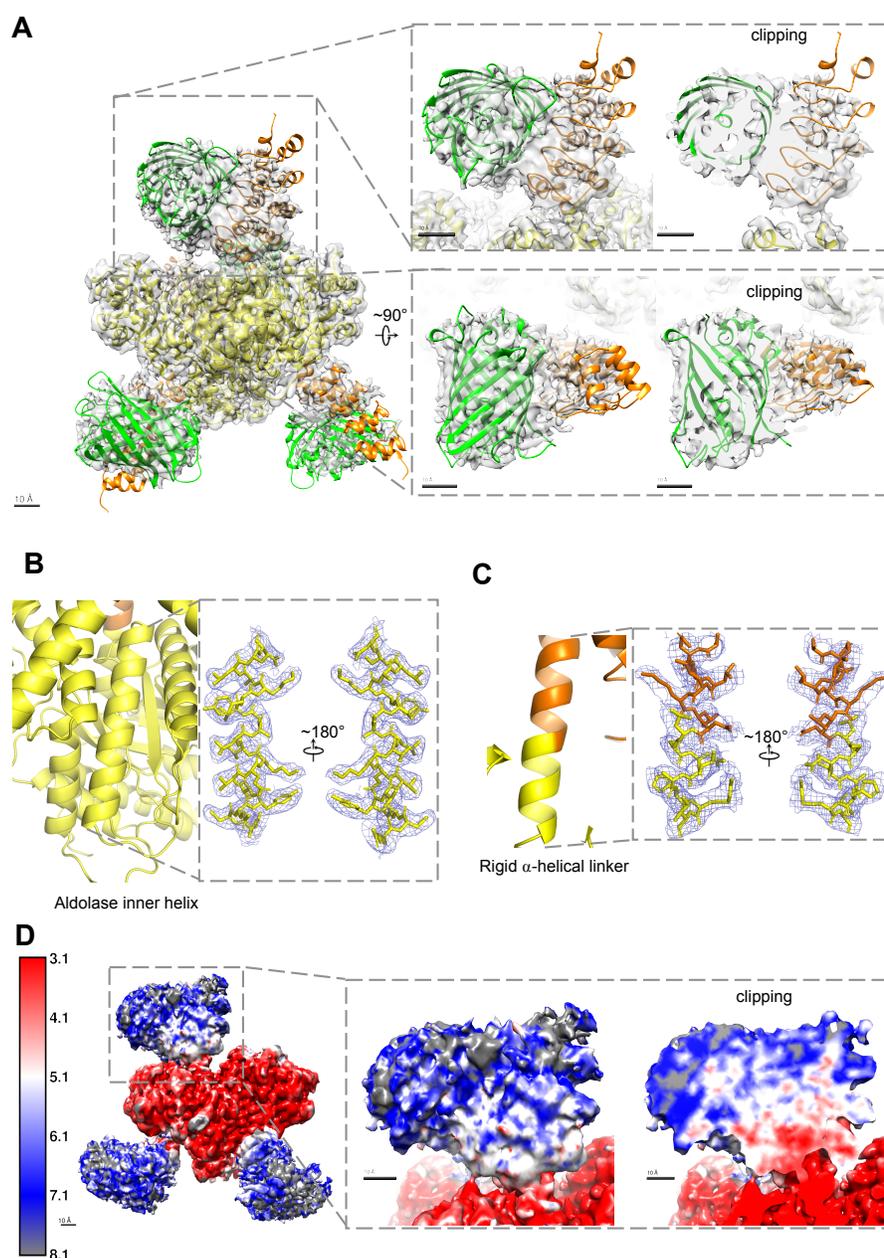


Figure 11 The cryoEM structure of the DARPin-aldolase platform in complex with the target GFP

(A) Surface of the sharpened 3 Å C1 reconstruction of DARPin-aldolase platform in complex with GFP. The crystal structures of GFP (green), the DARPin (orange) and aldolase (yellow) were docked into the cryoEM density. Two expanded views of the best resolved subunit are shown on the right. The expanded views in the dashed line boxes are shown from the top or side, and clipped halfway to indicate the quality of the fit. The cryoEM map is sharpened with Relion PostProcess. Chimera threshold, 0.005. Scale bars, 10 Å. (B) Ribbon diagram (left) and cryoEM density (right, blue mesh, zoned 1.8 Å within atoms) of an internal aldolase helix (residues Arg369 to Asp387). (C) Ribbon diagram (left) and cryoEM density (right, blue mesh, zoned 1.8 Å within atoms) of the helical linker (residues Ala176 to Ile191) between the DARPin (orange, residues Ala176 to

Lys181) and aldolase (yellow, residues Leu182 to Ile191). **(D)** ResMap local resolution estimate of the unsharpened cryoEM density (left) and of the best subunit (right). The expanded view in the dashed line box is shown from the side (left) and halfway into the GFP:DARPin density with clipping (right). The gradient color scale from red (3.1 Å) to grey (8.1 Å) was generated with ResMap. Chimera threshold 0.003. Scale bars, 10 Å. See also **Figure 9** and **Movie S2**.

Potential sources of heterogeneity

Because of the 5 to 8 Å local resolution range in the GFP portion of the map (**Figure 11D**), we suspected that part of the GFP:DARPin-aldolase complex was flexible. Since we observed some preferred orientation in our dataset, we hypothesized that some of this heterogeneity may be related to how the particles interact with the air-water interface. We collected tomograms in the centers of holes (**Figure 9E**). The particles formed a single layer in a thin sheet of ice that was about as thick as the particles themselves. This suggests that every particle interacts with the air-water interface (Downing and Glaeser 2018). We prepared samples on graphene oxide-coated grids to circumvent this issue, but these samples did not reach high resolution.

DARPin position caused conformational heterogeneity

To better understand the conformational heterogeneity in the data, Relion particle symmetry expansion was used to consider each subunit individually (**Figure 10**) (Zhou et al. 2015). The symmetry expanded particles were subjected to 3D classification without alignment, a strategy in which the orientation parameters determined in the previous refinement are used to classify the particles into subsets. For this focused classification, a spherical mask that encompassed a single GFP:DARPin region and some of the aldolase subunit was used to increase the signal. The resulting five classes showed reasonable GFP:DARPin conformations (**Figure 12A, S3 Movie**), but subsequent refinements were still limited to 5 to 6 Å overall, which suggested that additional conformational heterogeneity remained within the subsets. The appearance of each 3D class varied greatly with the threshold level in Chimera, so we have shown the classes at thresholds from 0.02 to 0.01 (**Figure 12A, S3 Movie**). At the highest threshold (0.02), a hollow GFP barrel with some density for the chromophore is clearly visible in four classes (1, 2, 4 and 5) (**Figure 12A, S3 Movie**). The majority of the particles (54%) were classified into class 2, which at higher threshold (0.02) appeared to lack a DARPin (**Figure 12A, S3 Movie**). However,

relaxing the Chimera threshold to 0.01 showed sparse density for the GFP:DARPin moiety (**Figure 12A** and **S3 Movie**). Class 2 was subjected to an additional round of 3D classification where it revealed several reasonable but lower resolution GFP:DARPin conformations (**Figure 12B**, **S3 Movie**). To investigate the heterogeneity in the focused classes, we compared each class to Class 4 (**Figure 12C-D**). In the different classes, the GFP:DARPin density shows a clear rocking around the Y axis (**Figure 12C**) and around the Z axis (**Figure 12D**) relative to the aldolase base. At this point, we wondered if any these displacements could be attributed to the aldolase subunit. We performed a similar focused classification experiment with a mask around the aldolase subunit and the helical linker, but no rotation or shift was observed in the resulting subsets. Thus, we concluded that some of the displacement we observed likely arose in the C-cap second helix that is fused into the helical linker, and other regions of the DARPin distal to the linker.

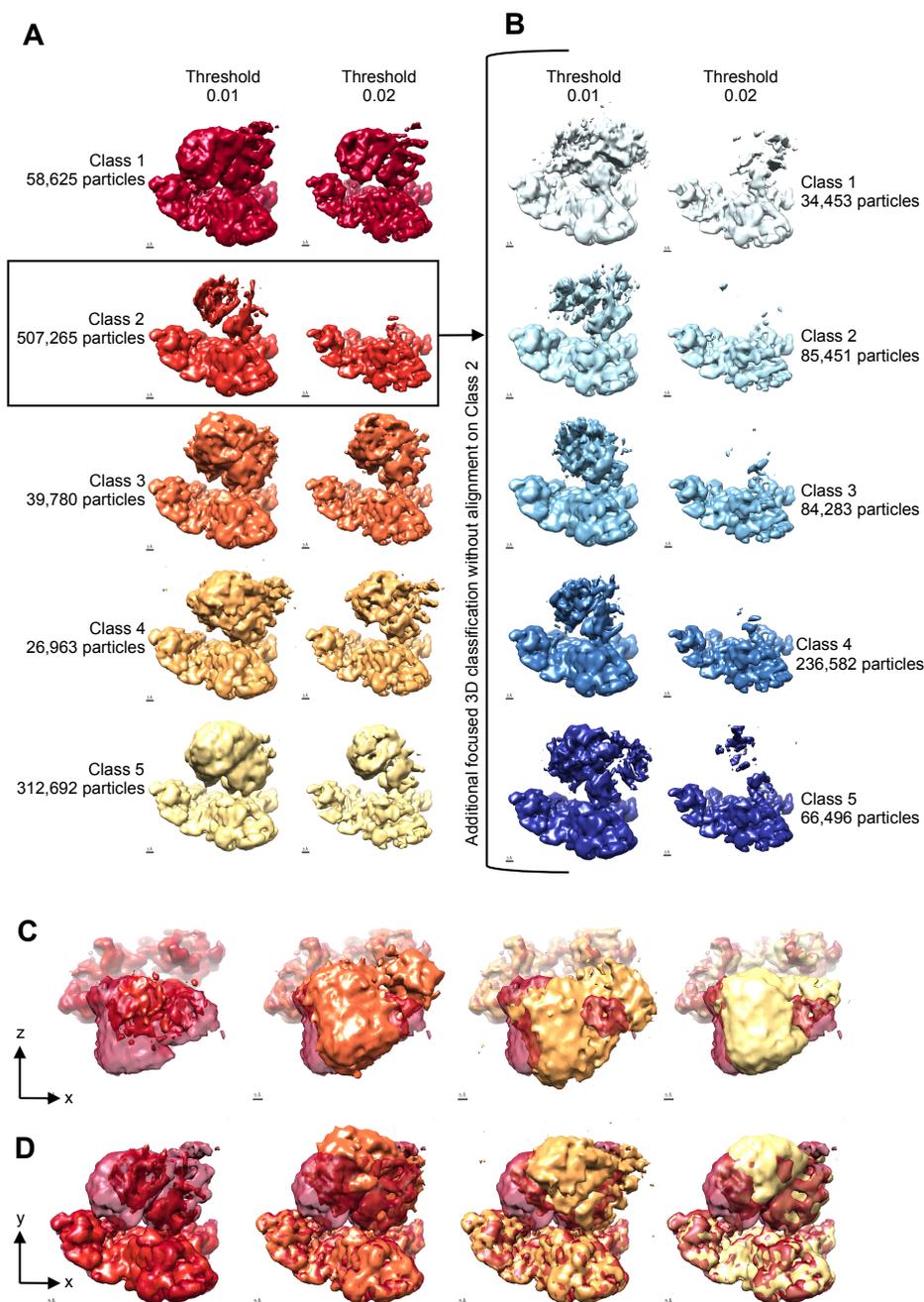


Figure 12 Symmetry expanded 3D classification of the GFP:DARPin region of the density.

(A) 3D classification without alignment of the symmetry expanded particles with five classes. The number of particles per class is indicated to the left of each class. Each class was viewed at two Chimera thresholds (0.01, 0.02) to facilitate direct comparisons. Scale bar, 5 Å. (B) An additional round of 3D classification without alignment was performed on Class 2 from (B). Scale bars, 5 Å. (C) The classes in (A) were each compared with Class 1 (dark red) to show the displacement between classes. The XZ plane is shown and the Y axis is perpendicular to the page. Class 1 was clearly displaced relative to the other classes. Chimera threshold, 0.01. Scale bar, 5 Å. (D) The

comparison from (C) is viewed looking down the Z axis. The Chimera threshold was 0.01. Scale bar, 5 Å. See also Figure S3 and Movie S3.

Discussion

In this study, we designed and tested a variety of platforms capable of non-covalently binding a small target protein via a selectable adapter for structure determination by single particle cryoEM. In our best construct, we resolved our target protein (GFP) to 5 to 8 Å resolution.

Our DARPin-aldolase platform has several advantages over other strategies. It is simple to express and purify. Aldolase has D2 symmetry and allows attachment of four targets without steric clash. Aldolase can be reconstructed to 2.6 Å resolution with even a 200 keV microscope (Herzik, Wu, and Lander 2017). Because DARPins can be readily generated against a wide range of small protein targets, the attachment of a DARPin to aldolase promises to be a generally applicable strategy. A recent study of the insulin degrading enzyme (IDE) bound to Fabs was able to isolate several IDE conformations using different Fabs (Z. Zhang et al. 2018). It stands to reason that different DARPins could also stabilize different conformations of the target. Because switching DARPins in the platform would be done by straight-forward DNA manipulations, our DARPin-aldolase platform has the potential to resolve a series of conformations of the target protein.

Our biochemistry experiments suggested that the purified GFP:DARPin-aldolase complex was very stable, and clear secondary structure was apparent in the 2D classes, yet heterogeneity remained. Potential sources of heterogeneity include the aldolase core, the angle of the linker helix relative to the aldolase core, and flexibility in the DARPin itself. The interaction of the particles with the air-water interface may also play a role. Since the air-water interface played a role in the heterogeneity observed in the structure reported here, designing a platform that would couch the target inside a cavity may be advantageous. Because the aldolase base and the helical linker region were resolved to near atomic resolution (**Figure 11B-C**), a large contribution to the heterogeneity we observed

likely stemmed from the DARPin C-cap. However, we cannot discount the affect of the other sources listed previously.

The DARPin against GFP used here was from a first generation DARPin library. The C-cap of the first generation DARPins was reported to be less stable than the other repeat modules (Seeger et al. 2013). While the crystal structure contained a well-resolved C-cap, the heterogeneity observed here suggests that it is not yet sufficiently rigid to serve as an attachment point in a cryoEM platform (**Figure 12**). Recent DARPin phage display libraries contain DARPins with reduced surface entropy and a more stable C-cap sequence(Seeger et al. 2013), however, and additional stabilizing surface interactions could be introduced in future designs (Kramer et al. 2010; Interlandi et al. 2008), or even a second attachment point of the DARPin to the base (at both N- and C-terminal caps of DARPin for instance). Together such improvements could allow the DARPin-aldolase platform to reveal the structures of many small proteins to near atomic resolution.

Acknowledgements

We thank Dr. Mark Herzik Jr. and Mengyu Wu at The Scripps Research Institute for help with sample preparation. We also thank Dr. Songye Chen, Dr. Andrey Malyutin, Dr. Rebecca Voorhees, and Dr. Bil Clemons at Caltech for technical assistance. We are also grateful to all members of the Jensen laboratory for discussion and technical assistance. In particular, we thank Dr. Lauren Ann Metskas and Dr. Ariana Peck. This work was supported by funds from NIH NIGMS P50 082545. CryoEM work was performed in the Caltech Beckman Institute Resource Center for Transmission Electron Microscopy.

Author Contributions

Conceptualization, Q.Y., S.J.W., G.J.J.; Methodology, Q.Y., S.J.W., J.Y.M; Investigation, Q.Y., S.J.W., J.Y.M; Data Curation, S.J.W.; Writing – Original Draft, S.J.W.; Writing – Review & Editing, S.J.W., Q.Y., G.J.J.; Visualization, S.J.W., Q.Y.; Supervision, G.J.J., Funding Acquisition, G.J.J.

Declaration of interests

The authors declare no competing interests.

STAR Methods

Contact for Reagent and Resource Sharing

Requests for further information as well as resources and reagents should be directed to and will be fulfilled by the Lead Contact, Grant Jensen (jensen@caltech.edu).

Experimental Model and Subject Details

Bacterial Strains

E. coli E. cloni and BL21(DE3) cells were cultured in Luria Broth or Autoinduction medium with appropriate antibiotics.

Method Details

Computational design

Designs were generated by examining the atomic coordinates of the base protein, the selectable adapter, and the target in UCSF Chimera(Pettersen et al. 2004) or COOT(Emsley and Cowtan 2004), manually adjusting the positions to approximate a fusion protein between the base and the selectable adapter, and assessing the design for potential steric hindrance. Promising designs left enough room for the target protein to bind the selectable adapter. The PDB identifiers of the models used in this process are summarized in the Key Resources Table. The sequences of the constructs tested are reported in **DataS1**.

The N- and C-termini of base proteins β -galactosidase and the EPN-01 nanocage are involved in multimerization, so the sequence of the selectable adapter was inserted within the base protein sequence. All other constructs utilized a terminal alpha helix to fuse the selectable adapter.

The computational α -helix fusion was generated by manually docking the rabbit muscle aldolase structure (PDB code: 5VY5) and the GFP/DARPin complex (PDB code: 5MA6) (Hansen et al. 2017; Herzik, Wu, and Lander 2017) in COOT. Because the DARPin sequence we selected has not been crystallized (DARPin 3G86.32(Brauchle et al. 2014)), residues in the 5MA6 PDB model were mutated in COOT to match our sequence. In order to rigidly join the aldolase and DARPin moiety together, we truncated the C-terminal flexible loop on DARPin and N-terminal flexible loop on aldolase, respectively, exposing the two terminal α -helices. The two terminal α -helices were manually concatenated and joined together to form an ideal α -helix using building α -helix tool in UCSF Chimera. The model was inspected for the orientation of DARPin relative to the aldolase, ensuring no steric clash and the providing enough space for target protein attachment. Structural design figures were generated using PyMOL1.8 (<https://pymol.org>) or UCSF Chimera.

Cloning, protein expression, and purification

The cDNA used for all base-selectable adapter fusions and for GFP was synthesized at Integrated DNA Technologies (IDT, Coralville, Iowa, USA) and cloned into an expression vector as indicated in the Key Resources Table.

The β -galactosidase sequence was derived from PDB 5A1A while the Protein A sequence and the single chain variable fragment (scFv) sequences were derived from PDB 1DEE(Bartesaghi et al. 2015; Graille et al. 2001). Graille et al. reported the crystal structure of *S. aureus* protein A bound to a human IgM Fab. To convert the Fab sequence into a scFv, the variable domains from the Fab immunoglobulin light chain (PDB 1DEE Chain A residues 1 to 106) and the Fab immunoglobulin heavy chain (PDB 1DEE Chain B residues 501 to 621) were connected using a flexible linker (see **DataS1** for sequences). The β -galactosidase fusion with Protein A was cloned into the pET21a vector with a C-terminal His tag. The scFv antibody against Protein A was cloned into the pET22b expression vector with a C-terminal His tag. The β -gal-PrA fusion and the scFv were each recombinantly expressed in *E. coli* BL21(DE3) cells overnight at 22 °C. Cells were lysed using sonication in a buffer containing 50 mM Tris-HCl pH 8.0 150 mM NaCl and 15 mM

imidazole, and purified using Ni-NTA affinity chromatography (Qiagen). About 300 μ g purified β -gal PrA fusion and 200 μ g scFv were mixed together in 1 ml incubation buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl). The mixture was incubated on ice overnight, and was subsequently analyzed by gel filtration using a Superdex 200 10/300 GL column (GE Healthcare) equilibrated with incubation buffer. Fractions for each peak were pooled and analyzed by SDS-PAGE.

The DARPin sequence was derived from DARPin 3G86.32 (Brauchle et al. 2014). The cDNA of GFP and of the DARPin-aldolase fusion were PCR-amplified and inserted into pACYCDuet and pET21b vectors respectively. The resulting GFP protein was untagged, while a C-terminal His-tag was used for the DARPin-aldolase chimeric fusion. GFP and DARPin-aldolase were co-expressed in *E. coli* BL21(DE3) using autoinduction medium with trace elements (Formedium) at 30 °C for overnight. Cells were harvested by centrifugation, lysed with sonication in a buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl and 15 mM imidazole, and then the complex was purified with Ni-NTA affinity chromatography. Gel filtration chromatography (Superdex 200 10/300 GL) was used to isolate the complete GFP:DARPin-aldolase complex. The gel filtration peaks were analyzed by SDS-PAGE. The purified GFP:DARPin-aldolase complex was concentrated to 2.5mg/ml in a buffer containing 25 mM Tris-HCl pH 8.0 and 150 mM NaCl.

CryoEM sample preparation and data collection

Electron microscopy grids were prepared at Scripps Research Institute. Briefly, 3 μ L sample of 2.5 mg/ml GFP-DARPin-aldolase complex was applied to a plasma cleaned Au UltraFoil Grid (200 Mesh, R2/2, Quantifoil) in a cold room (4°C, \geq 95% relative humidity). The grid was manually blotted with a filter paper (Whatman No.1) for approximately 3 seconds before plunging into liquid ethane using a manual plunger (Herzik et al., 2017). The grids were screened in Talos Arctica 200 kV with Falcon 3 (FEI) direct electron detector for ice thickness and sample distribution. Micrographs of GFP-DARPin-aldolase complex were collected on Titan Krios microscope (FEI) operating at 300 kV with energy filter (Gatan) and equipped with a K2 Summit direct electron detector (Gatan). For

untilted data, Serial EM was used for automated EM image acquisition (Mastronarde 2005). After calculating an efficiency score from early refinements using cryoEF (Naydenova and Russo 2017), additional data were collected at 26° using EPU software (FEI). A nominal magnification of 165,000x was used for data collection, corresponding to a pixel size of 0.865 Å at the specimen level, with the defocus ranging from -1.0 µm to -3.0 µm. Movies were recorded in superresolution mode, with a total dose of ~40 e-/Å², fractioned into 20 frames (0° tilt images) or 40 frames (26° tilt images) under the dose rate of 8.4 electron per pixel per second. Tilt series were collected using SerialEM on the Titan Krios at a nominal magnification of 165,000x, with a pixel size of 0.865 Å and at -4 µm target defocus (Mastronarde 2005). A bi-directional tilt scheme was used with a 3° tilt increment. The tilt series began at 0° and covered -45° to +45°. A 0.5 s movie with 5 frames was recorded for each tilt image in K2 super resolution mode using a dose rate of 8 electron per pixel per second.

Single particle cryoEM analysis

A summary of the major steps of data processing is available in **Figure 10**. Movies were decompressed and gain corrected with IMOD (Kremer, Mastronarde, and McIntosh 1996). Motion correction was performed using program MotionCor2 (Zheng et al. 2017), and exposure filtered in accordance with relevant radiation damage curves (Grant and Grigorieff 2015). Micrographs with high CTF Figure of Merit scores and promising maximum resolution (better than 3.6 Å for 0° tilt, and better than 6 Å for 26° tilt) were selected for further processing (1,548 micrographs). Several rounds of autopicking using combinations of different references and manual picking were analyzed to determine optimal settings, and all yielded similar results. In an early processing cycle, to generate 2D class references, a subset of 1,050 micrographs was autopicked with a model of aldolase (PDB: 5vy5) generated to 20 Å resolution in Chimera and subjected to an additional 10 Å low pass filter by Relion auto picking with D2 symmetry and 15° angular sampling. These particles experienced seven iterative rounds of 2D classification, subset selection, and re-extraction to generate the ten 2D classes used as a reference for the final round of autopicking. This yielded 851,776 particles from 1,548 micrographs. Particles were extracted in RELION

(Kimanius et al. 2016; Scheres 2012a; 2012b; Zivanov et al. 2018) and initial 2D classification was performed in cryoSPARC(Punjani et al. 2017). High quality 2D classes were selected for further processing (442,974 particles). The initial model was *de novo* generated and subsequent 3D refinement were performed using cryoSPARC. The UCSF PyEM package (<https://github.com/asarnow/pyem>) script was used to convert the cryoSPARC coordinates into Relion (268,905 particles). Duplicate particles were removed and the resulting 236,339 particles were analyzed by 3D refinement, Bayesian Particle Polishing and CTF Refinement in Relion. The data were binned to 1.5 Å/pixel, refined with D2 symmetry, and symmetry expanded. Symmetry expanded particles were used in 3D classification without alignment. All reconstructions were analyzed using the *relion_display* function and UCSF Chimera. The coordinate model was built by breaking our initial GFP:DARPin-aldolase PDB model into domains (GFP, DARPin and aldolase subunits) and rigidly docking these individual protein structures into the EM map using Chimera. Once the orientations were identified, the model was then fit and adjusted manually in UCSF Chimera and Coot. The figures were generated using UCSF Chimera, and local resolution and final Fourier shell correlation were calculated using ResMap (Kucukelbir, Sigworth, and Tagare 2014a) and cryoSPARC.

Tomography

Frames were motion corrected using MotionCor2 and reconstructed in IMOD. Because the grids lacked gold fiducial markers, patch tracking was used to align the tilt series. The weighted back projection method was used to reconstruct tilt series into tomograms.

Quantification and Statistical Analysis

Statistical analyses were performed within published software as described in *Single particle cryoEM analysis* method section.

Data and Software Availability

Data deposition

The cryoEM density map and coordinates of the GFP:DARPin-aldolase complex have been deposited in the Electron Microscopy Data Bank (EMDB) and the Protein Data Bank (PDB) with access codes: EMD-9277 and PDB 6MWQ.

Software availability

This study did not produce additional software. The software used here (summarized in Key Resources Table) has been published and is publically available. Software was accessed from developer websites, or from the SBGrid Consortium (Morin et al. 2013).

Multimedia files

S1 Movie. Model of DARPin-aldolase platform with GFP, and with four spheres (radius=60 Å) anchored at the DARPin binding sites, related to Figure 1

First the movie shows the design of the DARPin-aldolase platform in complex with GFP. Second, the movie shows a model of the DARPin-aldolase platform in complex with four spheres (radius=60 Å) anchored at the DARPin binding sites to demonstrate the space for target binding. Scale bars, 50 Å.

<https://www.cell.com/cms/10.1016/j.str.2019.04.003/attachment/277f0bf7-8524-4231-9adf-60cd618a6d02/mmc2.mp4>

S2 Movie. CryoEM structure of DARPin-aldolase platform in complex with GFP with model fit and local resolution, related to Figure 2

The cryoEM structure of the DARPin-aldolase platform in complex with GFP (EMD-9277) discussed in Figure 2A-B of the paper is shown with the atomic model (PDB 6MWQ) fit. The movie first shows the sharpened cryoEM map and model at Chimera threshold 0.04, then focuses in on the best resolved GFP:DARPin-aldolase subunit using Chimera threshold 0.06. Finally the local resolution of the unsharpened map from Figure 2D is shown at Chimera threshold 0.003. Scale bars, 10 Å.

<https://www.cell.com/cms/10.1016/j.str.2019.04.003/attachment/a34e80c0-deb4-4f93-bd63-cafa268f77b8/mmc3.mp4>

S3 Movie. Focused 3D classification of DARPin-aldolase subunit in complex with GFP, related to Figure 3

The focused classification classes from Figure 3 are shown at Chimera threshold 0.01 with clipping to demonstrate that each class has density and the GFP barrels are hollow. This movie accompanies Figure 3. Scale bar, 10 Å.

<https://www.cell.com/cms/10.1016/j.str.2019.04.003/attachment/1276eacb-c16c-411a-8aad-74252f5cbc1e/mmc4.mp4>

Data S1

Protein sequences of constructs used in this study.

<https://www.cell.com/cms/10.1016/j.str.2019.04.003/attachment/039530e1-8545-430e-985f-5f91dd9248ff/mmc5.pdf>

Chapter 3:

CRYOEM STRUCTURE OF THE VIBRIO CHOLERAE TYPE IV COMPETENCE PILUS SECRETIN PILQ

Sara J. Weaver¹, Matthew Sazinsky², Triana Dalia³, Ankur Dalia³, Grant J. Jensen⁴.

¹Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA, ²Chemistry, Pomona College, Pomona, CA, USA, ³Biology, Indiana University, Bloomington, IN, USA, ⁴Biology, California Institute of Technology, Pasadena, CA, USA.

Abstract

Natural competence is the process by which bacteria take up genetic material from their environment and integrate it into their genome using homologous recombination. In *Vibrio cholerae*, the Type IV competence pilus is thought to mediate DNA uptake by extending from the cell surface, binding DNA in the environment, and retracting to thread this DNA through the outer membrane secretin, PilQ, through which the pilus passes. A lack of structural information, however, has hindered our understanding of this process. PilQ is a 870 kDa outer membrane pore with C14 symmetry. Here, we purify the *V. cholerae* PilQ secretin from *V. cholerae* cells in amphipols for single particle cryogenic electron microscopy (cryoEM). We solve the structure to 3.0 Å and provide insight on the channel that DNA likely traverses during this mode of horizontal gene transfer.

Introduction

The gram negative bacterium *Vibrio cholerae* causes cholera, a water-borne disease marked by diarrhea (D. Hu et al. 2016). Since 1817, seven cholera pandemics featuring different strains have affected the world. Different strains arise in part because of horizontal gene transfer. Horizontal gene transfer contributes to the development of antibiotic resistance among a population and to the development of new serotypes of bacteria in a pandemic.

V. cholerae is known for high rates of horizontal gene transfer by natural transformation. In *V. cholerae*, natural transformation was first observed to occur in bacteria growing in the presence of chitin (Meibom et al. 2005). Chitin, a biopolymer found in the exoskeletons of crustaceans, induces transcription of the chitin regulon and expression of the Type IV competence pilus (Meibom et al. 2004). The Type IV competence pilus facilitates environmental DNA uptake by extending and retracting from the cell surface (Ellison et al. 2018).

Bacterial secretins are united by a common C-terminal secretin domain (Genin and Boucher 1994). Secretins are found in the Type II Secretion System (T2SS), the T3SS, the Type IV pili, and filamentous phage (Majewski, Worrall, and Strynadka 2018). While the C-terminal secretin domain is remarkably similar across these secretion systems, the N-terminal region varies. The T2SS secretins have N-terminal protein domains N0 to N3, followed by the secretin domain and then the S-domain. The T3SS secretins lack the N2 domain, while the Type IV competence pilus secretins lack the N1 and S-domains (**Figure 13**). Additionally, the N-terminus of Type IV competence pilus secretins typically include one or more AMIN domains that interact with the peptidoglycan (de Souza et al. 2008).



Figure 13 Protein domain organization in the Type IV Competence Pilus (T4P), the Type II Secretion system (T2SS) and the Type III Secretion System (T3SS).

Protein domain organization in the Type IV competence pilus machine, the Type II Secretion System (T2SS) and the Type III Secretion System (T3SS). The AMIN domain (red), the N0 (or STN) domain (blue), the Secretin_N domains N1 (purple), N2 (brown), and N3 (green), the secretin domain (yellow) and the S domain (pink).

The structure of the Type IV competence pilus secretin has been examined several times, though no work has provided high-resolution details. In 2012, a 26 Å structure of

Neisseria meningitidis PilQ was solved by cryoEM (Berry et al. 2012). NmPilQ was isolated from *N. meningitidis* outer membrane vesicles in DDM. In 2016, Koo et. al solved the structure of *Pseudomonas aeruginosa* PilQ to 7.4 Å (cyan in **Figure 14**). Koo et. al expressed His-tagged PilQ in *P. aeruginosa* cells, prepared a membrane preparation, and purified PilQ in Anzergent 3-14 detergent. PaPilQ appeared to have 14 peripheral spokes lining the beta barrel that did not demonstrate clear C14 symmetry, so C7 symmetry was assumed. The final structure of PaPilQ only used 6,919 particles in the final reconstruction, which may have limited their overall resolution. In a 2017 paper, D’Imprima et. al described the structure of *T. thermophilus* PilQ to ~7 Å resolution with reported C13 symmetry (yellow in **Figure 14**). D’Imprima et. al expressed His-tagged PilQ in *T. thermophilus*, isolated the membranes, and purified TtPilQ in DDM. D’Imprima et. al observed flexibility between the N-terminal domains in TtPilQ.

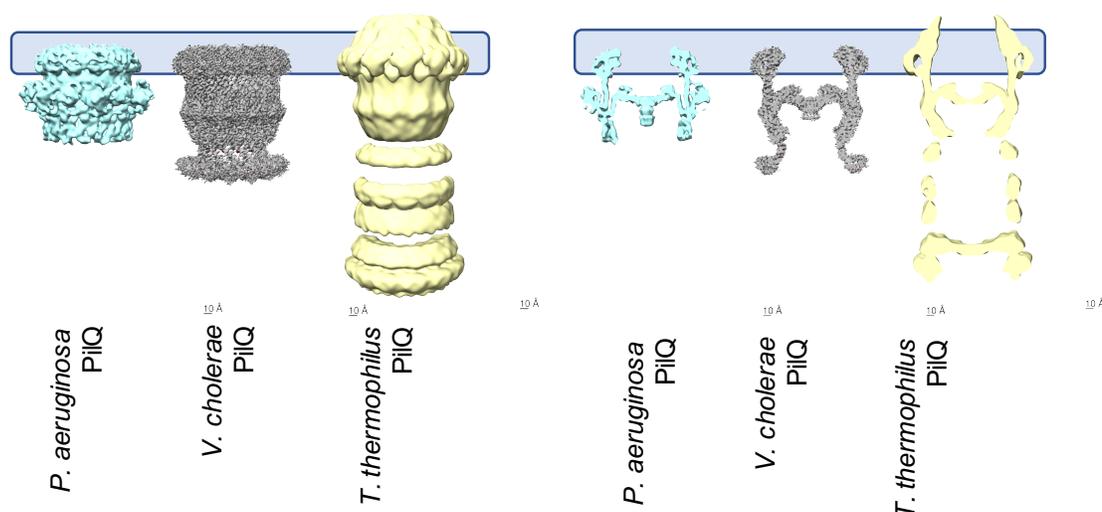


Figure 14 Structures of Type IV Pilus Secretins.

Comparison of cryoEM density of PilQ structures from *Pseudomonas aeruginosa* (left) (Koo et al. 2016), *V. cholerae* (center)(solved here), and *Thermus thermophilus* (D’Imprima et al. 2017).. CryoEM densities are shown from the side (left) and cut through the center (right). The membrane is blue. Scale, 10 Å.

Recently, several high resolution cryoEM structures of bacterial T2SS and T3SS secretins have been published, including *E. coli* GspD (Yan et al. 2017; Hay et al. 2018; Yin, Yan, and Li 2018), *V. cholerae* GspD (Yan et al. 2017; Yin, Yan, and Li 2018), *Pseudomonas*

aeruginosa XcpQ (Hay, Belousoff, and Lithgow 2017), *Klebsiella pneumoniae* PulD (Chernyatina and Low 2018), *A. hydrophila* (pilotin-independent ExeD), *V. vulnificus* (pilotin-dependent EpsD)(Howard et al. 2019), and *Salmonella typhimurium* (InvG)(Worrall et al. 2016). With the exception of *E. coli* K12 GspD, *P. aeruginosa* XcpQ, and *S. typhimurium* InvG, all of the secretin structures were solved after recombinant expression in *E. coli*. Our work expressing His-tagged *V. cholerae* PilQ in *V. cholerae* cells allowed us to demonstrate that our secretin was functional before purification.

While structures from the T2SS and T3SS secretins have revealed much about the secretin superfamily, they do not explain the peculiarities of the Type IV competence pilus secretins. To date, the highest resolution structures of Type IV competence pilus secretins were insufficient to build an atomic model (Koo et al. 2016; D’Imprima et al. 2017). Here we solve the structure of the *Vibrio cholerae* Type IV competence pilus secretin PilQ to 3 Å using a fully-functional, tagged variant that was expressed and purified from the native bacterium. Our work highlights differences between the Type IV competence pilus, the T2SS, and the T3SS secretins, and emphasizes the need for structures of different secretin family members. We demonstrate differences in the outer membrane spanning region, and provide insight into how the Type IV competence pilus could be accommodated within PilQ during natural transformation.

Results and Discussion

Purification of *V. cholerae* PilQ in amphipol

To ensure properly folded and fully-function Type IV competence pilus machinery, we chose to purify PilQ from *Vibrio cholerae* rather than a recombinant system. A chromosomal mutation was made to add a ten-histidine tag to the N-terminus of PilQ. The bacteria retained wildtype levels of natural transformation (**Figure 15**). Previous work demonstrated that similar N-terminal tags resulted in functional Type IV competence pili (Ellison et al. 2018).

Late-log chitin-independent transformation assay

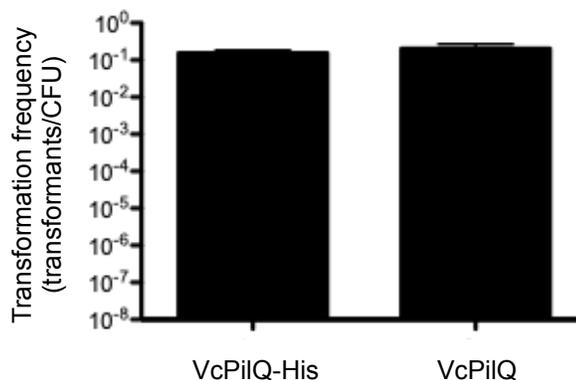


Figure 15 His-tagged PilQ *V. cholerae* transformation assay.

The strain of *V. cholerae* expressing the His-tagged variant of PilQ was subjected to a transformation assay to ensure that it was functional. Transformation frequency was measured by the ratio of transformants to colony forming units (CFUs).

Due to biosafety constraints, the *V. cholerae* cells expressing PilQ were lysed in the presence of the surfactant n-Dodecyl β -D-maltoside (DDM). After affinity purification, PilQ was exchanged into an amphipol environment. Amphipols are amphipathic polymers that can interact with hydrophobic protein surfaces and partition membrane proteins from complex mixtures (Tribet, Audebert, and Popot 1996; Popot 2010).

Secretins are detergent- and heat-resistant multimers (Chami et al. 2005; Guilvout et al. 2011). SDS page analysis of PilQ in amphipol showed a high molecular weight band near the top of the gel (marked with an asterisk in **Figure 16A and C**). Western blotting against the histidine tag on PilQ identified the high molecular weight band as PilQ (marked with an asterisk in **Figure 16B and D**). The gel electrophoresis also demonstrated some low molecular weight species in the protein purification. Some fragments show up on the Western blot, so they must be PilQ monomer or proteolyzed PilQ species that retained the N-terminal histidine tag. Others are not labeled in the Western, so they may represent PilQ with the N-terminal histidine tag degraded, or they may be contaminant proteins. Regardless, the size difference between these contaminants (<100 kDa) and the PilQ

multimer (~860 kDa) made it easy to distinguish PilQ from the milieu in electron micrographs (**Figure 17**).

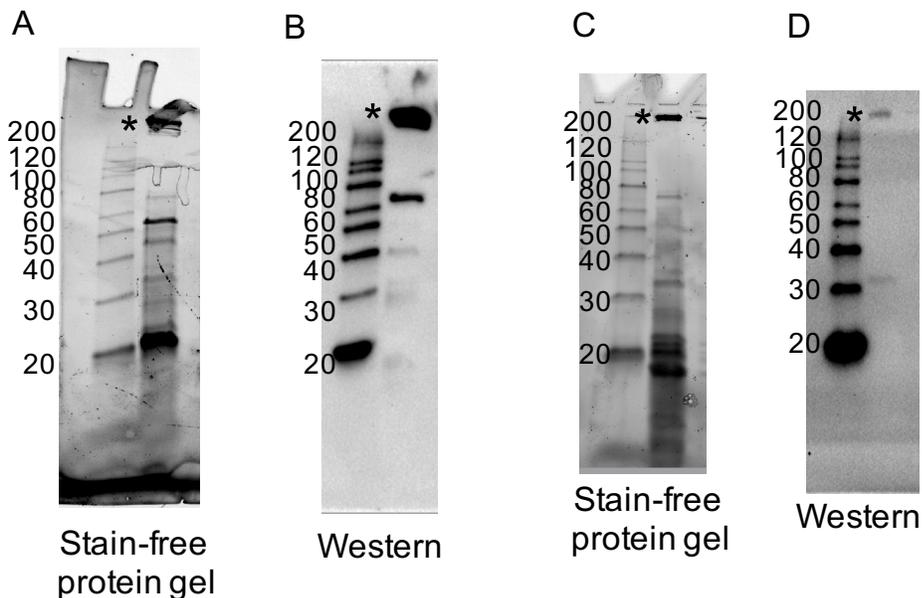


Figure 16 Protein electrophoresis of VcPilQ in amphipol.

Gel electrophoresis analysis of VcPilQ in amphipols by coomassie stain (A), (D), (E), and western blotting against anti-His tag (C), (F). The multimer band is indicated with an asterisk.

To confirm that full length PilQ was present in our samples (residues 30-571 after the N-terminal signal peptide (residues 1-29) is cleaved), we used gel band analysis with a trypsin digest and mass spectrometry to analyze the multimeric PilQ in amphipol from a SDS page gel (Band marked with an * in **Figure 16C** was analyzed). The results demonstrated 65% sequence coverage, with fragments identified in each domain of the folded protein (**Figure 34**).

Single particle cryoEM of the Type IV competence pilus secretin PilQ

Here we solve the first high-resolution structure of a Type IV competence pilus secretin using single particle cryoEM (**Figure 22**). The cryoEM data processing steps are summarized as a flow chart in **Figure 40** and in Methods.

Briefly, *V. cholerae* PilQ (VcPilQ) was plunge-frozen on Quantifoil Holey Carbon Grids with R2/2 spacing. This allowed us to capture multiple images per hole during data collection on a Titan Krios electron microscope. 3,808 movies were collected over the course of 15 hours using a three by three pattern using beam-image shift (A. Cheng et al. 2018). This collection scheme often hits the crystalline ice close to grid bars. The movies were motion corrected and screened for quality, which left 2,510 micrographs.

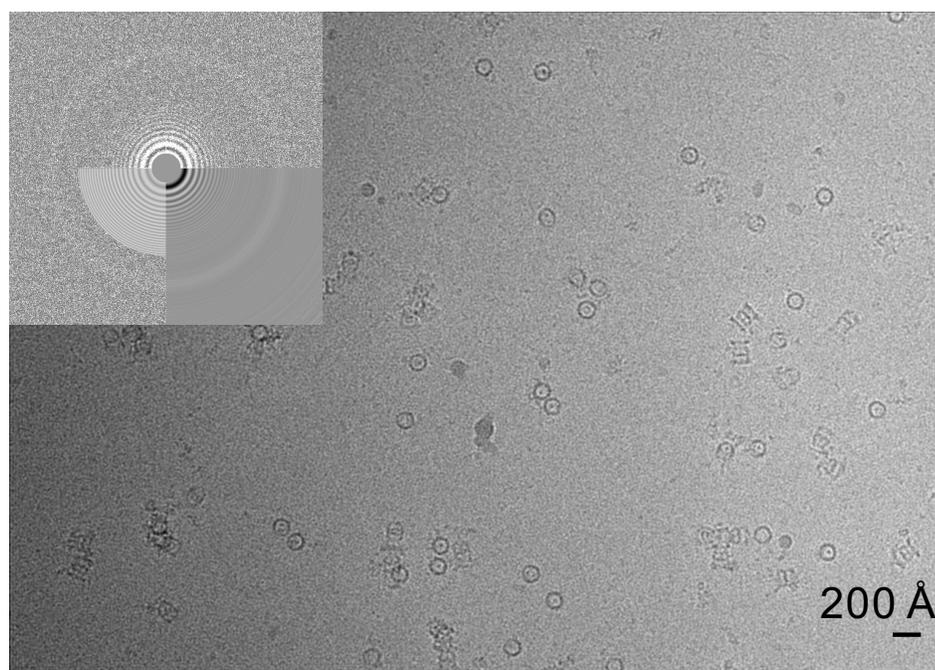


Figure 17 Example micrograph.

Representative micrograph with a 10 Å low pass filter applied. A representative Fourier transform and contrast transfer function (CTF) fit of each micrograph is shown in the upper left corner. Scale bar, 200 Å.

The cryoSPARC blob picker was used to identify 3,100,353 putative particles, which were inspected to select the top 252,319 particles for 2D classification. For VcPilQ, two-dimensional (2D) classification demonstrates a wide variety of orientations, the clear presence of secondary structure (**Figure 18**), and C14 symmetry (**Figure 19**).

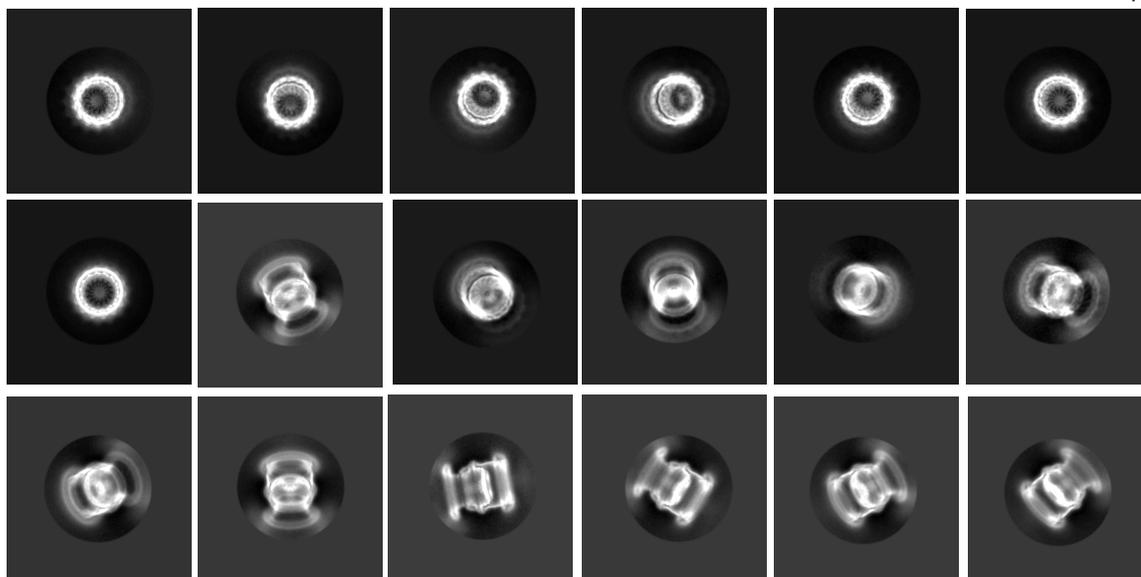


Figure 18 VcPilQ 2D classes.

Representative 2D classes calculated in Relion. The 2D class box is 441.6 Å wide.

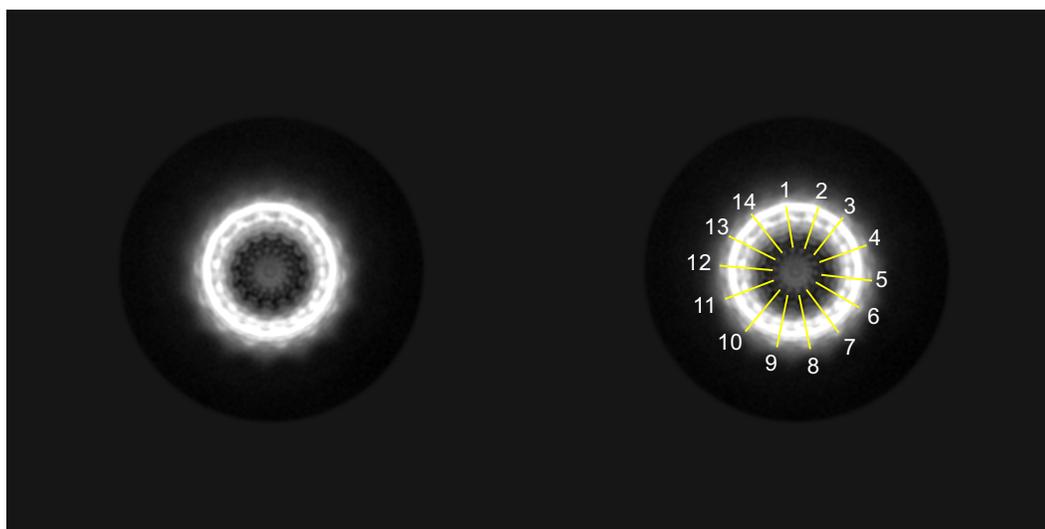


Figure 19 VcPilQ 2D class demonstrates C14 symmetry.

A top view 2D class from demonstrates that VcPilQ has C14 symmetry. Yellow lines point to inner spokes in the PilQ gate. Each monomer is numbered in white.

Several rounds of 3D classification and 3D refinement were used to select 100,543 particles for further analysis in Relion. After polishing and CTF Refinement, the cryoEM structure (C14 symmetry, overall resolution FSC@0.143 2.7 Å, FSC@0.5 3 Å, **Figure**

20C) reached near-atomic resolution and demonstrated isotropic resolution in the X, Y, and Z axes (3DFSC sphericity 0.96) (**Figure 35**)(Tan et al. 2017).

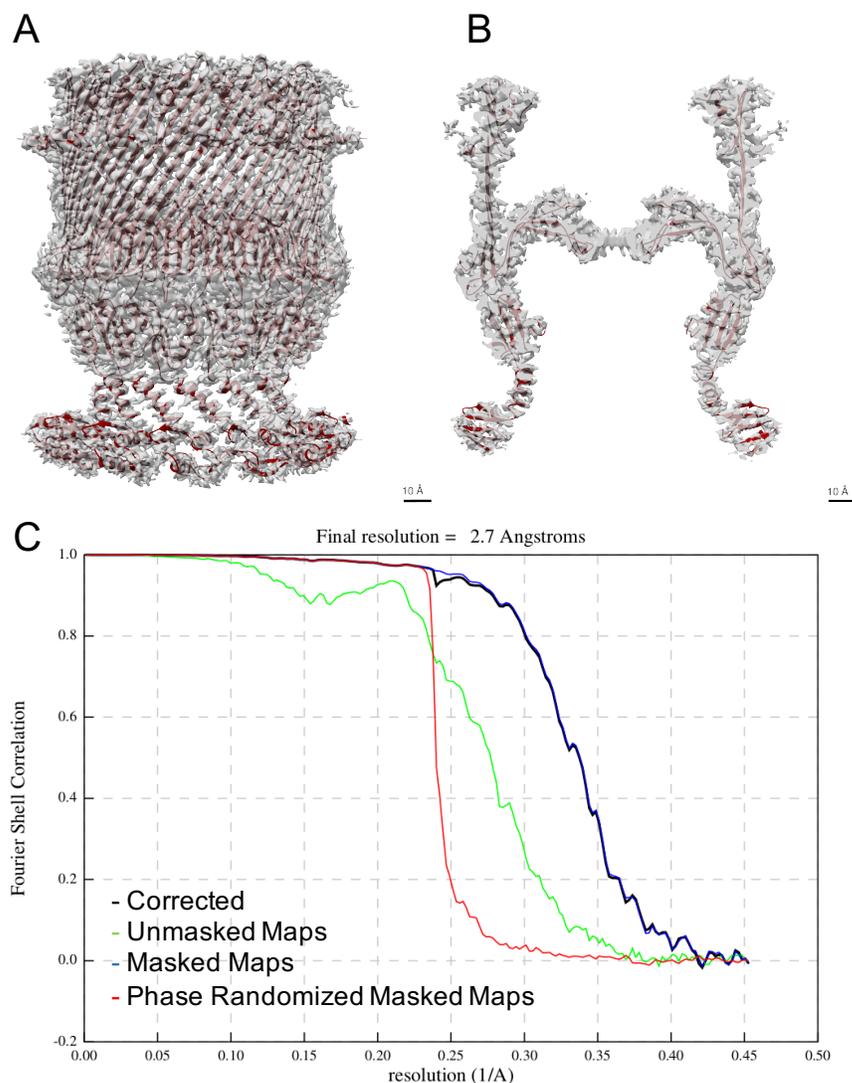


Figure 20 CryoEM density of VcPilQ.

Model and cryoEM density of symmetrized PilQ is shown from the side (A) and cut through the center (B). Scale, 10 Å. (C) Fourier shell correlation calculated in `relion_postprocess` with plots of Corrected Fourier Shell Correlation (black), Unmasked Fourier Shell Correlation (green), Masked Fourier Shell Correlation (Blue), and Corrected Fourier Shell Correlation of phase randomized masks (red).

Typically, the density must have a resolution greater than 3.2 Å to unambiguously identify the protein sequence (**Figure 3**). We modeled residues 160 to 571 of PilQ. Our mass

spectrometry results (**Figure 34**) demonstrated the presence of residues between 50 and 567. VcPilQ has one AMIN domain (residues 54-125) that is thought to interact with the peptidoglycan in the periplasm(de Souza et al. 2008). While we did not resolve the AMIN domain in our PilQ structure, it is present in our mass spectrometry (**Figure 34**). Additionally, hazy density is present in the 2D classification in the region we would expect to see the AMIN domain (marked with asterisks in **Figure 21**). The residues following the AMIN domain (residues 126-159) are predicted to be unstructured (homology modeling in **Figure 44**). If that region is unstructured, it's a likely source of the conformational heterogeneity we observe as a hazy density in **Figure 21**. My focused classification calculations were unable to resolve additional residues between 1 and 160 (data not shown). Because the AMIN domain binds the peptidoglycan it is also possible that *in situ*) this linker region would demonstrate a conformational preference. In *Neisseria meningitidis*, the lipoprotein PilP is thought to bind the AMIN domain and act as a bridge between the inner and outer membrane components of the Type IV pilus machinery(Balasingham et al. 2007).

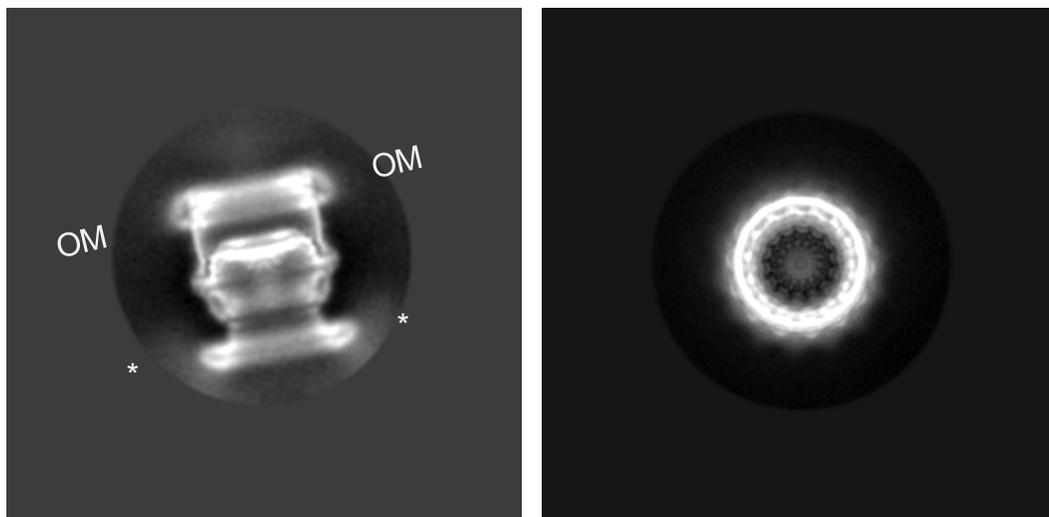


Figure 21 2D classification demonstrates hazy AMIN domain region.

Example 2D classes show PilQ from the side (left) and from the top (right). The outer membrane is indicated with OM. The putative AMIN domains are indicated by asterisks.

The structure of VcPilQ demonstrates similarities to other secretins

In each PilQ monomer, four beta strands come together to form a beta sheet (**Figure 22A**, left). Once assembled, PilQ forms a 56 strand double-walled beta-barrel (**Figure 42**). The inner barrel contains two beta hairpin (beta strand-turn-beta strand) gates (**Figure 41**). These regions agree well with the topology of published secretin structures (**Figure 23**). I used Chimera MatchMaker to align several published secretin structures to VcPilQ and calculate the root-mean-square deviation (RMSD) of each pair (Pettersen et al. 2004) (**Figure 23**). In each case, when the RMSD is calculated over the protein domains present in both structures (N3 and secretin domains only), the RMSD is about 1 Å (**Figure 23**). While most of the Secretin domain of VcPilQ agrees with T2SS and T3SS structures, the outer membrane region is significantly different in both angle and membrane spanning distance.

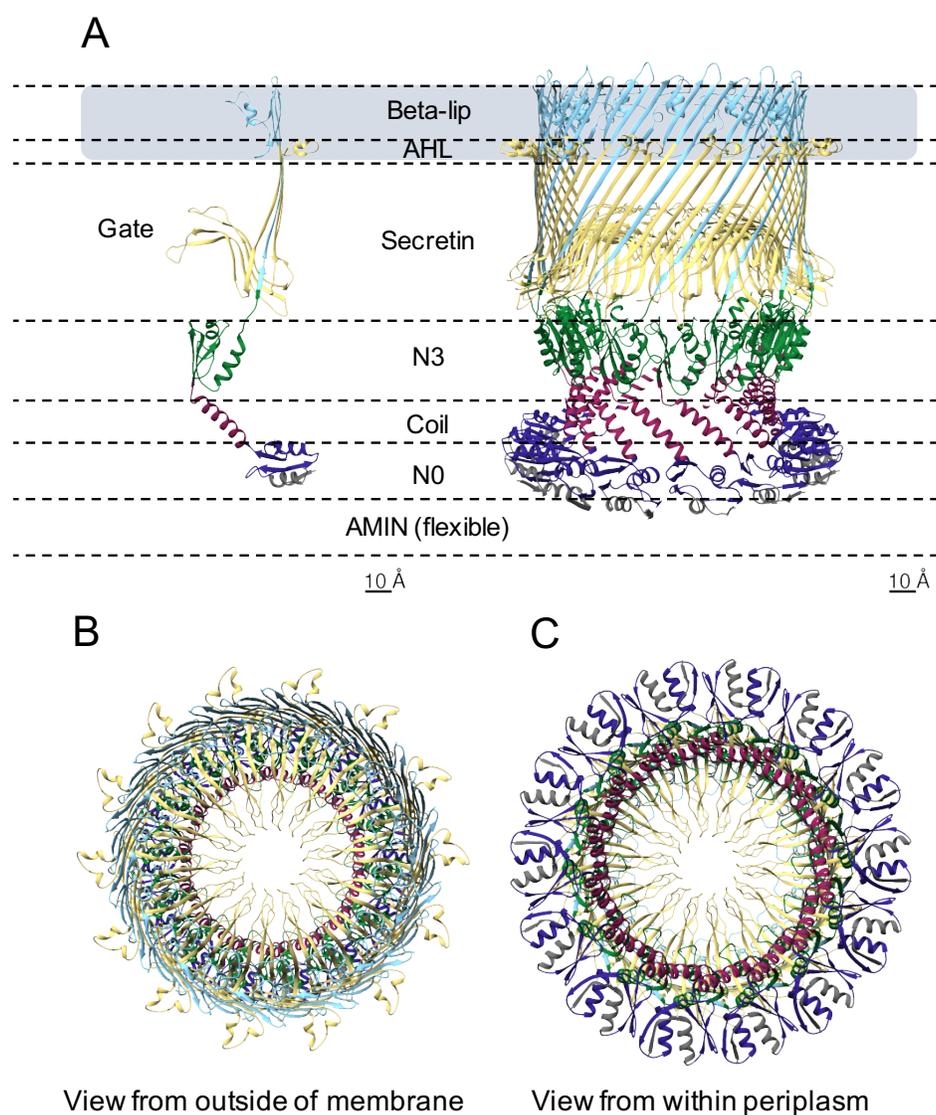
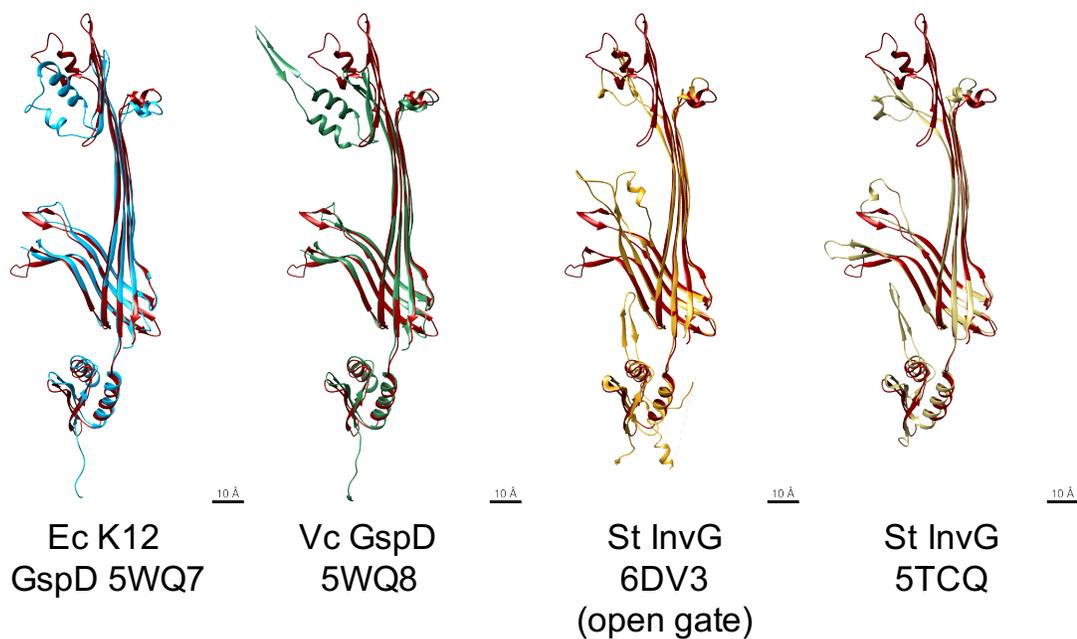


Figure 22 Atomic model of VcPilQ.

The atomic model of *V. cholerae* PilQ is shown in dark grey. One chain is highlighted within the grey multimer (left) and shown by itself (right). The outer membrane is represented in a light blue rectangle. Dashed lines represent the different domains of the PilQ structure (AMIN, N0, coil, N3, Secretin, Beta-lip). A central slice of the atomic model of *V. cholerae* PilQ. The inner cavity distances are depicted with dashed lines and labeled in ångstroms. Scale, 10 Å.

Vc PilQ (maroon) RMSD vs



VcPilQ Compared to	RMSD (Å)
<i>E. coli</i> K12 GspD (5WQ7)	1.2
<i>V. cholerae</i> GspD (5WQ8)	1.2
<i>S. typhimurium</i> InvG (6DV3, open gate)	0.97
<i>S. typhimurium</i> InvG (5TCQ)	1.1

Figure 23 RMSD of VcPilQ versus published secretin structures.

The monomer structure of VcPilQ is aligned to the *E. coli* K12 GspD (5WQ7), the *V. cholerae* GspD (5WQ8), and the *S. typhimurium* InvG (6DV3 and 5TCQ) (Worrall et al. 2016; Yan et al. 2017; J. Hu et al. 2018). Only the N3 and Secretin domains were compared. RMSD was calculated with MatchMaker in Chimera.

The N0 domain in VcPilQ agrees with previous structures

Only one of the seven T2SS secretin, and one of the T3SS structures resolves the N0 domain. Here, we can resolve the N0 domain (residues 160 to 227) of PilQ to 4 to 7 Å local resolution, which allowed us to build in a model based on homology to previously solved structures of N0 domains (**Figure 27**). A crystal structure of isolated N0 domain from *Neisseria meningitidis* was published (4AR0)(Berry et al. 2012) (**Figure 24A and B**). The N0 domain is resolved in the T3SS secretin InvG (**Figure 24C**)(Worrall et al. 2016) and in the cross-linked *Klebsiella oxytoca* PulD outer membrane complex structure (**Figure 24D**)(Chernyatina and Low 2019). Our VcPilQ N0 structure falls within 1 Å RMSD of each of the previous structures.

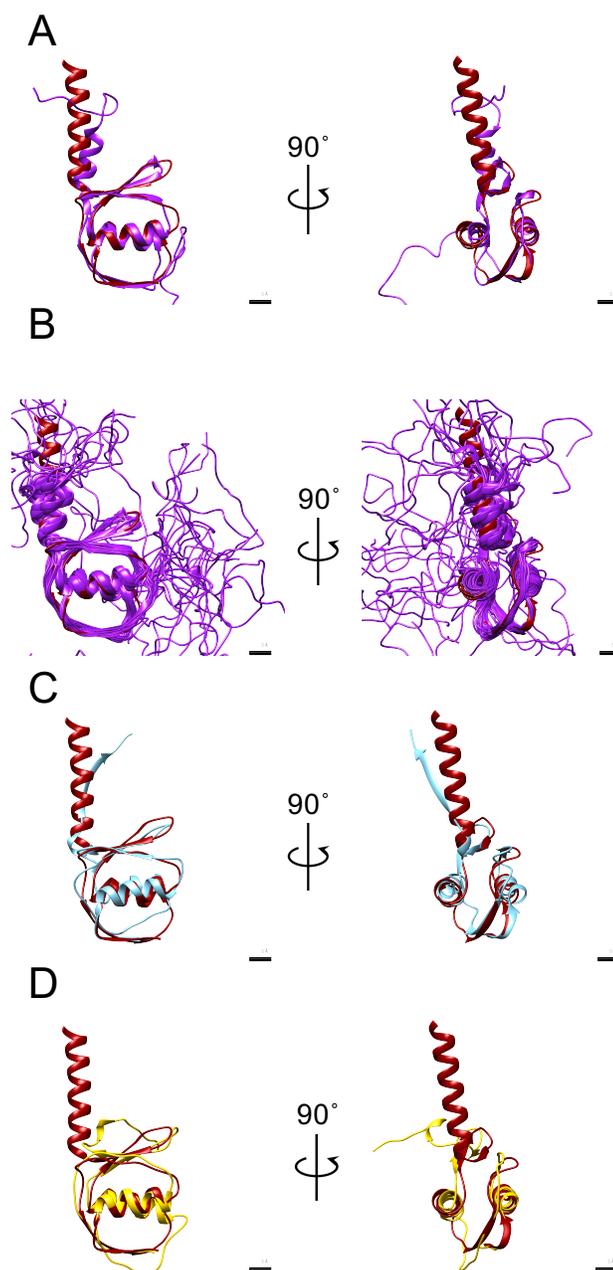


Figure 24 Alignment of VcPilQ N0 domain compared to others.

The VcPilQ coil and N0 domain (red) were compared to previously published N0 structures. The *N. meningitidis* PilQ N0 NMR structure (purple) is shown fit to the VcPilQ N0 (red) with the smallest RMSD structure (0.9 Å) (A) and with all structures (RMSD 0.9 to 1.1 Å) in (B). (C) *S. typhimurium* InvG (PDB 6DV3), RMSD 1.1 Å. (D) The N0 domain of *Klebsiella oxytoca* PulD (PDB ----, personal communication) RMSD 1.1 Å (Berry et al. 2012; Chernyatina and Low 2019; Worrall et al. 2016). Scale, 10 Å.

VcPilQ uses a novel helical coil to transition into the N3 domain

In VcPilQ, an alpha-helical coil follows the N0 domain (**Figure 22**). In the T2SS and T3SS secretins, the periplasmic protein domains are linked by unstructured loops. None of the T2SS or T3SS structures contain a helical coil to link N-terminal domains (**Figure 25**). The homology model of VcPilQ we built shows a bit of this alpha helix in some of the models, but the majority closely match the previously published T2SS and T3SS structures (**Figure 44**). ConSurf protein conservation analysis comparing VcPilQ to other Type IV competence pilus secretins suggests that this coil region is moderately conserved (**Figure 38**).

Next, the protein chain abruptly changes direction ($\sim 104^\circ$ angle) as the coil flows into the N3 domain (**Figure 25**). This dramatically reduces the channel diameter, from 90 Å at the bottom of the N0 domain to 60 Å across the N3 domain (**Figure 31**). In the T2SS and T3SS structures, the diameter of the channel is relatively constant.

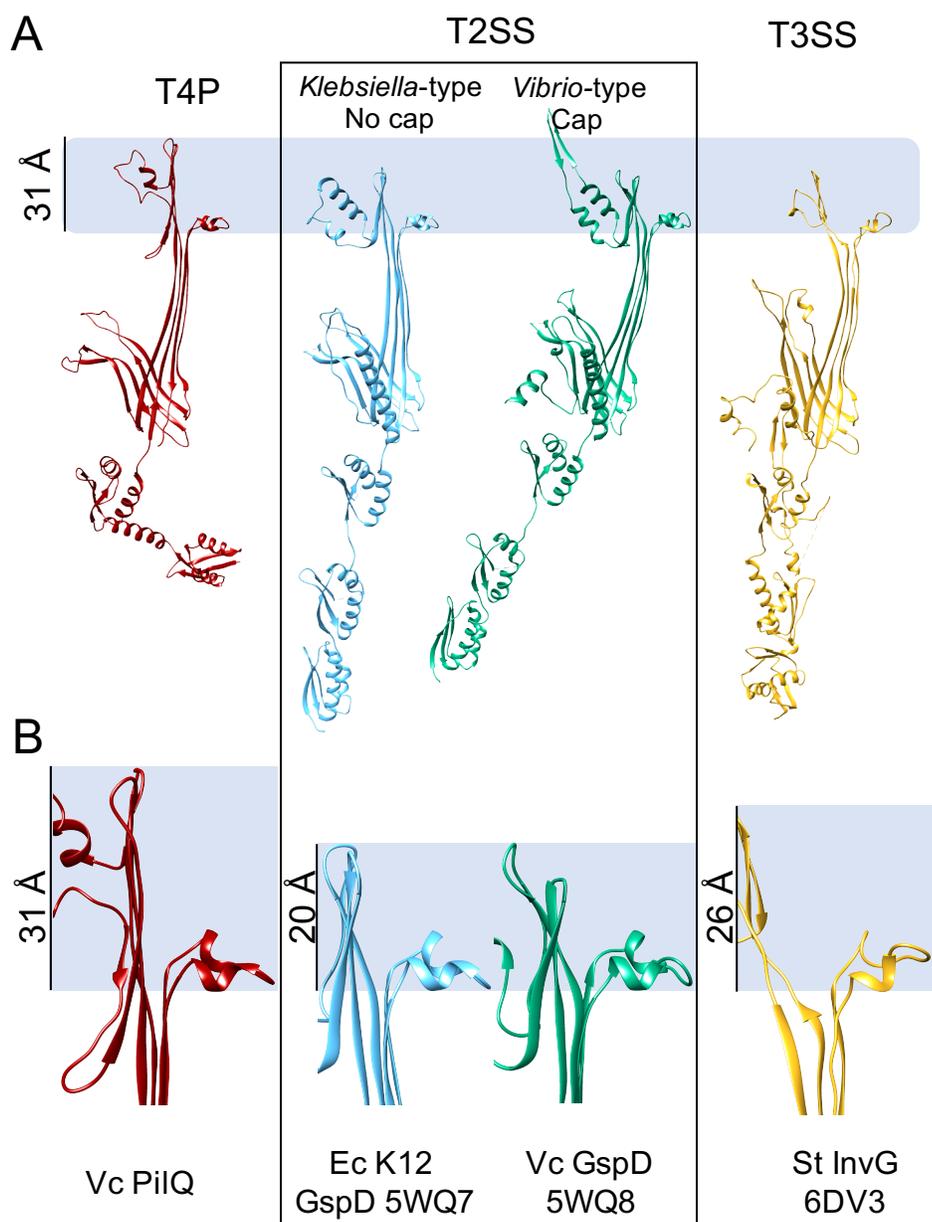


Figure 25 Secretin monomer atomic model.

Comparing the structure of *V. cholerae* PilQ (left, dark red) solved here with other secretin structures. Two example T2SS structures are shown: the *Klebsiella*-type *E. coli* K12 GspD (5WQ7) (blue) and the *Vibrio*-type *V. cholerae* GspD (5WQ8) (green). On the left, the T3SS secretin InvG from *Salmonella typhimurium* (6DV3) (yellow). The membrane is blue rectangle. The bottom half of the figure focuses on the outer membrane region of each structure.

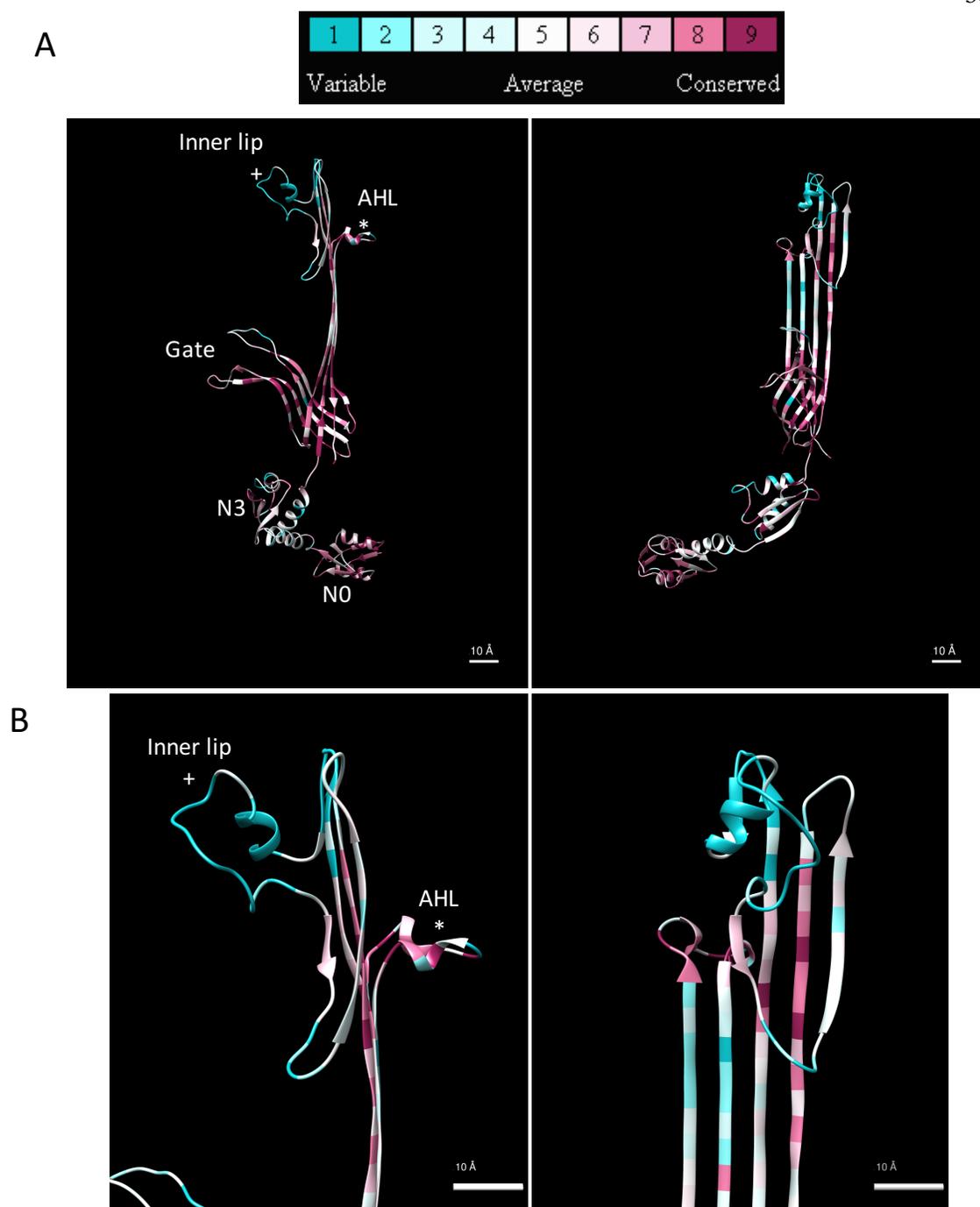


Figure 26 Conservation analysis of *V. cholerae* PilQ focusing on different protein domains.

Conservation of Type IV competence pilus machine residues in *V. cholerae* PilQ (**A**) ConSurf analysis identified if PilQ residues are variable (cyan), average (white), or conserved (dark magenta). This analysis was mapped onto a ribbon diagram of PilQ. The left image shows the PilQ monomer with features labeled (Inner lip with a +, Amphipathic helix lip (AHL) with a *, Gate, N3, and N0). The image on the right is a 90° rotation about the long axis. Scale, 10 Å. (**B**) The images from Figure A zoomed in to focus on the inner lip and AHL region.

When I started this project, I predicted the structure of VcPilQ using I-TASSER homology modeling (Y. Zhang 2008). Homology modeling for VcPilQ yielded five top structures (**Figure 44**). These predictions are based off of previously solved structures, so they can be sorted by their similarity to T2SS GspDs (with and without the cap) or the T3SS InvG. After solving the structure, it is clear that the putative outer membrane region of VcPilQ is significantly different (**Figure 25**). Because no structure of the Type IV Pilus secretin had been deposited, none of the homology models accurately predicted the thicker outer membrane domain of VcPilQ, or the extent of the alpha helix between the N0 and N3 domains.

The bulk of the secretin domain (**Figure 39**) is conserved among Type IV competence pilus secretins. Additionally, VcPilQ secretin domain shows a fold strong agreement with the published T2SS and T3SS structures (**Figure 23**, **Figure 25**). Additionally, the periplasmic side of the secretin chamber is negatively charged in VcPilQ, VcGspD and EcK12GspD (**Figure 43**). The outer surface of PilQ demonstrates bands of neutral, negative, and positive electrostatic potential (**Figure 27**). The inner lip of PilQ is highly negatively charged. The inner lip residues in VcPilQ demonstrate high variability (residues 347 to 385, **Figure 26**). In the *V. cholerae* GspD and *E. coli* K12 GspD structures, the inner lip has alternating negatively- and weakly positively-charged regions (**Figure 43**). These differences may be related to the specific function of the Type IV competence pilus secretin and the protein exporting secretins like GspD.

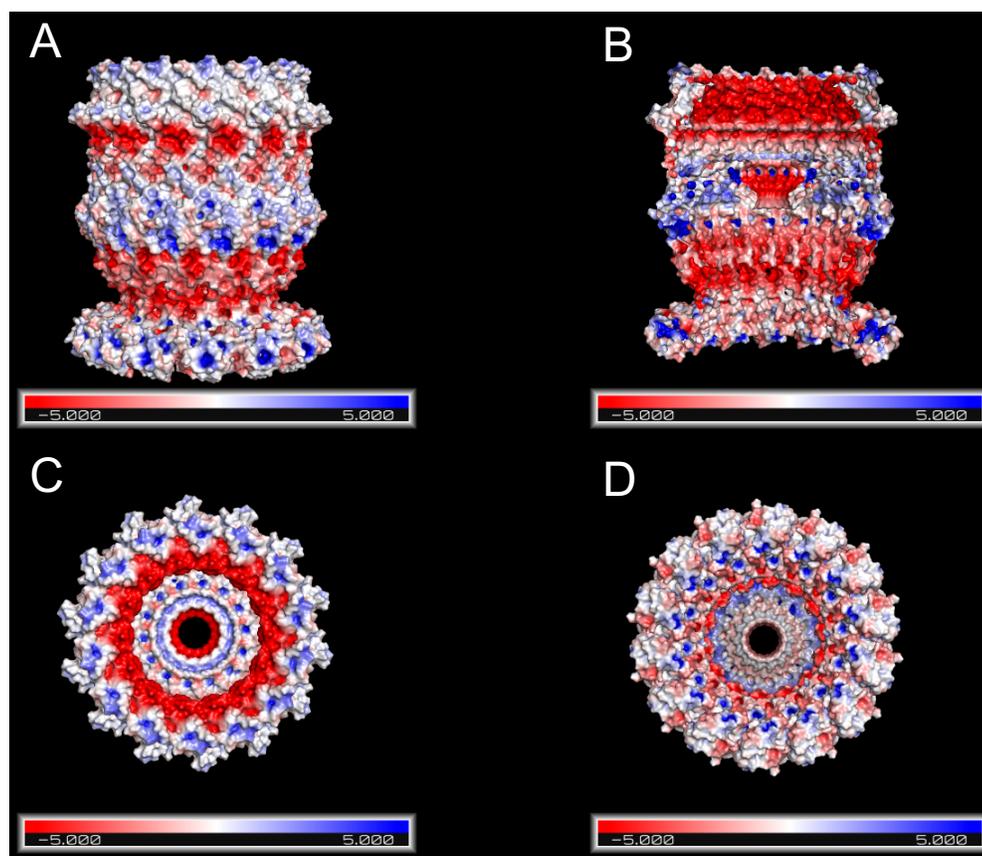


Figure 27 Electrostatic potential surface of VcPilQ.

PyMol Adaptive Poisson-Boltzmann Solver electrostatic potential calculation of the protein surface and center(Jurru et al. 2018).

The putative outer membrane region of VcPilQ is thicker than T2SS secretins

Bilayer membranes have three chemically-distinct regions: the polar membrane surface, the non-polar membrane interior, and the mid-polar region that makes up the interface between the other two(Hanshaw, Stahelin, and Smith 2008). A 2011 paper calculated the average hydrophobic belt diameter of 22 to 24 Å for all known bacterial beta-barrel outer membrane proteins (Lomize, Pogozeva, and Mosberg 2011). This hydrophobic belt would fall into the non-polar membrane interior. The bacterial outer membrane has an asymmetric lipid composition, with the extracellular leaflet dominated by lipopolysaccharide, while the periplasmic leaflet contains phospholipids, so the size of the membrane's mid-polar interfacial region is related to the specific lipid identity of a given membrane(Pogozeva, Mosberg, and Lomize 2014).

The secretin amphipathic helix lip (AHL) is thought to be a key determinant for secretin outer membrane insertion (Majewski, Worrall, and Strynadka 2018). Among Type IV pilus secretins, the AHL is conserved (**Figure 39**). The AHL is thought to mark the lower boundary of the outer membrane region of secretins (Worrall et al. 2016; Majewski, Worrall, and Strynadka 2018; Howard et al. 2019). The upper boundary of the outer membrane is somewhat ambiguous, with some scientists basing it off of the micelle in cryoEM density, while others use the position of aromatic residues (Hay et al. 2018).

In our VcPilQ atomic model, if we measure from the bottom of the AHL to the top of the beta strands, the putative outer membrane thickness is about 3 nm (**Figure 25**). As we saw in **Figure 25**, the putative outer membrane region of VcPilQ is about 10 Å taller than the same region in the T2SS structures previously published. To investigate more than just the residue locations, we generated a density map based only on the atomic model of our VcPilQ, and for two published secretins: the *E. coli* K12 GspD (EMD-6675, 5WQ7) and the *V. cholerae* GspD (EMD-6676, 5WQ8) (Yan et al. 2017). We created an inverted mask based on the atomic model cryoEM density and subtracted it from the empirical cryoEM density. As a result, we can visualize density (here called the “unmodeled density map”) in the cryoEM map that is not accounted for by the atomic model. The micelle density (in grey) blooms around the putative outer membrane region of the protein (**Figure 28** and **Figure 29**). This density appears both on the outside of the beta barrel, and inside, coating the inner lip of PilQ.

The bacterial outer membrane is about 4 nm thick, but that thickness can be attenuated by the presence of different integral membrane proteins (Phillips 2017). Based purely on atomic models, both our VcPilQ and the published T2SS structures have putative outer membrane regions that are shorter than this 4 nm membrane thickness estimate. This left us with two questions. First, why is the VcPilQ putative outer membrane region thicker than in the T2SS structures? Second, why don't any of these putative outer membrane regions span the entire 4 nm outer membrane?

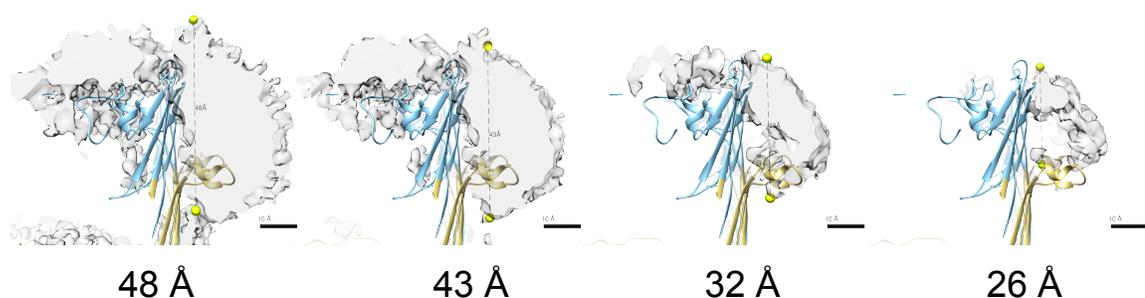


Figure 28 VcPilQ micelle density with putative outer membrane diameter labeled in Å. The non-protein cryoEM density of VcPilQ is shown in grey at four different Chimera thresholds (left to right). The atomic model of VcPilQ is shown in blue and yellow according to the color scheme of **Figure 22**. In each image, yellow spheres mark the approximate top and bottom of the micelle density. The distance between the yellow spheres (dashed line) is shown below each image in Ångstroms. Scale, 10 Å.

Why is the VcPilQ putative outer membrane region taller than in T2SS secretins?

To address the differences between VcPilQ and the T2SS secretins, we first asked how conserved the outer membrane region of VcPilQ is. If the beta-strand region (that is taller in VcPilQ than in GspD structures) was not conserved, we would wonder if our structure is an anomaly. So we overlaid our previous ConSurf analysis (comparing VcPilQ to other Type IV pilus secretins and excluding T2SS secretins) with our unmodeled density map (**Figure 29**). The AHL and the beta strands are relatively conserved in VcPilQ (maroon in **Figure 29**), but the inner lip region displays high variability (cyan in **Figure 29**). Thus, the most variable part of the outer membrane region is found in the inner lip. The inner lip is unlikely to be a determining factor in the membrane diameter of the VcPilQ. It is probably variable because it is specific to each Type IV pilus and its cargo.

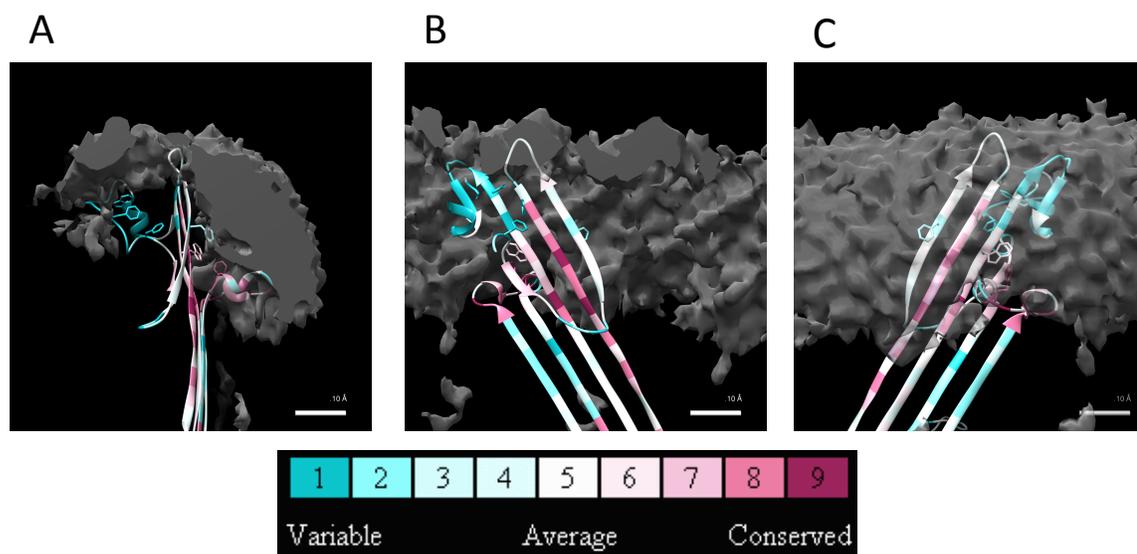


Figure 29 Micelle density of VcPilQ with ConSurf protein structure.

The atomic model of *V. cholerae* PilQ is colored by residue conservation, where the gradient from cyan to maroon represents a scale of variability to conservation. The grey isosurface represents the membrane micelle observed in the cryoEM structure. Three views are shown focusing on the outer membrane region.

Why don't any of the putative outer membrane regions span the entire 4 nm outer membrane?

The bacterial outer membrane is thought to be around 4 nm thick (Phillips 2017). The VcPilQ putative outer membrane region ranges from about 25 to 50 Å depending on where you mark the inner and outer face (Figure 28). But VcPilQ and all of the published T2SS secretin structures are purified proteins.

Recently, Chang et al. undertook an exhaustive cryoelectron tomography (cryoET) study of the *Myxococcus xanthus* Type IVa Pilus (T4aP) system *in situ* (Chang et al. 2016). *M. xanthus* glides across surfaces using cycles of extension and retraction of the T4aP. The *M. xanthus* secretin is called PilQ (MxPilQ) and shares approximately 20% identity with the *V. cholerae* PilQ (VcPilQ). Protein domain analysis and Phyre2 protein modeling of MxPilQ suggests a similar structure to VcPilQ (Figure 30A-B) (Kelley et al. 2015; Adebali, Ortega, and Zhulin 2015). Docking our VcPilQ structure into the sub-tomogram average of *M.*

xanthus Type IVa pilus (non-piliated) (EMD-3257, green in **Figure 30C**) and of the Δ PilP strain (EMD-3249, purple in **Figure 30D**) shows a similar overall architecture.

The Δ pilP sub-tomogram average was selected because biochemistry experiments suggested that only PilQ and TsaP are present in the T4aP machinery (Chang et al. 2016). The putative location of TsaP suggested in Chang *et al.* 2016 is marked with an asterisk in **Figure 30**. In *M. xanthus*, TsaP is a peptidoglycan-binding protein (Siewering et al. 2014). *M. xanthus* has two AMIN binding domains that are also thought to bind the peptidoglycan layer (**Figure 30A**) (de Souza et al. 2008). In *V. cholerae*, the TsaP homolog LysM has not been implicated in Type IV competence pilus function. VcPilQ has one AMIN domain to anchor it to the peptidoglycan layer.

The gate region of PilQ settles nicely into place in the *M. xanthus* secretin, which suggests we positioned the VcPilQ structure correctly. However, the putative outer membrane region in VcPilQ does not fill the entire outer membrane region of the MxT4aP sub-tomogram average. The thickness of the outer membrane in the sub-tomogram average is about 60 Å, whereas the amphipol micelle region in VcPilQ seems to be between 25 and 50 Å thick (**Figure 25, Figure 30**). However, this discrepancy may be partially explained by the observation that the outer membrane is about 5 nm thick in *M. xanthus* and 4 nm thick in *V. cholerae* (Phillips 2017).

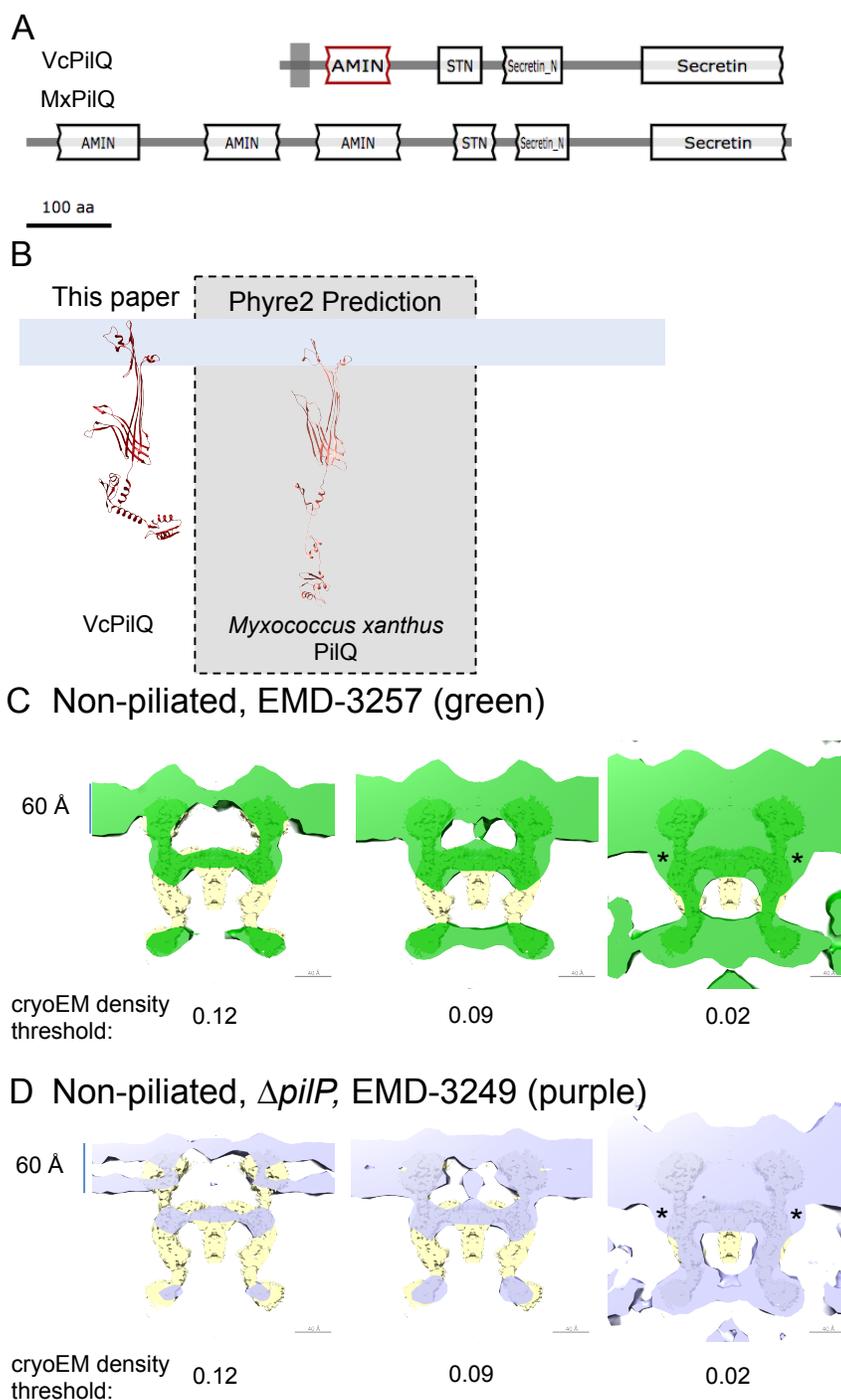


Figure 30 VcPilQ fit into sub-tomogram averaging of *M. xanthus* Type IVa pilus machinery. (A) The domain organization of *V. cholerae* PilQ (top) and *M. xanthus* PilQ (bottom) calculated in CDVist (Adebali, Ortega, and Zhulin 2015). (B) The structure of VcPilQ (solved here) in dark red is compared to the Phyre2 homology modeling prediction of the MxPilQ structure (right, pink)(Kelley et al. 2015). (C) The cryoEM density of the non-piliated *M. xanthus* Type IVa pilus structure (green, C) and in the $\Delta pilP$ strain (purple, D) (Chang et al. 2016). In the $\Delta pilP$ strain,

only PilQ and TsaP are thought to be present in the complex. The putative TsaP location is marked with an asterisk in the rightmost volume in (C) and (D). Each Mx sub-tomogram average structure is depicted at three different Chimera density thresholds (0.12, 0.09, 0.02) to emphasize different features. The structure of the VcPilQ solved here is shown in yellow. The outer membrane diameter is about 60 Å in both (C) and (D). Scale, 40 Å.

Will PilQ accommodate the Type IV competence pilus?

The Type IV competence pilus machinery extends and retracts the Type IV competence pilus through the PilQ outer membrane pore. Here we solved the structure of PilQ in the absence of the pilus and accessory proteins. In our structure, the inner diameter of the PilQ channel ranges from 25 to 108 Å (**Figure 31**).

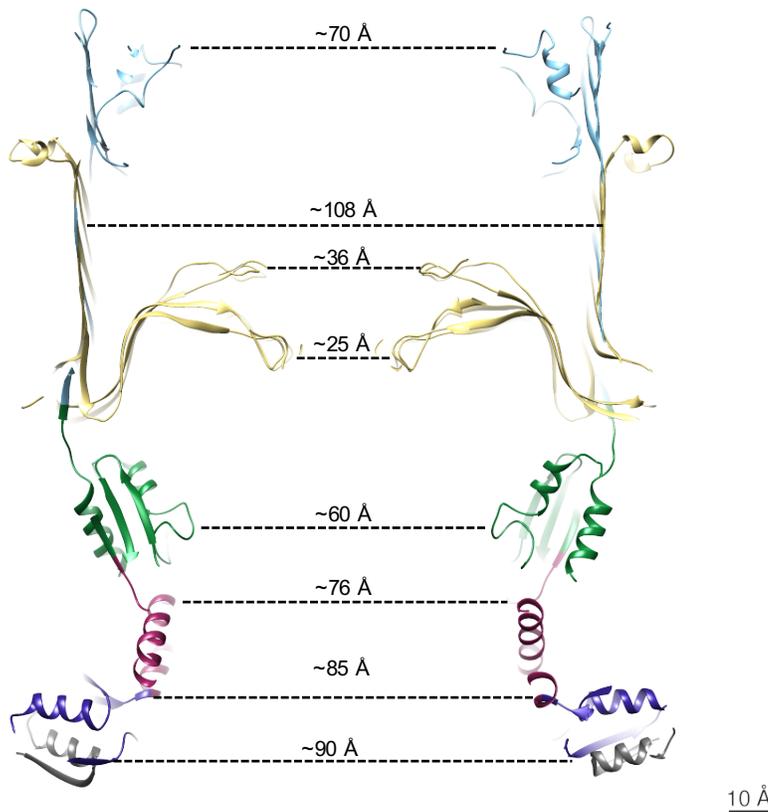


Figure 31 Approximate VcPilQ inner channel diameter distances.

A central slice through the ribbon diagram model of VcPilQ is shown. Dashed lines indicate the inner channel diameter in Angstroms. Scale, 10 Å.

The *V. cholerae* Type IV competence pilus is a Type IVa pilus. T4aP tend to be smaller than T4bP (Craig, Pique, and Tainer 2004). CryoEM structures of several Type IVa Pili (T4aP) have been solved from *N. meningitidis* (NmT4aP, 6 Å resolution, 6 to 7 nm diameter), *N. gonorrhoeae* (NgT4aP, 5 Å resolution, 6 nm diameter), *E. coli* (EHEC strain, EcT4aP, 8 Å resolution, 6 nm diameter), and *P. aeruginosa* (PaT4aP, 8 Å resolution, 5 nm diameter) (Kolappan et al. 2016; Wang et al. 2017; Bardiaux et al. 2019). Additionally, a recent paper on bioRxiv studying Type IV pili in *T. thermophilus* demonstrated the presence of narrow (4.5 nm) and wide (7 nm) pili (Neuhaus et al. 2019). At the time of submission, these structures were not available for download. In **Figure 32**, the VcPilQ structure (in black) is shown with fits of the NmT4aP, NgT4aP, PaT4aP and EcT4aP cryoEM structures. The gate region with diameters of 25 Å at the lower gate and 36 Å in the upper gate clashes with the pilus in each case. VcPilQ was solved in the absence of the pilus, so this gate conformation likely represents a “closed” state in cases where the pilus is absent or fully-retracted. The next narrowest region in the VcPilQ channel is across the periplasmic N3 ring (~60 Å in **Figure 31**). Some of the previously solved T4aP are 6 to 7 nm in diameter, which would be a tight fit in this region. The N3 domain is connected to the secretin domain by a loop, so it could possibly expand to accommodate a 60 Å pilus.

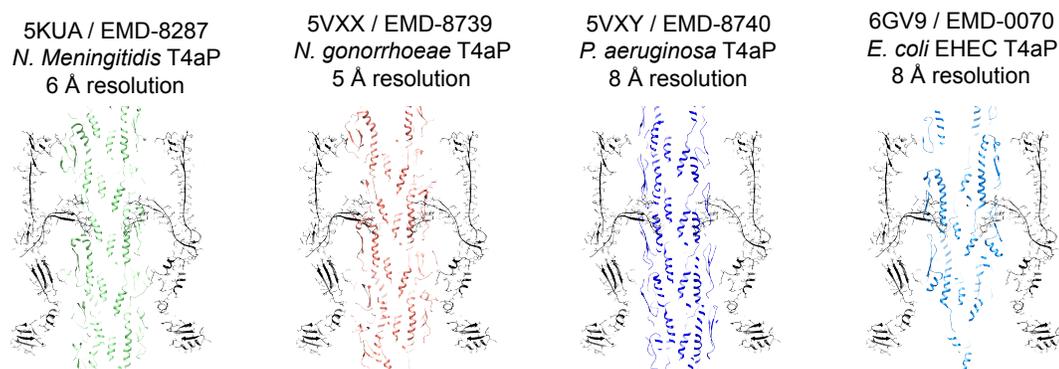


Figure 32 Type IVa pili cryoEM structures fit into VcPilQ.

VcPilQ (black) is shown with models of four different Type IVa pilus structures fit through its central channel. From left to right, the Type IVa pilus structures are from *N. meningitidis*, *N. gonorrhoeae*, *P. aeruginosa*, and *E. coli* (Kolappan et al. 2016; Wang et al. 2017; Bardiaux et al. 2019).

We used homology modeling to predict the VcT4aP monomer structure, but no experimental evidence yet suggests the diameter of the VcT4aP, so we did not attempt to model the pilus into our structure. Because Neuhaus et al. reported both wide and narrow pili based on a pilin subunit structure that is almost identical, we do not feel confident guessing the VcT4aP diameter. Regardless, the gates in VcPilQ would have to move to accommodate a pilus.

The VcPilQ gate region (residues 425-514) is highly conserved (**Figure 39**). In Yan et al., a glycine in the VcGspD gate (G453) was identified as a putative hinge point to facilitate gate opening (Yan et al. 2017). Mutagenesis of glycine 453 to alanine lead to a VcGspD structure in a partially open state. Within Type IV competence pilus secretins, the glycine identified as a hinge in *V. cholerae* GspD (G453A, 5WQ9) is conserved in VcPilQ (G439) (purple box in **Figure 39**). The inner channel distance between glycine 439 alpha carbons is about 8 nm (**Figure 33**). Thus, we hypothesize that a gate hinge mechanism could accommodate a pilus up to 7 nm in diameter.

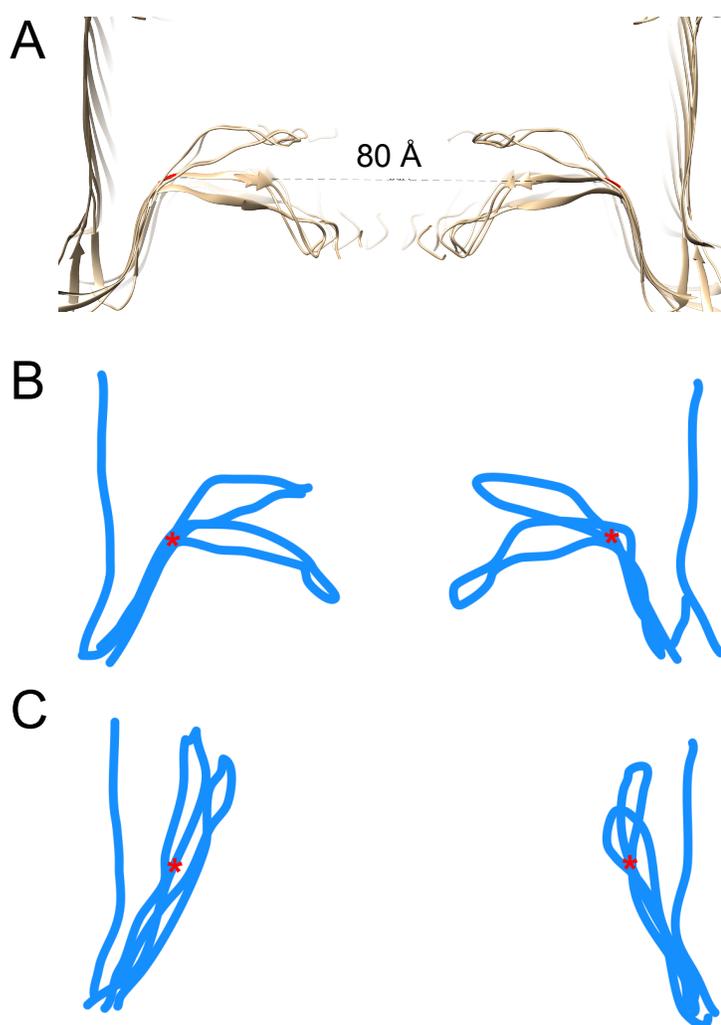


Figure 33 VcPilQ putative gate hinge distance.

The gate region of VcPilQ (A) is shown with the putative hinge residue Glycine 453 highlighted in red. The alpha carbon distance between opposite Glycine 439 residues is 80 Å. The structure from (A) is drawn as a schematic in (B) and a potential gate opening mechanism is shown in (C).

CONCLUSION

Presumably, to function in protein secretion, bacterial secretins must form a continuous channel across the outer membrane. Previous structures of bacterial secretins are only about a third of the way into the outer membrane (2 nm diameter in a 4-6 nm outer membrane in **Figure 25**), which suggests that current experiments do not tell the entire story.

Here we present the first high-resolution structure of a bacterial Type IV competence pilus secretin. We demonstrate that the outer membrane region of VcPilQ is a full 10 Å thicker than previously solved T2SS GspD structures (**Figure 25**). We suggest a structural rearrangement that would transition our closed VcPilQ into a piliated state that could accommodate previously solved structures of Type IV pili. We also compare our structure to previous T4aP sub-tomogram averaging results in *M. xanthus* and observe that our putative outer membrane region isn't thick enough to fill out the sub-tomogram average.

Perhaps an unidentified protein makes up the distance gap, or the structure of these proteins *in situ* is different than in purified structures. It is interesting that our structure of *V. cholerae* PilQ is a full 10 Å taller in the membrane domain than the T2SS and T3SS secretins. Additional structural work is required to address this mystery and fully elucidate the mechanism of bacterial secretins.

By solving the structure of *V. cholerae* PilQ to 3 Å resolution, we observed key differences in the outer membrane region and the periplasmic region. These differences are likely related to the function of PilQ in natural competence. For example, the inner cavity of PilQ is negatively charged (**Figure 27**). The DNA cargo of the Type IV competence pilus is negatively charged, so it's possible that this electrostatic repulsion will help the cargo pass through the cavity, rather than getting stuck. By comparison, the T2SS has a wide variety of protein cargo with varied electrostatic surfaces (**Figure 43**). In the inner cavity of several T2SS secretins, the charge alternates, which could likely accommodate a wider variety of cargo (**Figure 43**).

Appendix

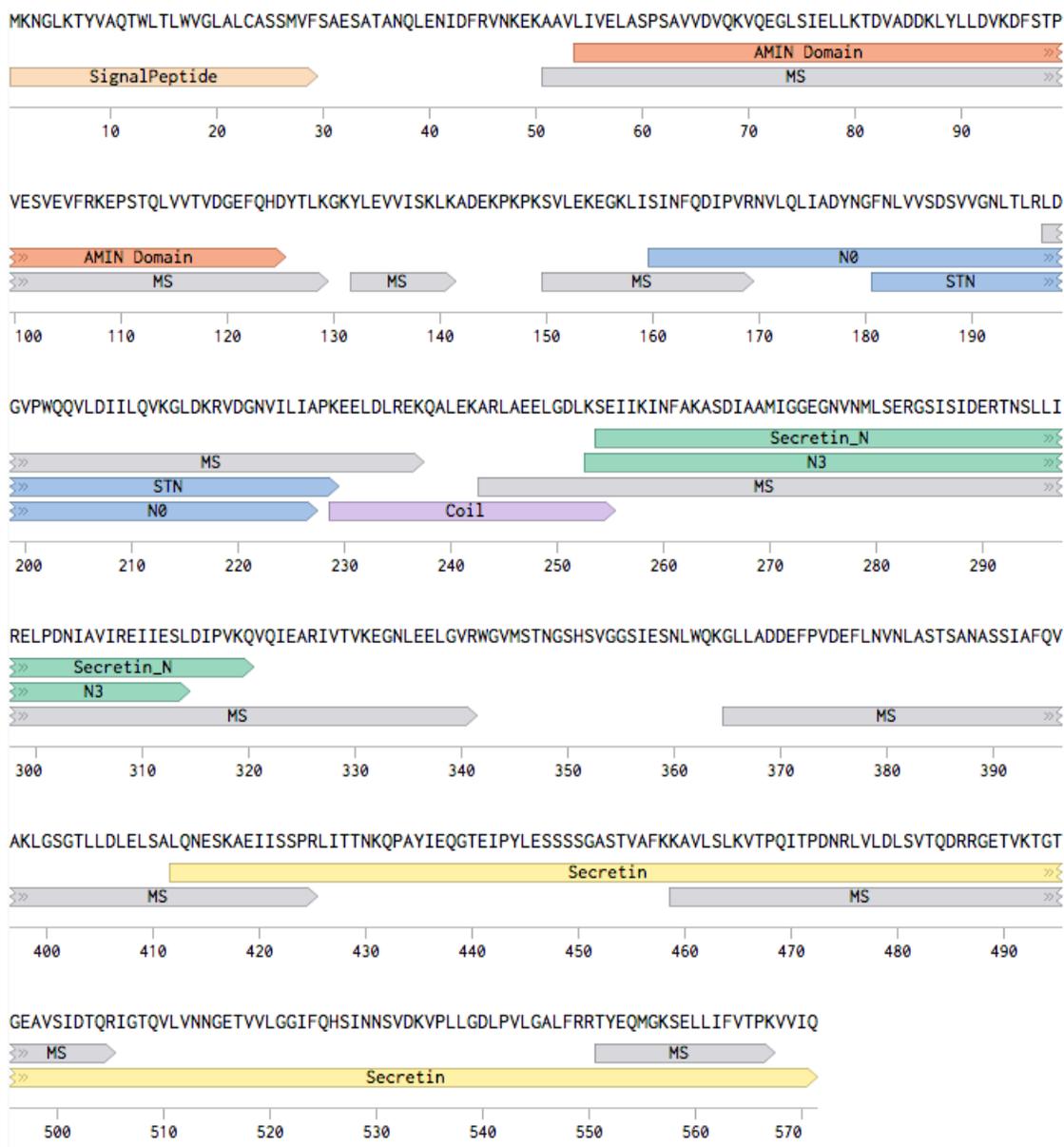


Figure 34 *V. cholerae* PilQ sequence with mass spectroscopy results.

Mass spectrometry coverage of purified PilQ: The sequence of VcPilQ (571 amino acids) is shown in black with residue numbers marked ever 10 residues. Regions of the sequence corresponding to protein motifs (Signal Peptide, AMIN Domain, STN, Coil, Secretin_N and Secretin) or folded regions of the protein structure (N0, N3) are highlighted with flags below the relevant sequence. The sequence coverage achieved by mass spectrometry is highlighted with grey flags labeled "MS". Approximately 65% of the sequence was represented in the peptides analyzed by mass spectrometry.

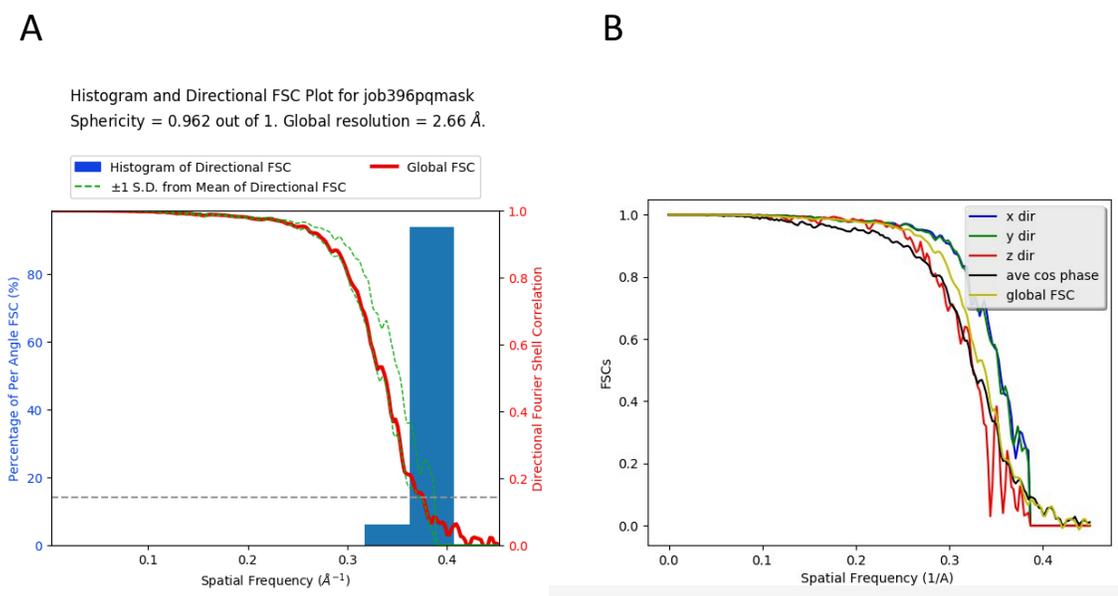


Figure 35 3D Fourier Shell Correlation Results.

The Salk institute 3DFSC server was used to examine the resolution in the X, Y and Z dimensions in the VcPilQ structure(Tan et al. 2017).

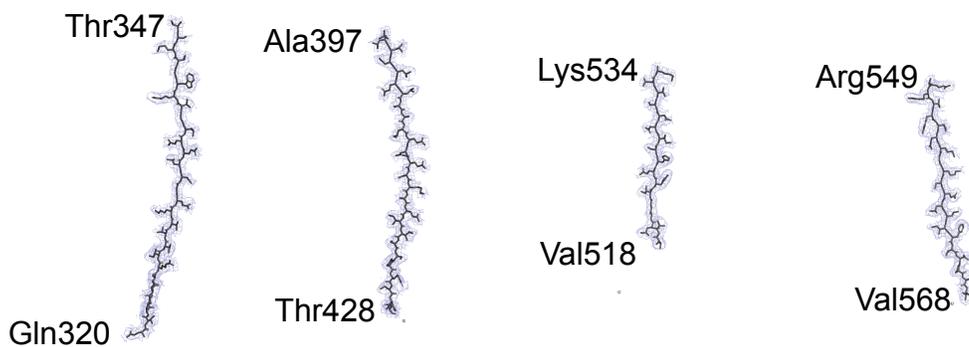


Figure 36 CryoEM density around VcPilQ beta strands.

CryoEM density around example beta strands was rendered in PyMol (carve=2.0 Å) to indicate map quality.

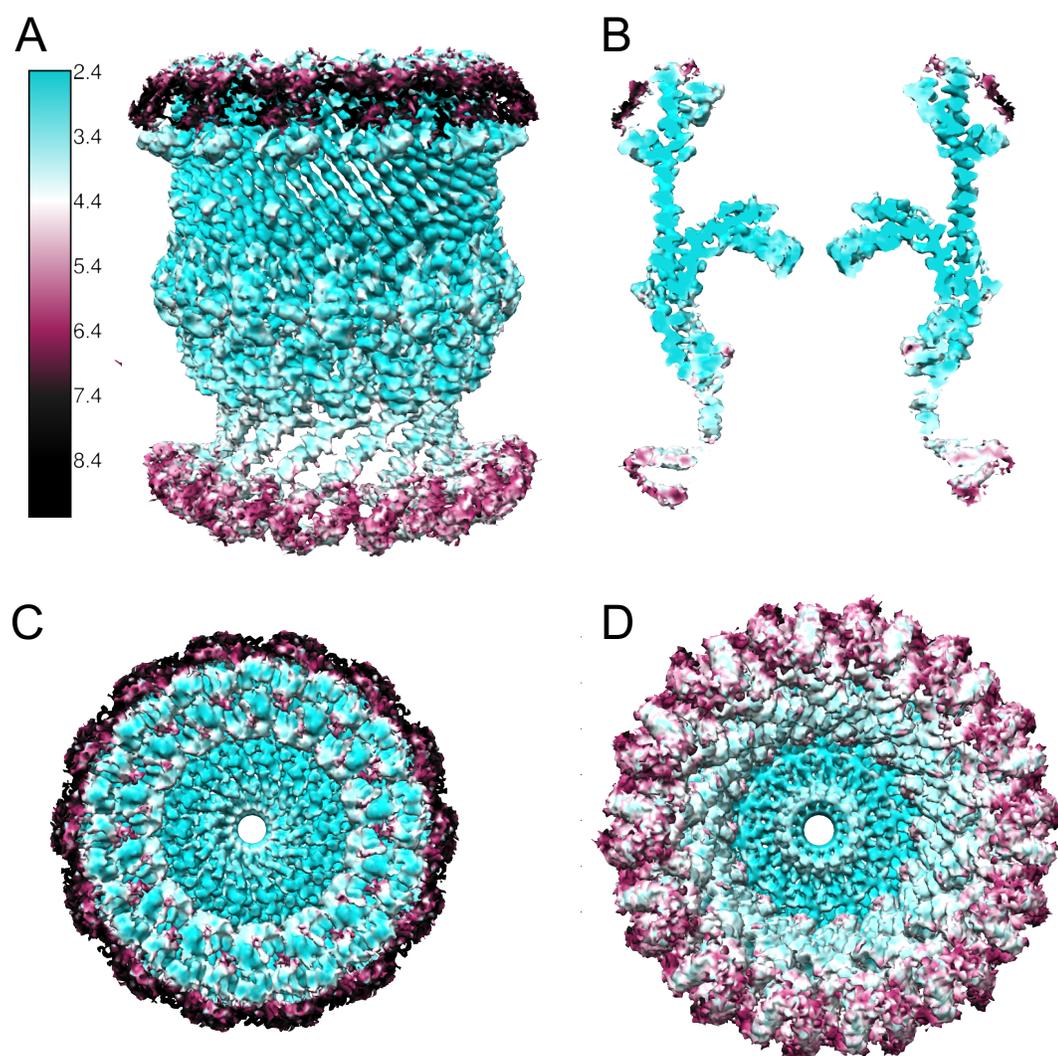


Figure 37 Local resolution of the VcPilQ structure.

ResMap was used to calculate the local resolution per voxel of the density map. Resolution is plotted from 2.4 Å (cyan) to 8.4 Å (black). VcPilQ is shown from the side (A), as a central slice (B), from the top (extracellular side, C) and from the bottom (periplasmic side, D).

The conservation scale:**Signal Sequence (residues 1-29)**

1 11 21
 MFNGLKTYVA QTWLTLLWVGL ALCASSMVF

Linker (Residues 30-53)

31 41 51
 S AESATANQLE NIDFRVNKEK AAV

AMIN Domain (Residues 54-125)

54 61 71 81 91
 LIVELAS PSAVVDVQKV QEGLSIELLK TDVADDKLYL LDVKDFSTPV

101 111 121
 ESVEVFRKEP STQLVVTVDG EFQHD

Linker (126-159)

131 141 151
 YTRG KYLEVVISKL KADEKPKPKS VLEKEGKLI

N0 Domain (160-227)

161 171 181 191
 S INFQDIPVRN VLQLIADYNG FNLVSDSVV GNLTLRLDGV

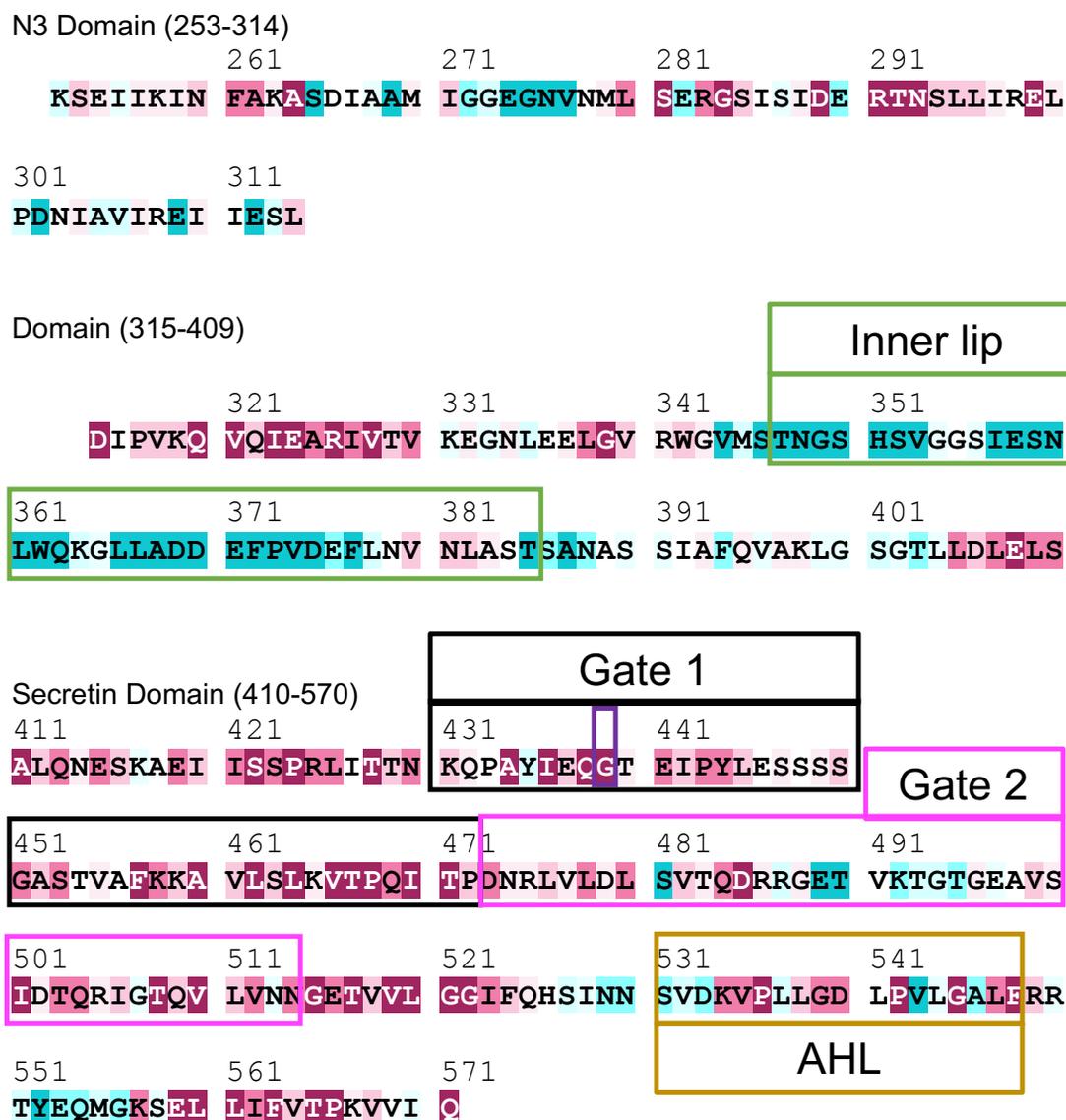
201 211 221
 PWQQVLDIIL QVKGLDKRVD GNVILIA

Coil (228-252)

231 241 251
 PKE ELDLREKQAL EKARLAEELG DL

Figure 38 ConSurf residue conservation analysis on VcPilQ residues 1 to 252.

Conservation of Type IV competence pilus machine residues in *V. cholerae* PilQ was calculated using ConSurf(Ashkenazy et al. 2010). The sequence of VcPilQ is shown with each residue colored by its degree of variability (blue) or conservation (magenta). Residues are numbered every 10 residues. Residues with insufficient signal for the ConSurf Analysis are shown in yellow (residues 1 to 20). These are in the signal sequence (residues 1-29) that gets cleaved during PilQ processing.



The conservation scale:

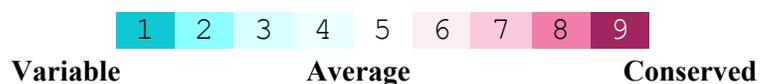


Figure 39 ConSurf residue conservation analysis on VcPilQ residues 253-571.

Conservation of Type IV competence pilus machine residues in *V. cholerae* PilQ was calculated using ConSurf(Ashkenazy et al. 2010). The sequence of VcPilQ is shown with each residue

colored by its degree of variability (blue) or conservation (magenta). Residues are numbered every 10 residues.

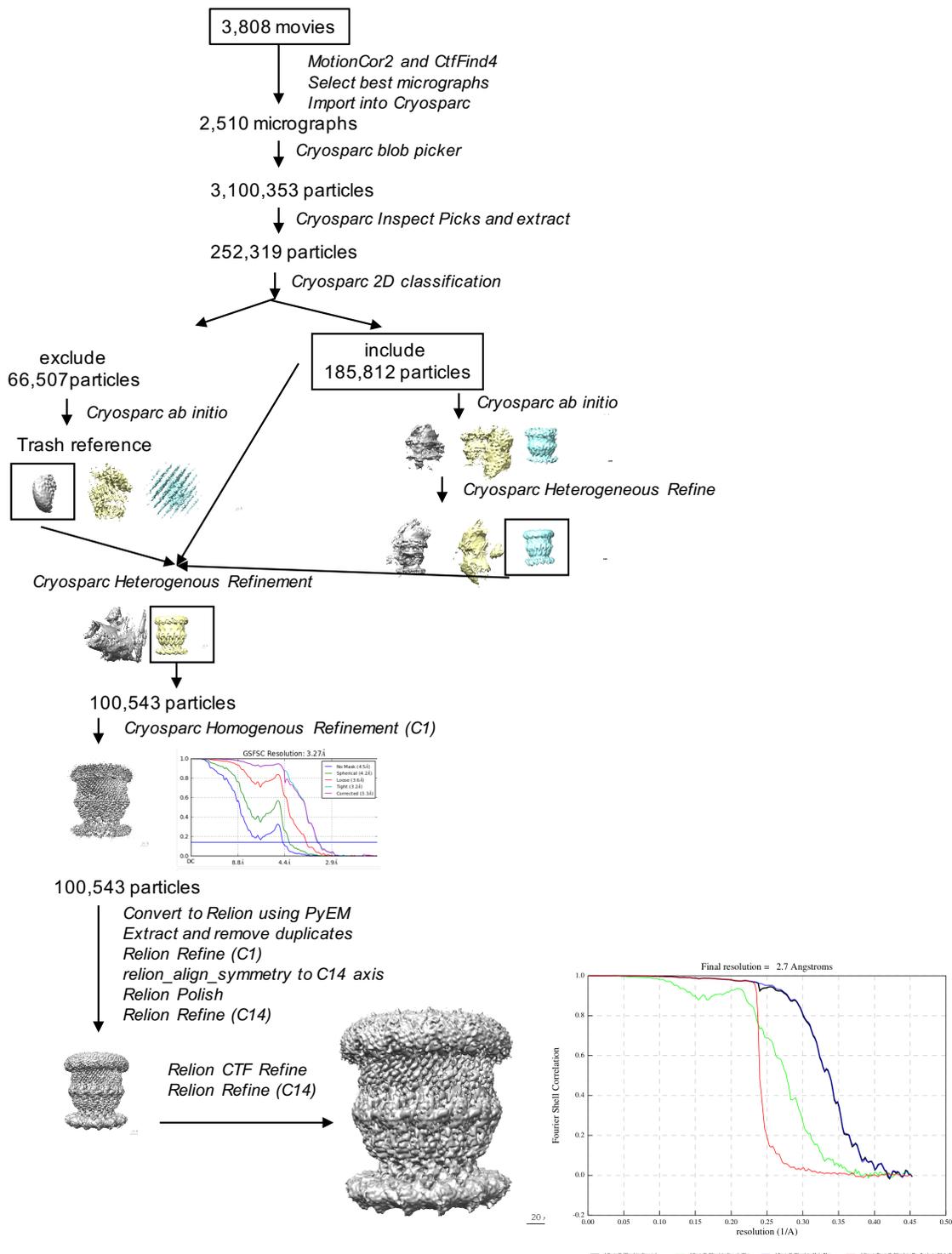


Figure 40 cryoEM image processing summary flow chart.

The major steps in data processing are summarized in the flow chart.

Table 1. Summary of single-particle data collection, 3D reconstruction, and model refinement

Imaging parameters and 3D reconstruction	
Acceleration voltage (kV)	300
Magnification (X)	81,000
Pixel size (Å)	1.104
Frame rate (s ⁻¹)	0.092
Exposure time (s)	3.7
Total exposure (e ⁻ / Å)	60
Particles	
Micrographs used for selection	2,510
Defocus range (µm)	-0.5 to -3.5
Windowed	252,319
In final 3D reconstruction	100,543
Resolution	
‘Gold-standard’ at FSC 0.5 (Å)	3.0 Å
‘Gold-standard’ at FSC 0.143 (Å)	2.7 Å
Map-sharpening B factor (Å ²)	-69
Model refinement	
Resolution in phenix.real_space_refine (Å)	3.0
Model-to-map fit (CC mask)	0.745
Number of atoms/residues/molecules	
NCS restrained chains	14
Protein atoms, residues (per chain)	43736, 412
Ramachandran angles (%)	
Favored	92.21
Allowed	7.79
Outliers	0
r.m.s. deviations	
Bond lengths (Å) Bond angles (°)	0.006
Bond angles (°)	0.847
Molprobit	
Score	2.81
Clashscore	10.53
Rotamer outliers (%)	10.91
EMRinger score	3.13

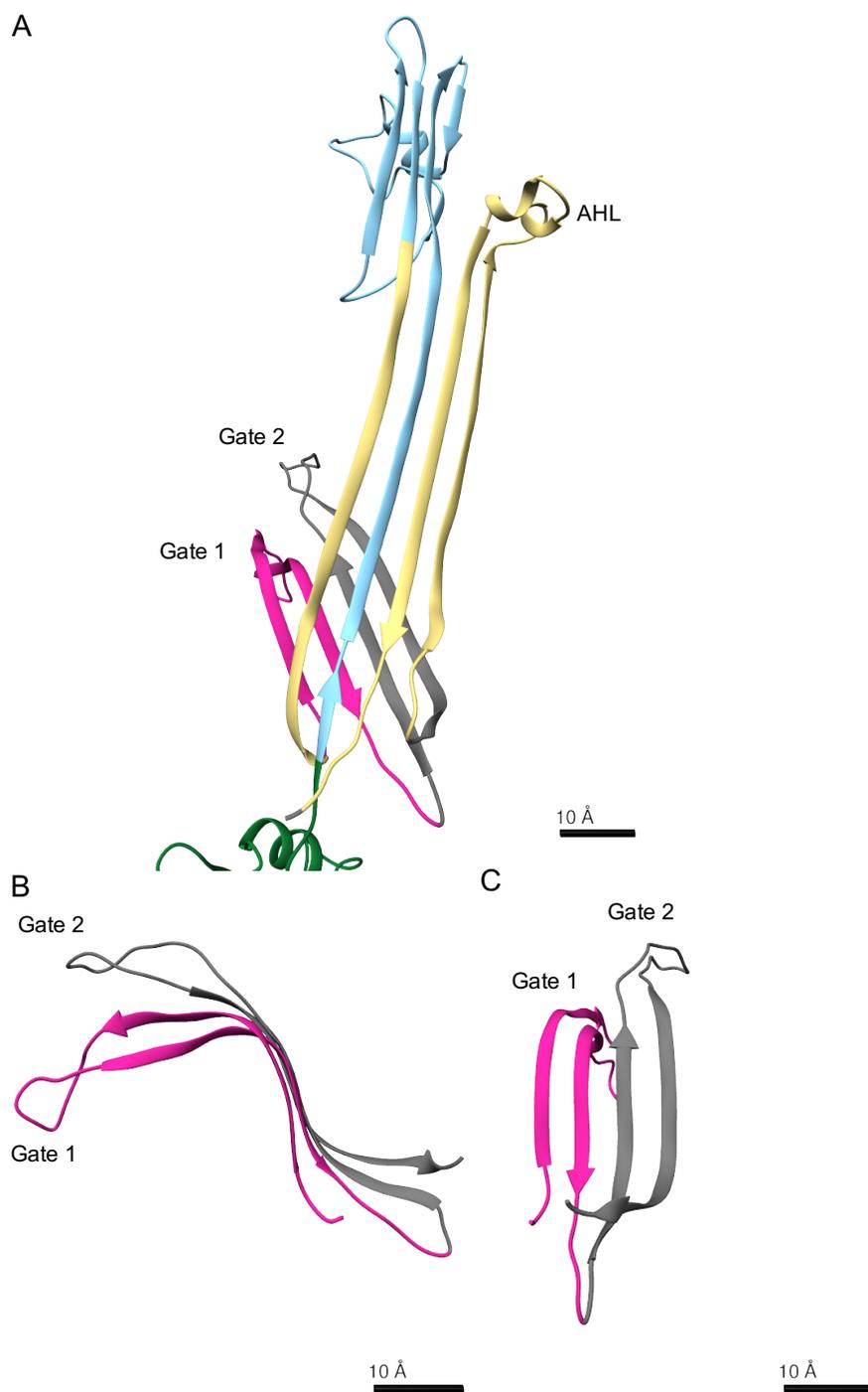


Figure 41 Ribbon diagram of VcPilQ beta sheets.

In (A) the coloring matches **Figure 22** except that gates 1 and 2 are colored separately here. In (B) and (C) two views of the isolated gate residues are shown. Scale, 10 Å.

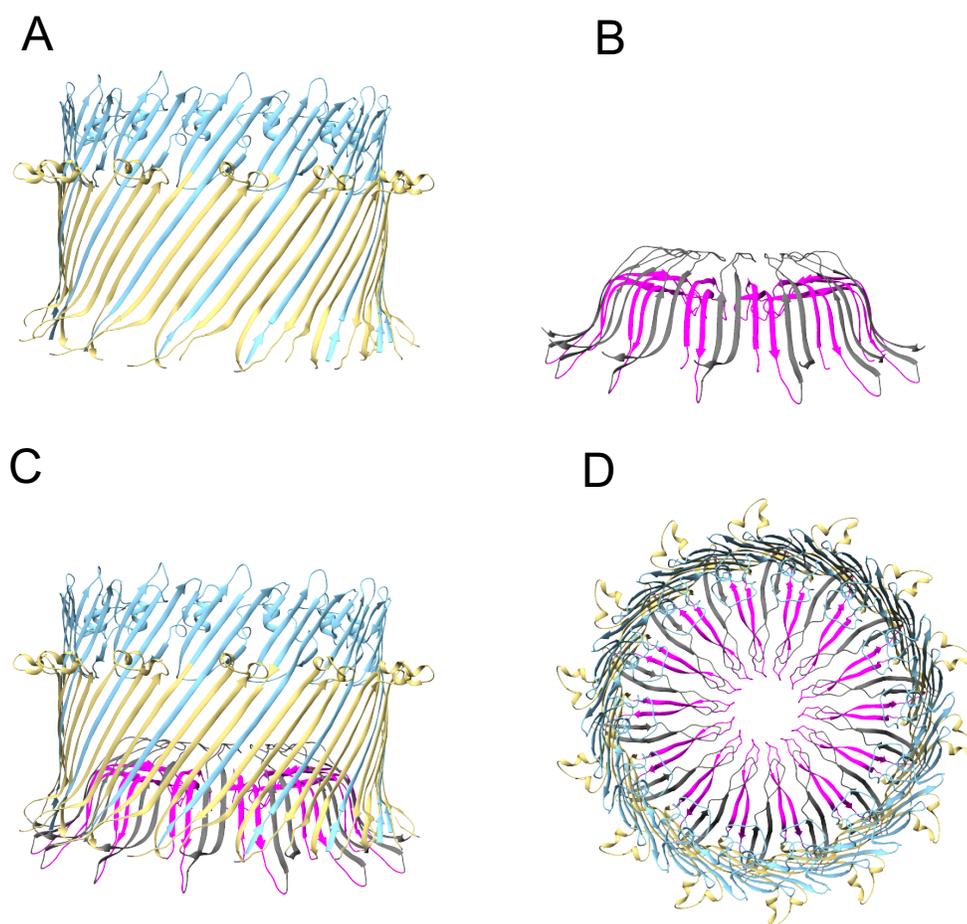


Figure 42 VcPilQ is a double beta barrel.
Ribbon diagrams focusing on the secretin domain beta barrel and the gate domain inner beta barrel are shown with coloring matching **Figure 41**.

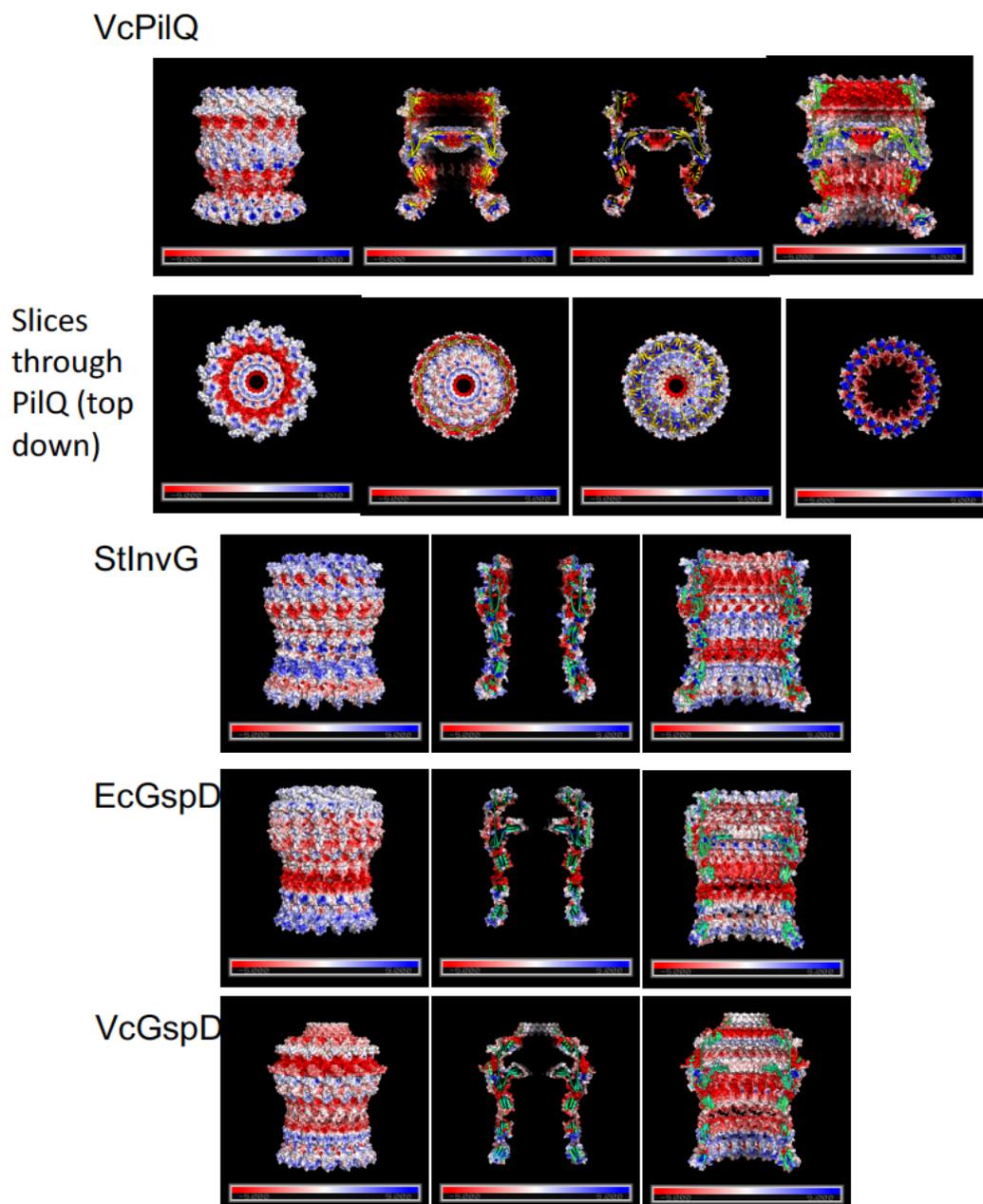


Figure 43 Electrostatic potential of secretins.

PyMol Adaptive Poisson-Boltzmann Solver was used to calculate the electrostatic potential calculation of VcPilQ (this study), *S. typhimurium* InvG (PDB), *E. coli* K12 GspD (5WQ7) and *V. cholerae* GspD (5WQ8) (Jurrus et al. 2018; Worrall et al. 2016; Yan et al. 2017).

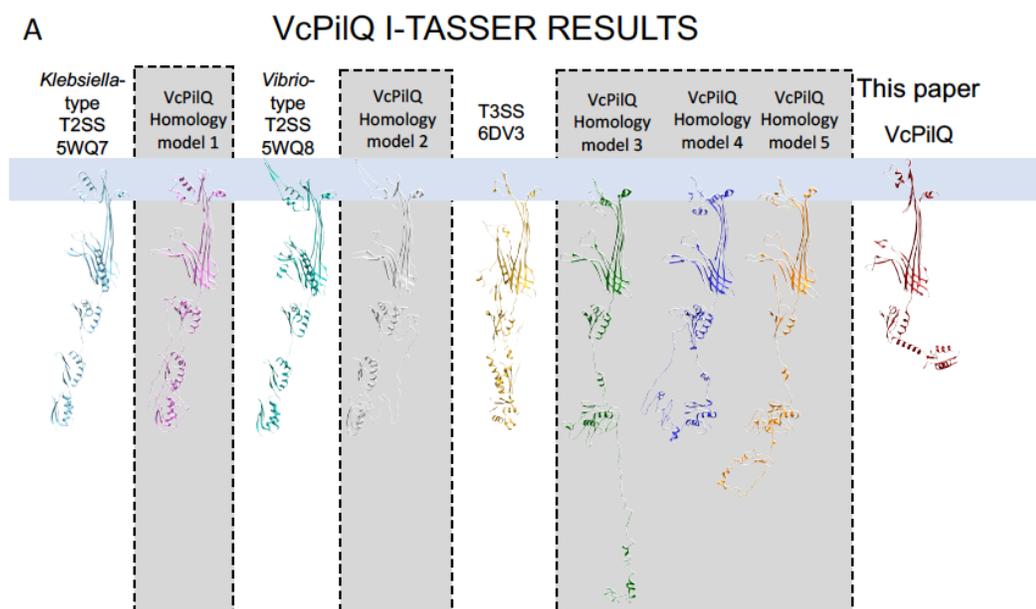


Figure 44 Homology modeling of *V. cholerae* PilQ.

I-TASSER was used to predict the structure of *V. cholerae* PilQ. I-TASSER reported five potential models (labeled VcPilQ Homology model 1 to 5 and shown with gray background). The I-TASSER models are based on structures deposited in the PDB, so the five models can be sorted by their similarity to the T2SS with the cap (*Vibrio*-type), without the cap (*Klebsiella*-type) or the T3SS, so the models are shown next to the corresponding secretin type. On the right the structure of VcPilQ solved in this paper is shown in dark red. The membrane is blue.

METHODS

Cloning of the His-tagged secretin constructs

The 10xHis-PilQ, Δ VC1807::SpecR, lacZ::lacIq, comEA-mCherry, Δ luxO, Ptac-tfoX, Δ TCP::ZeoR, Δ MSHA::CarbR, Δ CTX::KanR *V. cholerae* strain was prepared as described in (Ellison et al. 2018).

Transformation assay

The transformation assay prepared as described in (Ellison et al. 2018).

Expression

Vibrio cholerae expressing His-tagged PilQ (10xHis-PilQ, Δ VC1807::SpecR, lacZ::lacIq, comEA-mCherry, Δ luxO, Ptac-tfoX, Δ TCP::ZeoR, Δ MSHA::CarbR, Δ CTX::KanR) was streaked on Luria Broth agar plates and grown overnight at 30°C. Small cultures were seeded (5 mL) and grown over night at 30°C. The next day, 500 mL cultures were seeded with the 5 mL culture. LB broth was supplemented with 20 mM MgCl₂, 10 mM CaCl₂ and 100 μ M IPTG. The large cultures were grown overnight at 30°C in beveled flasks. The following day, cultures were spun down (6,000 rpm, 4°C, 20 minutes) and the cell paste was weighed, aliquoted, and stored at -80°C.

Purification

Cell pellet (15 g) was resuspended in lysis buffer (50 mM Tris HCl, pH 8, 300 mM NaCl, 1% DDM, 20 mM imidazole) supplemented with lysozyme (40 mg/mL in 50% glycerol/water), DNase I (4 mg/mL in 50% glycerol/water), and EDTA-free Protease Inhibitor tablet (Roche, 11697498001). Lysis proceeded with stirring at 4°C for 20 hours. Lysate was clarified by ultracentrifugation (Beckman L8-M ultracentrifuge, Rotor Type 45 Ti, 35,000rpm, 1 hour). The supernatant was mixed with Ni NTA agarose beads (Anatrace, SUPER-NINTA25) and incubated with stirring (4°C, 8 hours). In a gravity column at 4°C, proteins conjugated to Ni NTA agarose beads were washed (50 mM Tris HCl, pH 8, 300 mM NaCl, 0.05% DDM, 70 mM imidazole), (50 mM Tris HCl, pH 8, 300 mM NaCl, 0.05% DDM, 300 mM imidazole), and eluted (50 mM Tris HCl, pH 8, 300 mM

NaCl, 0.05% DDM, 1 M imidazole). Eluant was concentrated to ~1 mg/L (EMD Millipore Amicon Ultra-15, 30 kDa cutoff, UFC903024). Concentrated PilQ (150 μ L of ~1 mg/mL protein) was exchanged into Amphipol A8-35 (0.585 mg for a 3:1 ratio, Anatrace, A835) and allowed to incubate at 4 C for 1 hr. Excess DDM was removed using Bio-Beads SM2 (Bio-Rad, 1523920) by incubating overnight at 4°C. The protein was concentrated.

Electron microscopy

For cryoEM, Quantifoil R2/2 300 Mesh grids (EMS, Q33100CR2) were glow discharged (Pelco EasiGlow, 20 mA, 60 seconds). PilQ in amphipol (3 μ L of ~0.8 mg/L) was frozen on a Mark IV Vitrobot (FEI, 20 °C, 100% relative humidity, blot force -6, blot time 4 s). Micrographs were collected on a 300 kV Titan Krios microscope (FEI) with energy filter (Gatan) and equipped with a K3 direct electron detector (Gatan). Data was collected using Serial EM software with a pixel size of 1.104 Å (81,000x magnification) and a defocus range from -1.0 μ m to -3.0 μ m (Mastronarde 2005). A fluence of 19.8 electrons/pixel/second was used with a 3.7 s exposure time to collect 60 e-/Å².

Image Processing

The cryoEM image processing workflow is summarized in **Figure 40**. MotionCor2 was used for motion correction and dose weighting of 3,808 movies (Zheng et al. 2017). CTF correction was used to evaluate micrograph quality (Rohou and Grigorieff 2015). Cryosparc blob picking on 2,510 micrographs yielded 3,100,353 potential particles (Punjani et al. 2017). After inspection, the 252,319 particles were analyzed by several rounds of 2D classification and 3D classification to yield 100,543 particles. These particles were moved to Relion using the UCSF PyEM package (<https://github.com/asarnow/pyem>) script. In Relion, several rounds of 3D refinement, polishing and CTF refinement were used (Kimanius et al. 2016; Zivanov et al. 2018; Zivanov, Nakane, and Scheres 2019). ResMap was used to calculate local resolution (Kucukelbir, Sigworth, and Tagare 2014b).

Model Building and Refinement

The initial model (residues 230-571) was auto-built using Buccaneer(Cowtan 2006). Subsequent building and model adjustments were performed by hand using COOT(Emsley et al. 2010). A homology model of the N0 domain (residues 160-229) was created using I-TASSER and manually docked using COOT(Roy, Kucukural, and Zhang 2010; Yang et al. 2015; Y. Zhang 2008). Coulombic potential density for residues 1-159 were not observed. The model was refined in PHENIX version 1.16-dev3549 using phenix.real_space_refine with the resolution set to 3 Å (Adams et al. 2010). NCS constraints were applied for the 14 subunits and were automatically detected and refined. Automatically determined secondary structure restraints, rotamer restraints, and Ramachandran restraints were applied as well. The quality of the model was evaluated using EMRinger(Barad et al. 2015) and Molprobit (Chen et al. 2010) (Table 1).

Mass Spectrometry

After running a BioRad Stain Free gel and performing a coomassie staining, the band of interest was excised with a clean razor blade. The gel piece was destained with ammonium bicarbonate and reduced with DTT (50°C, 30 minutes). Next the sample was alkylated with iodoacetamide (room temperature, dark, 20 minutes). The gel pieces were then dehydrated. Trypsin was used to digest the protein in the gel (37°C, overnight). Peptides were extracted from the gel matrix, dried, and de salted with a zip tip.

The in-gel-digested samples were subjected to LC-MS/MS analysis on a nanoflow LC system, EASY-nLC 1200, (Thermo Fisher Scientific) coupled to a QExactive HF Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a Nanospray Flex ion source.

Samples were directly loaded onto a C18 Aurora series column (Ion Opticks, Parkville, Australia). The 25cm x 50µm ID column (1.6 µm) was heated to 45° C. The peptides were separated with a 60 min gradient at a flow rate of 350 nL/min. The gradient was as follows: 2–6% Solvent B (3.5 min), 6-25% B (42.5 min), and 25-40% B (14.5min), to 100% B

(1min) and 100% B (12min). Solvent A consisted of 97.8% H₂O, 2% ACN, and 0.2% formic acid and solvent B consisted of 19.8% H₂O, 80% ACN, and 0.2% formic acid.

The QExactive HF Orbitrap was operated in data dependent mode. Spray voltage was set to 1.8 kV, S-lens RF level at 50, and heated capillary at 275 °C. Full scan resolution was set to 60,000 at m/z 200. Full scan target was 3×10^6 with a maximum injection time of 15 ms (profile mode). Mass range was set to 300–1650 m/z. For data dependent MS2 scans the loop count was 12, target value was set at 1×10^5 , and intensity threshold was kept at 1×10^5 . Isolation width was set at 1.2 m/z and a fixed first mass of 100 was used. Normalized collision energy was set at 28. Peptide match was set to off, and isotope exclusion was on. Ms2 data was collected in centroid mode.

Raw data were analyzed using MaxQuant (v. 1.6.5.0)(Cox et al. 2011; Cox and Mann 2008). Spectra were searched against UniProt *V. cholerae* entries (3784 sequences) and a contaminant protein database (246 sequences). Trypsin was specified as the digestion enzyme and up to two missed cleavages were allowed. Precursor mass tolerance was 4.5 ppm after recalibration and fragment mass tolerance was 20 ppm. Variable modifications included oxidation of methionine and protein N-terminal acetylation. Carbamidomethylation of cysteine was specified as a fixed modification. A decoy database was used to set score thresholds to ensure a 1% false discovery rate at the protein and peptide level. Protein abundances were estimated using iBAQ and the fractional abundance was calculated as the protein abundance divided by the sum of all non-contaminant protein abundances(Schwanhäusser et al. 2011).

Protein sequence alignment

The ConSurf web server was used to identify proteins related to PilQ and perform multiple sequence analysis(Ashkenazy et al. 2016; Celniker et al. 2013; Ashkenazy et al. 2010). Because our atomic model of PilQ starts at residue 160, an initial ConSurf run was performed using the full PilQ sequence to obtain a starting MSA(Berezin et al. 2004).

Next, a ConSurf run using that MSA and our atomic model of PilQ was performed to map conserved residues onto the PilQ model(Landau et al. 2005; Glaser et al. 2003).

Acknowledgements

Cryo Electron microscopy was performed in the Beckman Institute Resource Center for Transmission Electron Microscopy at Caltech. Dr. Songye Chen, Dr. Andrey Malyutin assisted with data collection. Dr. Spiros D. Garbis, Dr. Annie Moradian, Dr. Michael Sweredoski, Dr. Brett Lomenick at the Caltech Proteome Exploration Laboratory (PEL) performed and analyzed mass spectrometry results. Dr. Naima Sharaf, Jeffery Lai, and Prof. Doug Rees provided invaluable advice on membrane protein biochemistry and instrumentation. Jane Ding and Welison Floriano provided computational support. Dr. Debnath Ghosal, Dr. Mohammed Kaplan, Dr. Davi Ortega, Dr. Catherine Oikonomou, Dr. Lauren Ann Metskas, Dr. Christopher Barnes, Claudia Jette, and Andrew Schacht provided feedback and advice.

BIBLIOGRAPHY

- Adams, P. D., P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, et al. 2010. "PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution." *Acta Crystallographica Section D: Biological Crystallography* 66 (2): 213–21. <https://doi.org/10.1107/S0907444909052925>.
- Adebali, Ogun, Davi R. Ortega, and Igor B. Zhulin. 2015. "CDvist: A Webserver for Identification and Visualization of Conserved Domains in Protein Sequences." *Bioinformatics* 31 (9): 1475–77. <https://doi.org/10.1093/bioinformatics/btu836>.
- 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>.
- Ashkenazy, Haim, Elana Erez, Eric Martz, Tal Pupko, and Nir Ben-Tal. 2010. "ConSurf 2010: Calculating Evolutionary Conservation in Sequence and Structure of Proteins and Nucleic Acids." *Nucleic Acids Research* 38 (suppl_2): W529–33. <https://doi.org/10.1093/nar/gkq399>.
- Bai, Xiao-chen, Eeson Rajendra, Guanghui Yang, Yigong Shi, and Sjors HW Scheres. 2015. "Sampling the Conformational Space of the Catalytic Subunit of Human γ -Secretase." *ELife* 4: e11182. <https://doi.org/10.7554/elife.11182>.
- Balasingham, Seetha V, Richard F Collins, Reza Assalkhou, Håvard Homberset, Stephan A Frye, Jeremy P Derrick, and Tone Tønjum. 2007. "Interactions between the Lipoprotein PilP and the Secretin PilQ in Neisseria Meningitidis." *Journal of Bacteriology* 189: 5716–5727. <https://doi.org/10.1128/jb.00060-07>.
- Barad, Benjamin A, Nathaniel Echols, Ray Wang, Yifan Cheng, DiMaio Frank, Paul D Adams, and James S Fraser. 2015. "EMRinger: Side Chain-Directed Model and Map Validation for 3D Cryo-Electron Microscopy." <https://doi.org/10.1038/nmeth.3541>.
- Bardiaux, Benjamin, Gisele C de Amorim, Areli Rico, Weili Zheng, Ingrid Guilvout, Camille Jollivet, Michael Nilges, Edward H Egelman, Nadia Izadi-Pruneyre, and Olivera Francetic. 2019. "Structure and Assembly of the Enterohemorrhagic Escherichia Coli Type 4 Pilus." *Structure (London, England : 1993)*. <https://doi.org/10.1016/j.str.2019.03.021>.
- Bartesaghi, Alberto, Alan Merk, Soojay Banerjee, Doreen Matthies, Xiongwu Wu, Jacqueline LS Milne, and Sriram Subramaniam. 2015. "2.2 Å Resolution Cryo-EM Structure of β -Galactosidase in Complex with a Cell-Permeant Inhibitor" 348 (6239): 1147–51. <https://doi.org/10.1126/science.aab1576>.

- Berezin, Carine, Fabian Glaser, Josef Rosenberg, Inbal Paz, Tal Pupko, Piero Fariselli, Rita Casadio, and Nir Ben-Tal. 2004. "ConSeq: The Identification of Functionally and Structurally Important Residues in Protein Sequences." *Bioinformatics* 20 (8): 1322–24. <https://doi.org/10.1093/bioinformatics/bth070>.
- Berman, Helen M., John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. 2000. "The Protein Data Bank." *Nucleic Acids Research* 28 (1): 235–42. <https://doi.org/10.1093/nar/28.1.235>.
- Berry, Jamie-Lee, Marie M Phelan, Richard F Collins, Tomas Adomavicius, Tone Tønjum, Stefan A Frye, Louise Bird, et al. 2012. "Structure and Assembly of a Trans-Periplasmic Channel for Type IV Pili in *Neisseria Meningitidis*." *Plos Pathog* 8 (9): e1002923. <https://doi.org/10.1371/journal.ppat.1002923>.
- Blow, David. 2002. *Outline of Crystallography for Biologists*. OUP Oxford.
- Brauchle, Michael, Simon Hansen, Emmanuel Caussin, Anna Lenard, Amanda Ochoa-Espinosa, Oliver Scholz, Simon G Sprecher, Andreas Plückthun, and Markus Affolter. 2014. "Protein Interference Applications in Cellular and Developmental Biology Using DARPins That Recognize GFP and MCherry." *Biology Open* 3: 1252–1261. <https://doi.org/10.1242/bio.201410041>.
- Brocchieri, Luciano, and Samuel Karlin. 2005. "Protein Length in Eukaryotic and Prokaryotic Proteomes" 33 (10): 3390–3400. <https://doi.org/10.1093/nar/gki615>.
- Campbell, Melody G, Anchi Cheng, Axel F Brilot, Arne Moeller, Dmitry Lyumkis, David Veesler, Junhua Pan, et al. 2012. "Movies of Ice-Embedded Particles Enhance Resolution in Electron Cryo-Microscopy" 20 (11): 1823–28. <https://doi.org/10.1016/j.str.2012.08.026>.
- Celniker, Gershon, Guy Nimrod, Haim Ashkenazy, Fabian Glaser, Eric Martz, Itay Mayrose, Tal Pupko, and Nir Ben-Tal. 2013. "ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function." *Israel Journal of Chemistry* 53 (3–4): 199–206. <https://doi.org/10.1002/ijch.201200096>.
- Chami, Mohamed, Ingrid Guilvout, Marco Gregorini, Hervé W Rémigy, Shirley A Müller, Marielle Valerio, Andreas Engel, Anthony P Pugsley, and Nicolas Bayan. 2005. "Structural Insights into the Secretin PulD and Its Trypsin-Resistant Core." *Journal of Biological Chemistry* 280: 37732–37741. <https://doi.org/10.1074/jbc.m504463200>.
- Chang, Yi-Wei, Lee A Rettberg, Anke Treuner-Lange, Janet Iwasa, Lotte Sogaard-Andersen, and Grant J Jensen. 2016. "Architecture of the Type IVa Pilus Machine." *Science* 351: aad2001. <https://doi.org/10.1126/science.aad2001>.
- Chen, V. B., W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson. 2010. "MolProbity: All-Atom Structure Validation for Macromolecular Crystallography." *Acta Crystallographica Section D: Biological*

- Crystallography* 66 (1): 12–21.
<https://doi.org/10.1107/S0907444909042073>.
- Cheng, Anchi, Edward T. Eng, Lambertus Alink, William J. Rice, Kelsey D. Jordan, Laura Y. Kim, Clinton S. Potter, and Bridget Carragher. 2018. “High Resolution Single Particle Cryo-Electron Microscopy Using Beam-Image Shift.” *Journal of Structural Biology* 204 (2): 270–75. <https://doi.org/10.1016/j.jsb.2018.07.015>.
- Cheng, Yifan, Nikolaus Grigorieff, Pawel A Penczek, and Thomas Walz. 2015. “A Primer to Single-Particle Cryo-Electron Microscopy.” *Cell* 161 (3): 438–49. <https://doi.org/10.1016/j.cell.2015.03.050>.
- Chernyatina, Anastasia A, and Harry H Low. 2018. “Architecture of a Bacterial Type II Secretion System.” *BioRxiv*, 397794. <https://doi.org/10.1101/397794>.
- Chernyatina, Anastasia A., and Harry H. Low. 2019. “Core Architecture of a Bacterial Type II Secretion System.” *BioRxiv*, September, 397794. <https://doi.org/10.1101/397794>.
- Coscia, Francesca, Leandro F Estrozi, Fabienne Hans, H el ene Malet, Noirclerc-Savoie Marjolaine, Guy Schoehn, and Carlo Petosa. 2016. “Fusion to a Homo-Oligomeric Scaffold Allows Cryo-EM Analysis of a Small Protein.” *Sci Rep-Uk* 6 (1): srep30909. <https://doi.org/10.1038/srep30909>.
- Cowtan, K. 2006. “The Buccaneer Software for Automated Model Building. 1. Tracing Protein Chains.” *Acta Crystallographica Section D: Biological Crystallography* 62: 1002–11. <https://doi.org/10.1107/s0907444906022116>.
- Cox, J urgen, and Matthias Mann. 2008. “MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification.” *Nature Biotechnology* 26 (12): 1367–72. <https://doi.org/10.1038/nbt.1511>.
- Cox, J urgen, Nadin Neuhauser, Annette Michalski, Richard A. Scheltema, Jesper V. Olsen, and Matthias Mann. 2011. “Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment.” *Journal of Proteome Research* 10 (4): 1794–1805. <https://doi.org/10.1021/pr101065j>.
- Craig, Lisa, Michael E Pique, and John A Tainer. 2004. “Type IV Pilus Structure and Bacterial Pathogenicity.” *Nature Reviews Microbiology* 2: 363. <https://doi.org/10.1038/nrmicro885>.
- Danev, Radostin, Haruaki Yanagisawa, and Masahide Kikkawa. 2019. “Cryo-Electron Microscopy Methodology: Current Aspects and Future Directions.” *Trends in Biochemical Sciences* 44 (10): 837–48. <https://doi.org/10.1016/j.tibs.2019.04.008>.
- D’Imprima, Edoardo, Davide Floris, Mirko Joppe, Ricardo S anchez, Martin Grininger, and Werner K uhlbrandt. 2019. “Protein Denaturation at the Air-Water Interface and How to Prevent It.” *ELife* 8. <https://doi.org/10.7554/elife.42747>.
- D’Imprima, Edoardo, Ralf Salzer, Ramachandra M Bhaskara, Ricardo S anchez, Ilona Rose, Lennart Kirchner, Gerhard Hummer, Werner K uhlbrandt, Janet Vonck, and Beate Averhoff. 2017. “Cryo-EM Structure of the Bifunctional

- Secretin Complex of *Thermus Thermophilus*.” *ELife* 6: e30483. <https://doi.org/10.7554/elife.30483>.
- Downing, Kenneth H, and Robert M Glaeser. 2018. “Estimating the Effect of Finite Depth of Field in Single-Particle Cryo-EM.” *Ultramicroscopy* 184: 94–99. <https://doi.org/10.1016/j.ultramic.2017.08.007>.
- Ellison, Courtney K, Triana N Dalia, Alfredo Ceballos, Joseph Wang, Nicolas Biais, Yves V Brun, and Ankur B Dalia. 2018. “Retraction of DNA-Bound Type IV Competence Pili Initiates DNA Uptake during Natural Transformation in *Vibrio Cholerae*.” *Nature Microbiology* 3: 773–780. <https://doi.org/10.1038/s41564-018-0174-y>.
- Emsley, P, and K Cowtan. 2004. “Coot: Model-Building Tools for Molecular Graphics.” *Acta Crystallographica Section D: Biological Crystallography* 60: 2126–32. <https://doi.org/10.1107/S0907444904019158>.
- Emsley, P., B. Lohkamp, W. G. Scott, and K. Cowtan. 2010. “Features and Development of Coot.” *Acta Crystallographica Section D: Biological Crystallography* 66 (4): 486–501. <https://doi.org/10.1107/S0907444910007493>.
- Fan, Xiao, Jia Wang, Xing Zhang, Zi Yang, Jin-Can Zhang, Lingyun Zhao, Hai-Lin Peng, Jianlin Lei, and Hong-Wei Wang. 2019. “Single Particle Cryo-EM Reconstruction of 52 KDa Streptavidin at 3.2 Angstrom Resolution.” *Nature Communications* 10: 2386. <https://doi.org/10.1038/s41467-019-10368-w>.
- Genin, S., and C.A. Boucher. 1994. “A Superfamily of Proteins Involved in Different Secretion Pathways in Gram-Negative Bacteria: Modular Structure and Specificity of the N-Terminal Domain.” *Molecular and General Genetics MGG* 243 (1): 112–18. <https://doi.org/10.1007/BF00283883>.
- Glaser, Fabian, Tal Pupko, Inbal Paz, Rachel E. Bell, Dalit Bechor-Shental, Eric Martz, and Nir Ben-Tal. 2003. “ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information.” *Bioinformatics* 19 (1): 163–64. <https://doi.org/10.1093/bioinformatics/19.1.163>.
- Graille, Marc, Enrico A Stura, Nicholas G Housden, Jennifer A Beckingham, Stephen P Bottomley, Dennis Beale, Michael J Taussig, Brian J Sutton, Michael G Gore, and Jean-Baptiste Charbonnier. 2001. “Complex between *Peptostreptococcus Magnus* Protein L and a Human Antibody Reveals Structural Convergence in the Interaction Modes of Fab Binding Proteins.” *Structure* 9 (8): 679–87. [https://doi.org/10.1016/s0969-2126\(01\)00630-x](https://doi.org/10.1016/s0969-2126(01)00630-x).
- Grant, Timothy, and Nikolaus Grigorieff. 2015. “Measuring the Optimal Exposure for Single Particle Cryo-EM Using a 2.6 Å Reconstruction of Rotavirus VP6.” *ELife* 4: e06980. <https://doi.org/10.7554/eLife.06980>.
- Guilvout, Ingrid, Nicholas N. Nickerson, Mohamed Chami, and Anthony P. Pugsley. 2011. “Multimerization-Defective Variants of Dodecameric Secretin PulD.” *Research in Microbiology* 162 (2): 180–90. <https://doi.org/10.1016/j.resmic.2011.01.006>.

- Hansen, Simon, Jakob C. Stüber, Patrick Ernst, Alexander Koch, Daniel Bojar, Alexander Batyuk, and Andreas Plückthun. 2017. "Design and Applications of a Clamp for Green Fluorescent Protein with Picomolar Affinity." *Scientific Reports* 7: 16292. <https://doi.org/10.1038/s41598-017-15711-z>.
- Hanshaw, Roger G., Robert V. Stahelin, and Bradley D. Smith. 2008. "Noncovalent Keystone Interactions Controlling Biomembrane Structure." *Chemistry (Weinheim an Der Bergstrasse, Germany)* 14 (6): 1690–97. <https://doi.org/10.1002/chem.200701589>.
- Hay, Iain D, Matthew J Belousoff, Rhys A Dunstan, Rebecca S Bamert, and Trevor Lithgow. 2018. "Structure and Membrane Topography of the Vibrio-Type Secretin Complex from the Type 2 Secretion System of Enteropathogenic Escherichia Coli." *Journal of Bacteriology* 200. <https://doi.org/10.1128/JB.00521-17>.
- Hay, Iain D., Matthew J. Belousoff, and Trevor Lithgow. 2017. "Structural Basis of Type 2 Secretion System Engagement between the Inner and Outer Bacterial Membranes." *MBio* 8: e01344–17. <https://doi.org/10.1128/mbio.01344-17>.
- Henderson, R. 1995. "The Potential and Limitations of Neutrons, Electrons and X-Rays for Atomic Resolution Microscopy of Unstained Biological Molecules." 28 (2): 171–93. <https://doi.org/10.1017/s003358350000305x>.
- Henderson, Richard. 1992. "Image Contrast in High-Resolution Electron Microscopy of Biological Macromolecules: TMV in Ice" 46 (1–4): 118. [https://doi.org/10.1016/0304-3991\(92\)90003-3](https://doi.org/10.1016/0304-3991(92)90003-3).
- Herzik, Mark A., Mengyu Wu, and Gabriel C. Lander. 2017. "Achieving Better-than-3-Å Resolution by Single-Particle Cryo-EM at 200 KeV." *Nature Methods* 14 (11): 1075–78. <https://doi.org/10.1038/nmeth.4461>.
- Herzik, Mark A, Mengyu Wu, and Gabriel C Lander. 2019. "High-Resolution Structure Determination of Sub-100 KDa Complexes Using Conventional Cryo-EM." *Nature Communications* 10: 1032. <https://doi.org/10.1038/s41467-019-08991-8>.
- Howard, Peter S, Leandro F Estrozi, Quentin Bertrand, Carlos Contreras-Martel, Timothy Strozen, Viviana Job, Alexandre Martins, Daphna Fenel, Guy Schoehn, and Andréa Dessen. 2019. "Structure and Assembly of Pilotin-Dependent and -Independent Secretins of the Type II Secretion System." *PLOS Pathogens* 15: e1007731. <https://doi.org/10.1371/journal.ppat.1007731>.
- Hu, Dalong, Bin Liu, Lu Feng, Peng Ding, Xi Guo, Min Wang, Boyang Cao, Peter R. Reeves, and Lei Wang. 2016. "Origins of the Current Seventh Cholera Pandemic." *Proceedings of the National Academy of Sciences* 113 (48): E7730–39. <https://doi.org/10.1073/pnas.1608732113>.
- Hu, J., L. J. Worrall, C. Hong, M. Vuckovic, C. E. Atkinson, N. Caveney, Z. Yu, and N. C. J. Strynadka. 2018. "Cryo-EM Analysis of the T3S Injectisome Reveals the Structure of the Needle and Open Secretin." *Nature Communications* 9: 3840. <https://doi.org/10.1038/s41467-018-06298-8>.

- Interlandi, Gianluca, Svava K Wetzl, Giovanni Settanni, Andreas Plückthun, and Amedeo Caflisch. 2008. "Characterization and Further Stabilization of Designed Ankyrin Repeat Proteins by Combining Molecular Dynamics Simulations and Experiments." *Journal of Molecular Biology* 375: 837–854. <https://doi.org/10.1016/j.jmb.2007.09.042>.
- Jensen, Grant J, and Roger D Kornberg. 1998. "Single-Particle Selection and Alignment with Heavy Atom Cluster-Antibody Conjugates." *Proceedings of the National Academy of Sciences* 95: 9262–9267. <https://doi.org/10.1073/pnas.95.16.9262>.
- Jeong, Woo H, Haerim Lee, Dong H Song, Jae-Hoon H Eom, Sun C Kim, Hee-Seung S Lee, Hayyoung Lee, and Jie-Oh O Lee. 2016. "Connecting Two Proteins Using a Fusion Alpha Helix Stabilized by a Chemical Cross Linker." *Nat Commun* 7 (1): 11031. <https://doi.org/10.1038/ncomms11031>.
- Jurrus, Elizabeth, Dave Engel, Keith Star, Kyle Monson, Juan Brandi, Lisa E. Felberg, David H. Brookes, et al. 2018. "Improvements to the APBS Biomolecular Solvation Software Suite." *Protein Science* 27 (1): 112–28. <https://doi.org/10.1002/pro.3280>.
- Kelley, Lawrence A., Stefans Mezulis, Christopher M. Yates, Mark N. Wass, and Michael J. E. Sternberg. 2015. "The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis." *Nature Protocols* 10 (6): 845–58. <https://doi.org/10.1038/nprot.2015.053>.
- Khoshouei, Maryam, Mazdak Radjainia, Wolfgang Baumeister, and Radostin Danev. 2017. "Cryo-EM Structure of Haemoglobin at 3.2 Å Determined with the Volta Phase Plate." *Nature Communications* 8: ncomms16099. <https://doi.org/10.1038/ncomms16099>.
- Kimanius, Dari, Bjorn O Forsberg, Sjors Scheres, and Erik Lindahl. 2016. "Accelerated Cryo-EM Structure Determination with Parallelisation Using GPUs in RELION-2," 059717. <https://doi.org/10.1101/059717>.
- Kolappan, Subramania, Mathieu Coureuil, Xiong Yu, Xavier Nassif, Edward H. Egelman, and Lisa Craig. 2016. "Structure of the Neisseria Meningitidis Type IV Pilus." *Nature Communications* 7: 13015. <https://doi.org/10.1038/ncomms13015>.
- Koo, Jason, Ryan P Lamers, John L Rubinstein, Lori L Burrows, and Lynne P Howell. 2016. "Structure of the Pseudomonas Aeruginosa Type IVa Pilus Secretin at 7.4 Å." *Structure* 24: 1778–1787. <https://doi.org/10.1016/j.str.2016.08.007>.
- Kramer, Michaela A, Svava K Wetzl, Andreas Plückthun, Peer Mittl, and Markus G Grütter. 2010. "Structural Determinants for Improved Stability of Designed Ankyrin Repeat Proteins with a Redesigned C-Capping Module." *Journal of Molecular Biology* 404: 381–391. <https://doi.org/10.1016/j.jmb.2010.09.023>.
- Kratz, Peter A., Bettina Böttcher, and Michael Nassal. 1999. "Native Display of Complete Foreign Protein Domains on the Surface of Hepatitis B Virus Capsids." *Proc National Acad Sci* 96 (5): 1915–20. <https://doi.org/10.1073/pnas.96.5.1915>.

- Kremer, James R, David N Mastronarde, and J.Richard McIntosh. 1996. "Computer Visualization of Three-Dimensional Image Data Using IMOD." *Journal of Structural Biology* 116: 71–76. <https://doi.org/10.1006/jsbi.1996.0013>.
- Kucukelbir, Alp, Fred J Sigworth, and Hemant D Tagare. 2014a. "Quantifying the Local Resolution of Cryo-EM Density Maps." *Nature Methods* 11 (1): 63–65. <https://doi.org/10.1038/nmeth.2727>.
- Kucukelbir, Alp, Fred J. Sigworth, and Hemant D. Tagare. 2014b. "Quantifying the Local Resolution of Cryo-EM Density Maps." *Nature Methods* 11 (1): 63–65. <https://doi.org/10.1038/nmeth.2727>.
- Kudryashev, Mikhail, Ray Wang, Maximilian Brackmann, Sebastian Scherer, Timm Maier, David Baker, DiMaio Frank, Henning Stahlberg, Edward H Egelman, and Marek Basler. 2015. "Structure of the Type VI Secretion System Contractile Sheath" 160 (5): 952–62. <https://doi.org/10.1016/j.cell.2015.01.037>.
- Landau, Meytal, Itay Mayrose, Yossi Rosenberg, Fabian Glaser, Eric Martz, Tal Pupko, and Nir Ben-Tal. 2005. "ConSurf 2005: The Projection of Evolutionary Conservation Scores of Residues on Protein Structures." *Nucleic Acids Research* 33 (suppl_2): W299–302. <https://doi.org/10.1093/nar/gki370>.
- Liu, Yuxi, Shane Gonen, Tamir Gonen, and Todd O Yeates. 2018. "Near-Atomic Cryo-EM Imaging of a Small Protein Displayed on a Designed Scaffolding System." *Proceedings of the National Academy of Sciences* 115: 201718825. <https://doi.org/10.1073/pnas.1718825115>.
- Liu, Yuxi, Duc T Huynh, and Todd O Yeates. 2019. "A 3.8 Å Resolution Cryo-EM Structure of a Small Protein Bound to an Imaging Scaffold." *Nature Communications* 10: 1864. <https://doi.org/10.1038/s41467-019-09836-0>.
- Liu, Yuxi, Duc Huynh, and Todd O Yeates. 2018. "A 3.8 Å Resolution Cryo-EM Structure of a Small Protein Bound to a Modular Imaging Scaffold." *BioRxiv*, 505792. <https://doi.org/10.1101/505792>.
- Lomize, Andrei L., Irina Pogozheva, and Henry I Mosberg. 2011. "Anisotropic Solvent Model of the Lipid Bilayer. 2. Energetics of Insertion of Small Molecules, Peptides and Proteins in Membranes." *Journal of Chemical Information and Modeling* 51 (4): 930–46. <https://doi.org/10.1021/ci200020k>.
- Majewski, Dorothy D, Liam J Worrall, and Natalie CJ Strynadka. 2018. "Secretins Revealed: Structural Insights into the Giant Gated Outer Membrane Portals of Bacteria." *Current Opinion in Structural Biology* 51: 61–72. <https://doi.org/10.1016/j.sbi.2018.02.008>.
- Mastronarde, David N. 2005. "Automated Electron Microscope Tomography Using Robust Prediction of Specimen Movements." *Journal of Structural Biology* 152: 36–51. <https://doi.org/10.1016/j.jsb.2005.07.007>.
- Meibom, Karin L, Melanie Blokesch, Nadia A Dolganov, Cheng-Yen Wu, and Gary K Schoolnik. 2005. "Chitin Induces Natural Competence in *Vibrio*

- Cholerae*.” *Science* 310: 1824–1827.
<https://doi.org/10.1126/science.1120096>.
- Meibom, Karin L, Xibing B Li, Alex T Nielsen, Cheng-Yen Wu, Saul Roseman, and Gary K Schoolnik. 2004. “The *Vibrio Cholerae* Chitin Utilization Program.” *Proceedings of the National Academy of Sciences of the United States of America* 101: 2524–2529.
<https://doi.org/10.1073/pnas.0308707101>.
- Morin, Andrew, Ben Eisenbraun, Jason Key, Paul C Sanschagrin, Michael A Timony, Michelle Ottaviano, and Piotr Sliz. 2013. “Collaboration Gets the Most out of Software.” *ELife* 2: e01456. <https://doi.org/10.7554/elife.01456>.
- Naydenova, Katerina, and Christopher J. Russo. 2017. “Measuring the Effects of Particle Orientation to Improve the Efficiency of Electron Cryomicroscopy.” *Nature Communications* 8: 629.
<https://doi.org/10.1038/s41467-017-00782-3>.
- Neuhaus, Alexander, Muniyandi Selvaraj, Ralf Salzer, Julian D. Langer, Kerstin Kruse, Kelly Sanders, Bertram Daum, Beate Averhoff, and Vicki A. M. Gold. 2019. “A New Twist on Bacterial Motility – Two Distinct Type IV Pili Revealed by CryoEM.” *BioRxiv*, August, 720938.
<https://doi.org/10.1101/720938>.
- Noeske, Jonas, Michael R Wasserman, Daniel S Terry, Roger B Altman, Scott C Blanchard, and Jamie HD Cate. 2015. “High-Resolution Structure of the *Escherichia Coli* Ribosome” 22 (4): 336–41.
<https://doi.org/10.1038/nsmb.2994>.
- Orlova, EV, and HR Saibil. 2011. “Structural Analysis of Macromolecular Assemblies by Electron Microscopy” 111 (12): 7710–48.
<https://doi.org/10.1021/cr100353t>.
- Padilla, Jennifer E, Christos Colovos, and Todd O Yeates. 2001. “Nanohedra: Using Symmetry to Design Self Assembling Protein Cages, Layers, Crystals, and Filaments” 98 (5): 2217–21.
<https://doi.org/10.1073/pnas.041614998>.
- 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: 1605–1612.
<https://doi.org/10.1002/jcc.20084>.
- Phillips, Rob. 2017. “Membranes by the Numbers.” *ArXiv:1703.02066 [Cond-Mat, Physics:Physics, q-Bio]*, March. <http://arxiv.org/abs/1703.02066>.
- Pogozheva, Irina D, Henry I Mosberg, and Andrei L Lomize. 2014. “Life at the Border: Adaptation of Proteins to Anisotropic Membrane Environment.” *Protein Science: A Publication of the Protein Society* 23 (9): 1165–96.
<https://doi.org/10.1002/pro.2508>.
- Popot, Jean-Luc. 2010. “Amphipols, Nanodiscs, and Fluorinated Surfactants: Three Nonconventional Approaches to Studying Membrane Proteins in Aqueous Solutions.” *Annual Review of Biochemistry* 79: 737–775.
<https://doi.org/10.1146/annurev.biochem.052208.114057>.

- 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. <https://doi.org/10.1038/nmeth.4169>.
- “RCSB PDB: Homepage.” n.d. Accessed October 1, 2019. <https://www.rcsb.org/>.
- Rohou, Alexis, and Nikolaus Grigorieff. 2015. “CTFFIND4: Fast and Accurate Defocus Estimation from Electron Micrographs” 192 (2): 216–21. <https://doi.org/10.1016/j.jsb.2015.08.008>.
- Rosenthal, Peter B, and Richard Henderson. 2003. “Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-Particle Electron Cryomicroscopy.” *Journal of Molecular Biology* 333: 721–745. <https://doi.org/10.1016/j.jmb.2003.07.013>.
- Roy, Ambrish, Alper Kucukural, and Yang Zhang. 2010. “I-TASSER: A Unified Platform for Automated Protein Structure and Function Prediction.” *Nature Protocols* 5 (4): 725–38. <https://doi.org/10.1038/nprot.2010.5>.
- Scheres, Sjors. 2012a. “A Bayesian View on Cryo-EM Structure Determination” 415 (2): 406–18. <https://doi.org/10.1016/j.jmb.2011.11.010>.
- . 2012b. “RELION: Implementation of a Bayesian Approach to Cryo-EM Structure Determination” 180 (3): 519–30. <https://doi.org/10.1016/j.jsb.2012.09.006>.
- Schwanhäusser, Björn, Dorothea Busse, Na Li, Gunnar Dittmar, Johannes Schuchhardt, Jana Wolf, Wei Chen, and Matthias Selbach. 2011. “Global Quantification of Mammalian Gene Expression Control.” *Nature* 473 (7347): 337–42. <https://doi.org/10.1038/nature10098>.
- Seeger, Markus A, Reto Zbinden, Andreas Flütsch, Petrus GM Gutte, Sibylle Engeler, Heidi Roschitzki-Voser, and Markus G Grütter. 2013. “Design, Construction, and Characterization of a Second-generation DARPin Library with Reduced Hydrophobicity.” *Protein Science* 22: 1239–1257. <https://doi.org/10.1002/pro.2312>.
- Shoji, Shinichiro, Corey M Dambacher, Zahra Shajani, James R Williamson, and Peter G Schultz. 2011. “Systematic Chromosomal Deletion of Bacterial Ribosomal Protein Genes” 413 (4): 751–61. <https://doi.org/10.1016/j.jmb.2011.09.004>.
- Siewering, Katja, Samta Jain, Carmen Friedrich, Mariam T Webber-Birungi, Dmitry A Semchonok, Ina Binzen, Alexander Wagner, et al. 2014. “Peptidoglycan-Binding Protein TsaP Functions in Surface Assembly of Type IV Pili.” *Proceedings of the National Academy of Sciences* 111: E953–E961. <https://doi.org/10.1073/pnas.1322889111>.
- Souza, Robson Francisco de, Vivek Anantharaman, Sandro José de Souza, L. Aravind, and Frederico J. Gueiros-Filho. 2008. “AMIN Domains Have a Predicted Role in Localization of Diverse Periplasmic Protein Complexes.” *Bioinformatics* 24 (21): 2423–26. <https://doi.org/10.1093/bioinformatics/btn449>.
- Stillman, T. J., P. D. Hempstead, P. J. Artymiuk, S. C. Andrews, A. J. Hudson, A. Treffry, J. R. Guest, and P. M. Harrison. 2001. “The High-Resolution X-

- Ray Crystallographic Structure of the Ferritin (EcFtnA) of Escherichia Coli; Comparison with Human H Ferritin (HuHF) and the Structures of the Fe(3+) and Zn(2+) Derivatives.” *Journal of Molecular Biology* 307 (2): 587–603. <https://doi.org/10.1006/jmbi.2001.4475>.
- Tan, Yong, 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: 793–796. <https://doi.org/10.1038/nmeth.4347>.
- Taylor, KA, and RM Glaeser. 2008. “Retrospective on the Early Development of Cryoelectron Microscopy of Macromolecules and a Prospective on Opportunities for the Future.”
- “The Nobel Prize in Chemistry 2017.” 2017. Nobel Media AB. <https://www.nobelprize.org/prizes/chemistry/2017/summary/>.
- Tribet, Christophe, Roland Audebert, and Jean-Luc Popot. 1996. “Amphipols: Polymers That Keep Membrane Proteins Soluble in Aqueous Solutions” 93 (26): 15047–50. <https://doi.org/10.1073/pnas.93.26.15047>.
- Uchański, Tomasz, Simonas Masiulis, Baptiste Fischer, Valentina Kalichuk, Alexandre Wohlkönig, Thomas Zögg, Han Remaut, et al. 2019. “Megabodies Expand the Nanobody Toolkit for Protein Structure Determination by Single-Particle Cryo-EM.” *BioRxiv*, October, 812230. <https://doi.org/10.1101/812230>.
- Ullmann, A., F. Jacob, and J. Monod. 1967. “Characterization by in Vitro Complementation of a Peptide Corresponding to an Operator-Proximal Segment of the Beta-Galactosidase Structural Gene of Escherichia Coli.” *Journal of Molecular Biology* 24 (2): 339–43. [https://doi.org/10.1016/0022-2836\(67\)90341-5](https://doi.org/10.1016/0022-2836(67)90341-5).
- Votteler, Jörg, Cassandra Ogohara, Sue Yi, Yang Hsia, Una Nattermann, David M Belnap, Neil P King, and Wesley I Sundquist. 2016. “Designed Proteins Induce the Formation of Nanocage-Containing Extracellular Vesicles.” *Nature* 540 (7632): 292–95. <https://doi.org/10.1038/nature20607>.
- Wang, Fengbin, Mathieu Coureuil, Tomasz Osinski, Albina Orlova, Tuba Altindal, Gaël Gesbert, Xavier Nassif, Edward H Egelman, and Lisa Craig. 2017. “Cryoelectron Microscopy Reconstructions of the Pseudomonas Aeruginosa and Neisseria Gonorrhoeae Type IV Pili at Sub-Nanometer Resolution.” *Structure* 25: 1423–1435.e4. <https://doi.org/10.1016/j.str.2017.07.016>.
- Worrall, LJ, C Hong, M Vuckovic, W Deng, JRC Bergeron, DD Majewski, RK Huang, et al. 2016. “Near-Atomic-Resolution Cryo-EM Analysis of the Salmonella T3S Injectisome Basal Body.” *Nature* 540: 597. <https://doi.org/10.1038/nature20576>.
- Wu, Shenping, Agustin Avila-Sakar, JungMin Kim, David S Booth, Charles H Greenberg, Andrea Rossi, Maofu Liao, Xueming Li, Akram Alian, and Sarah L Griner. 2012. “Fabs Enable Single Particle CryoEM Studies of Small Proteins.” *Structure* 20: 582–592.

- Yan, Zhaofeng, Meng Yin, Dandan Xu, Yongqun Zhu, and Xueming Li. 2017. “Structural Insights into the Secretin Translocation Channel in the Type II Secretion System” 24 (2): 177–83. <https://doi.org/10.1038/nsmb.3350>.
- Yang, Jianyi, Renxiang Yan, Ambrish Roy, Dong Xu, Jonathan Poisson, and Yang Zhang. 2015. “The I-TASSER Suite: Protein Structure and Function Prediction.” *Nature Methods* 12 (1): 7–8. <https://doi.org/10.1038/nmeth.3213>.
- Yao, Qing, Qiuhe Lu, Xiaobo Wan, Feng Song, Yue Xu, Mo Hu, Alla Zamyatina, et al. 2014. “A Structural Mechanism for Bacterial Autotransporter Glycosylation by a Dodecameric Heptosyltransferase Family.” Edited by Wilfred van der Donk. *ELife* 3 (October): e03714. <https://doi.org/10.7554/eLife.03714>.
- Yin, Meng, Zhaofeng Yan, and Xueming Li. 2018. “Structural Insight into the Assembly of the Type II Secretion System Pilotin–Secretin Complex from Enterotoxigenic Escherichia Coli.” *Nature Microbiology* 3: 581–587. <https://doi.org/10.1038/s41564-018-0148-0>.
- Zhang, Yang. 2008. “I-TASSER Server for Protein 3D Structure Prediction.” *BMC Bioinformatics* 9 (1): 40. <https://doi.org/10.1186/1471-2105-9-40>.
- Zhang, Zhening, Wenguang G Liang, Lucas J Bailey, Yong Zi Tan, Hui Wei, Andrew Wang, Mara Farcasanu, et al. 2018. “Ensemble CryoEM Elucidates the Mechanism of Insulin Capture and Degradation by Human Insulin Degrading Enzyme.” *ELife* 7. <https://doi.org/10.7554/elife.33572>.
- 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*. <https://doi.org/10.1038/nmeth.4193>.
- Zhou, Mengying, Yini Li, Qi Hu, Xiao-Chen Bai, Weiyun Huang, Chuangye Yan, Sjors H. W. Scheres, and Yigong Shi. 2015. “Atomic Structure of the Apoptosome: Mechanism of Cytochrome c- and DATP-Mediated Activation of Apaf-1.” *Genes & Development* 29 (22): 2349–61. <https://doi.org/10.1101/gad.272278.115>.
- 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: e42166. <https://doi.org/10.7554/elife.42166>.
- Zivanov, Jasenko, Takanori Nakane, and Sjors H. W. Scheres. 2019. “Estimation of High-Order Aberrations and Anisotropic Magnification from Cryo-EM Datasets in RELION-3.1.” *BioRxiv*, October, 798066. <https://doi.org/10.1101/798066>.